PRINCETON SERIES IN FINANCE

QUANTITATIVE RISK MANAGEMENT

CONCEPTS, TECHNIQUES AND TOOLS

Alexander J. McNeil, Rüdiger Frey and Paul Embrechts

REVISED EDITION

Quantitative Risk Management

Quantitative Risk Management: Concepts, Techniques and Tools
IS A PART OF THE
PRINCETON SERIES IN FINANCE

SERIES EDITORS

Darrell Duffie — *Stanford University*
Stephen Schaefer — *London Business School*

Finance as a discipline has been growing rapidly. The numbers of researchers in academy and industry, of students, of methods and models have all proliferated in the past decade or so. This growth and diversity manifests itself in the emerging cross-disciplinary as well as cross-national mix of scholarship now driving the field of finance forward. The intellectual roots of modern finance, as well as the branches, will be represented in the Princeton Series in Finance.

Titles in this series will be scholarly and professional books, intended to be read by a mixed audience of economists, mathematicians, operations research scientists, financial engineers, and other investment professionals. The goal is to provide the finest cross-disciplinary work in all areas of finance by widely recognized researchers in the prime of their creative careers.

OTHER BOOKS IN THIS SERIES

Financial Econometrics: Problems, Models, and Methods by Christian Gourieroux and Joann Jasiak

Credit Risk: Pricing, Measurement, and Management by Darrell Duffie and Kenneth J. Singleton

Microfoundations of Financial Economics: An Introduction to General Equilibrium Asset Pricing by Yvan Lengwiler

Credit Risk Modeling: Theory and Applications by David Lando

Quantitative Risk Management: Concepts, Techniques and Tools, revised edition, by Alexander J. McNeil, Rüdiger Frey and Paul Embrechts

Quantitative Risk Management

Concepts, Techniques and Tools

Revised Edition

Alexander J. McNeil

Rüdiger Frey

Paul Embrechts

Princeton University Press
Princeton and Oxford

Published by Princeton University Press,
41 William Street, Princeton, New Jersey 08540

In the United Kingdom: Princeton University Press,
6 Oxford Street, Woodstock, Oxfordshire OX20 1TW

press.princeton.edu

Jacket art © Michael D. Brown/Shutterstock

ISBN 978-0-691-16627-8

Library of Congress Control Number: 2015934589

British Library Cataloguing-in-Publication Data is available

This book has been composed in Times and typeset
by T&T Productions Ltd, London

Printed on acid-free paper ♾

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To Janine, Alexander and Calliope
Alexander

Für Catharina, Sebastian und Michaela
Rüdiger

Voor Levi, Mila, Ben en Marlon
Paul

Contents

Preface

Why have we written this book? In recent decades the field of financial risk management has developed rapidly in response to both the increasing complexity of financial instruments and markets and the increasing regulation of the financial services industry. This book is devoted specifically to *quantitative* modelling issues arising in this field. As a result of our own discussions and joint projects with industry professionals and regulators over a number of years, we felt there was a need for a textbook treatment of quantitative risk management (QRM) at a technical yet accessible level, aimed at both industry participants and students seeking an entrance to the area.

We have tried to bring together a body of methodology that we consider to be core material for any course on the subject. This material and its mode of presentation represent the blending of our own views, which come from the perspectives of financial mathematics, insurance mathematics and statistics. We feel that a book combining these viewpoints fills a gap in the existing literature and emphasises the fact that there is a need for quantitative risk managers in banks, insurance companies and beyond to have broad, interdisciplinary skills.

What is new in this second edition? The second edition of this book has been extensively revised and expanded to reflect the continuing development of QRM methodology since the 2005 first edition. This period included the 2007–9 financial crisis, during which much of the methodology was severely tested. While we have added to the detail, we are encouraged that we have not had to revise the main messages of the first edition in the light of the crisis. In fact, many of those messages—the importance of extremes and extremal dependence, systematic risk and the model risk inherent in portfolio credit models—proved to be central issues in the crisis.

Whereas the first edition had a Basel and banking emphasis, we have added more material relevant to Solvency II and insurance in the second edition. Moreover, the methodological chapters now start at the natural starting point: namely, a discussion of the balance sheets and business models of a bank and an insurer.

This edition contains an extended treatment of credit risk in four chapters, including new material on portfolio credit derivatives and counterparty credit risk. There is a new market-risk chapter, bringing together more detail on mapping portfolios to market-risk factors and applying and backtesting statistical methods. We have also extended the treatment of the fundamental topics of risk measures and risk aggregation.

We have revised the structure of the book to facilitate teaching. The chapters are a little shorter than in the first edition, with more advanced or specialized material

now placed in a series of "Special Topics" chapters at the end. The book is split into four parts: (I) An Introduction to Quantitative Risk Management, (II) Methodology, (III) Applications, (IV) Special Topics.

Who was this book written for? This book is primarily a textbook for courses on QRM aimed at advanced undergraduate or graduate students and professionals from the financial industry. *A knowledge of probability and statistics at least at the level of a first university course in a quantitative discipline and familiarity with undergraduate calculus and linear algebra are fundamental prerequisites*. Though not absolutely necessary, some prior exposure to finance, economics or insurance will be beneficial for a better understanding of some sections.

The book has a secondary function as a reference text for risk professionals interested in a clear and concise treatment of concepts and techniques that are used in practice. As such, we hope it will facilitate communication between regulators, end-users and academics.

A third audience for the book is the community of researchers that work in the area. Most chapters take the reader to the frontier of current, practically relevant research and contain extensive, annotated references that guide the reader through the vast literature.

Ways to use this book. The material in this book has been tested on many different audiences, including undergraduate and postgraduate students at ETH Zurich, the Universities of Zurich and Leipzig, Heriot-Watt University, the London School of Economics and the Vienna University of Economics and Business. It has also been used for professional training courses aimed at risk managers, actuaries, consultants and regulators. Based on this experience we can suggest a number of ways of using the book.

A taught course would generally combine material from Parts I, II and III, although the exact choice of material from Parts II and III would depend on the emphasis of the course. Chapters 2 and 3 from Part I would generally be core taught modules, whereas Chapter 1 might be prescribed as background reading material.

A general course on QRM could be based on a complete treatment of Parts I–III. This would require a minimum of two semesters, with 3–4 hours of taught courses per week for an introductory course and longer for a detailed treatment. A quantitative course on enterprise risk management for actuaries would follow a very similar selection, probably omitting material from Chapters 11 and 12, which contain Basel-specific details of portfolio credit risk modelling and an introduction to portfolio credit derivatives.

For a course on credit risk modelling, there is a lot of material to choose from. A comprehensive course spanning two semesters would include Part I (probably omitting Chapter 3), Chapters 6 and 7 from Part II, and Chapters 10–12 from Part III. Material on counterparty credit risk (Chapter 17) might also be included from Part IV.

A one-semester, specialized course on market risk could be based on Part I, Chapters 4–6 from Part II, and Chapter 9 from Part III. An introduction to risk

management for financial econometricians could follow a similar selection but might cover all the chapters in Part II.

It is also possible to devise more specialized courses, such as a course on risk-measurement and aggregation concepts based on Chapters 2, 7 and 8. Moreover, material from various chapters could be used as interesting examples to enliven statistics courses on subjects like multivariate analysis, time-series analysis and generalized linear modelling. In Part IV there are a number of potential topics for seminars at postgraduate and PhD level.

What we have not covered. We have not been able to address all the topics that a reader might expect to find under the heading of QRM. Perhaps the most obvious omission is the lack of a section on the risk management of derivatives by hedging. Here we felt that the relevant techniques, and the financial mathematics required to understand them, are already well covered in a number of excellent textbooks. Other omissions include modelling techniques for price liquidity risk and models for systemic risk in national and global networks of financial firms, both of which have been areas of research since the 2007–9 crisis. Besides these larger areas, many smaller issues have been neglected for reasons of space but are mentioned with suggestions for further reading in the "Notes and Comments" sections, which should be considered as integral parts of the text.

Acknowledgements. The origins of this book date back to 1996, when A.M. and R.F. began postdoctoral studies in the group of P.E. at the Federal Institute of Technology (ETH) in Zurich. All three authors are grateful to ETH for providing the environment in which the project initially flourished. A.M. and R.F. thank Swiss Re and UBS, respectively, for providing the financial support for their postdoctoral positions. P.E. thanks the Swiss Finance Institute, which continues to provide support through a Senior SFI Professorship.

A.M. acknowledges the support of Heriot-Watt University, where he now works, and thanks the Isaac Newton Institute for Mathematical Sciences for a Visiting Fellowship during which he was able to put the finishing touches to the manuscript. R.F. has subsequently held positions at the Swiss Banking Institute of the University of Zurich, the University of Leipzig and the Vienna University of Economics and Business, and he is grateful to all these institutions for their support. P.E. thanks the London School of Economics, where he enjoyed numerous fruitful discussions with colleagues during his time as Centennial Professor of Finance, and the Oxford-Man Institute at the University of Oxford for hospitality during his visit as OMI Visiting Man Chair.

The Forschungsinstitut für Mathematik of ETH Zurich provided generous financial support throughout the project and facilitated meetings of all three authors in Zurich on many occasions. At a crucial juncture in early 2004 the Mathematisches Forschungsinstitut Oberwolfach was the venue for a memorable week of progress. We also acknowledge the invaluable contribution of RiskLab Zurich to the enterprise: the agenda for the book was strongly influenced by joint projects and discussions

with the RiskLab sponsors UBS, Credit Suisse and Swiss Re. We have also benefited greatly from the NCCR FINRISK research programme in Switzerland, which funded doctoral and postdoctoral research on topics in the book.

We are indebted to numerous readers who have commented on various parts of the manuscript, and to colleagues in Zurich, Leipzig, Edinburgh, Vienna and beyond who have helped us in our understanding of QRM and the mathematics underlying it. These include Stefan Altner, Philippe Artzner, Jochen Backhaus, Guus Balkema, Michał Barski, Uta Beckmann, Reto Baumgartner, Wolfgang Breymann, Reto Bucher, Hans Bühlmann, Peter Bühlmann, Valérie Chavez-Demoulin, Dominik Colangelo, Marius Costeniuc, Enrico De Giorgi, Freddy Delbaen, Rosario Dell'Aquila, Stefano Demarta, Stefan Denzler, Alexandra Dias, Catherine Donnelly, Douglas Dwyer, Damir Filipovic, Tom Fischer, Gabriel Frahm, Hansjörg Furrer, Rajna Gibson, Kay Giesecke, Michael Gordy, Bernhard Hodler, Marius Hofert, Andrea Höing, Kurt Hornik, Friedrich Hubalek, Christoph Hummel, Edgars Jakobsons, Alessandro Juri, Roger Kaufmann, Philipp Keller, Erwan Koch, Hans Rudolf Künsch, Filip Lindskog, Hans-Jakob Lüthi, Natalia Markovich, Benoît Metayer, Andres Mora, Alfred Müller, Johanna Nešlehová, Monika Popp, Giovanni Puccetti, Lars Rösler, Wolfgang Runggaldier, David Saunders, Hanspeter Schmidli, Sylvia Schmidt, Thorsten Schmidt, Uwe Schmock, Philipp Schönbucher, Martin Schweizer, Torsten Steiger, Daniel Straumann, Dirk Tasche, Hideatsu Tsukahara, Laura Vana, Eduardo Vilela, Marcel Visser, Ruodu Wang, Jonathan Wendin and Mario Wüthrich. For their help in preparing the manuscript we thank Gabriele Baltes and Edgars Jakobsons. We should also, of course, thank the scores of (unnamed) students who took QRM lectures based on this book and contributed through their critical questions and remarks.

We thank the team at Princeton University Press for all their help in the production of this book, particularly our editors Richard Baggaley (first edition) and Hannah Paul (second edition). We are also grateful to anonymous referees who provided us with exemplary feedback, which has shaped this book for the better. Special thanks go to Sam Clark at T&T Productions Ltd, who took our uneven LaTeX code and turned it into a more polished book with remarkable speed and efficiency.

We are delighted that the first edition appeared in a Japanese translation and thank the translation team (Hideatsu Tsukahara, Shun Kobayashi, Ryozo Miura, Yoshinori Kawasaki, Hiroaki Yamauchi and Hidetoshi Nakagawa) for their work on this project, their valuable feedback on the book and their hospitality to A.M. at the Japanese book launch.

To our wives, Janine, Catharina and Gerda, and our families our sincerest debt of gratitude is due. Though driven to distraction no doubt by our long contemplation of risk, without obvious reward, their support was constant.

Further resources. Readers are encouraged to visit the book's homepage at

www.pupress.princeton.edu/titles/8056.html

and the site

www.qrmtutorial.org

where they will find supplementary resources for this book. We are particularly grateful to Marius Hofert, not only for his proofreading, but also for his help in developing slides, exercises and R libraries and scripts to illustrate many of the topics in the book.

Special abbreviations. A number of abbreviations for common terms in probability are used throughout the book; these include "rv" for random variable, "df" for distribution function, "iid" for independent and identically distributed and "se" for standard error.

Part I

An Introduction to Quantitative Risk Management

1
Risk in Perspective

In this chapter we provide a non-mathematical discussion of various issues that form the background to the rest of the book. In Section 1.1 we begin with the nature of risk itself and discuss how risk relates to randomness; in the financial context (which includes insurance) we summarize the main kinds of risks encountered and explain what it means to measure and manage such risks.

A brief history of financial risk management and the development of financial regulation is given in Section 1.2, while Section 1.3 contains a summary of the regulatory framework in the financial and insurance industries.

In Section 1.4 we take a step back and attempt to address the fundamental question of why we might want to measure and manage risk at all. Finally, in Section 1.5 we turn to quantitative risk management (QRM) explicitly and set out our own views concerning the nature of this discipline and the challenge it poses. This section in particular should give more insight into our choice of methodological topics in the rest of the book.

1.1 Risk

The *Concise Oxford English Dictionary* defines risk as "hazard, a chance of bad consequences, loss or exposure to mischance". In a discussion with students taking a course on financial risk management, ingredients that are typically discussed are events, decisions, consequences and uncertainty. It is mostly only the downside of risk that is mentioned, rarely a possible upside, i.e. the potential for a gain. While for many people risk has largely negative connotations, it may also represent an opportunity. Much of the financial industry would not exist were it not for the presence of financial risk and the opportunities afforded to companies that are able to create products and services that offer more financial certainty to their clients.

For financial risks no single one-sentence definition of risk is entirely satisfactory. Depending on context, one might arrive at notions such as "any event or action that may adversely affect an organization's ability to achieve its objectives and execute its strategies" or, alternatively, "the quantifiable likelihood of loss or less-than-expected returns".

1.1.1 Risk and Randomness

Regardless of context, risk strongly relates to uncertainty, and hence to the notion of randomness. Randomness has eluded a clear, workable definition for many centuries;

it was not until 1933 that the Russian mathematician A. N. Kolmogorov gave an axiomatic definition of randomness and probability (see Kolmogorov 1933). This definition and its accompanying theory provide the language for the majority of the literature on risk, including this book.

Our reliance on probability may seem unsatisfactorily narrow to some. It bypasses several of the current debates on risk and uncertainty (Frank Knight), the writings on probabilistic thinking within economics (John Maynard Keynes), the unpredictability of unprecedented financial shocks, often referred to as Black Swans (Nassim Taleb), or even the more political expression of the known, the unknown and the unknowable (Donald Rumsfeld); see the Notes and Comments section for more explanation. Although these debates are interesting and important, at some point clear definitions and arguments are called for and this is where mathematics as a language enters. The formalism of Kolmogorov, while not the only possible approach, is a tried-and-tested framework for mathematical reasoning about risk.

In Kolmogorov's language a probabilistic model is described by a triplet $(\Omega, \mathcal{F}, P)$. An element ω of Ω represents a realization of an experiment, in economics often referred to as a state of nature. The statement "the probability that an event A occurs" is denoted (and in Kolmogorov's axiomatic system defined) as $P(A)$, where A is an element of $\mathcal{F}$, the set of all events. P denotes the probability measure. For the less mathematically trained reader it suffices to accept that Kolmogorov's system translates our intuition about randomness into a concise, axiomatic language and clear rules.

Consider the following examples: an investor who holds stock in a particular company; an insurance company that has sold an insurance policy; an individual who decides to convert a fixed-rate mortgage into a variable one. All of these situations have something important in common: the investor holds today an asset with an uncertain future value. This is very clear in the case of the stock. For the insurance company, the policy sold may or may not be triggered by the underlying event covered. In the case of a mortgage, our decision today to enter into this refinancing agreement will change (for better or for worse) the future repayments. So randomness plays a crucial role in the valuation of current products held by the investor, the insurance company and the home owner.

To model these situations a mathematician would now define the value of a risky position X to be a function on the probability space $(\Omega, \mathcal{F}, P)$; this function is called a *random variable*. We leave for the moment the range of X (i.e. its possible values) unspecified. Most of the modelling of a risky position X concerns its *distribution function* $F_X(x) = P(X \leqslant x)$: the probability that by the end of the period under consideration the value of the risk X is less than or equal to a given number x. Several risky positions would then be denoted by a random vector $(X_1, \ldots, X_d)$, also written in bold face as $\boldsymbol{X}$; time can be introduced, leading to the notion of random (or so-called stochastic) processes, usually written $(\boldsymbol{X}_t)$. Throughout this book we will encounter many such processes, which serve as essential building blocks in the mathematical description of risk.

We therefore expect the reader to be at ease with basic notation, terminology and results from elementary *probability and statistics*, the branch of mathematics dealing with *stochastic* models and their application to the real world. The word "stochastic" is derived from the Greek "stochazesthai", the art of guessing, or "stochastikos", meaning skilled at aiming ("stochos" being a target). In discussing stochastic methods for risk management we hope to emphasize the skill aspect rather than the guesswork.

1.1.2 Financial Risk

In this book we discuss risk in the context of finance and insurance (although many of the tools introduced are applicable well beyond this context). We start by giving a brief overview of the main risk types encountered in the financial industry.

The best-known type of risk is probably *market risk*: the risk of a change in the value of a financial position or portfolio due to changes in the value of the underlying components on which that portfolio depends, such as stock and bond prices, exchange rates, commodity prices, etc. The next important category is *credit risk*: the risk of not receiving promised repayments on outstanding investments such as loans and bonds, because of the "default" of the borrower. A further risk category is *operational risk*: the risk of losses resulting from inadequate or failed internal processes, people and systems, or from external events.

The three risk categories of market, credit and operational risk are the main ones we study in this book, but they do not form an exhaustive list of the full range of possible risks affecting a financial institution, nor are their boundaries always clearly defined. For example, when a corporate bond falls in value this is market risk, but the fall in value is often associated with a deterioration in the credit quality of the issuer, which is related to credit risk. The ideal way forward for a successful handling of financial risk is a *holistic* approach, i.e. an integrated approach taking all types of risk and their interactions into account.

Other important notions of risk are *model risk* and *liquidity risk*. The former is the risk associated with using a misspecified (inappropriate) model for measuring risk. Think, for instance, of using the Black–Scholes model for pricing an exotic option in circumstances where the basic Black–Scholes model assumptions on the underlying securities (such as the assumption of normally distributed returns) are violated. It may be argued that model risk is always present to some degree.

When we talk about liquidity risk we are generally referring to price or market liquidity risk, which can be broadly defined as the risk stemming from the lack of marketability of an investment that cannot be bought or sold quickly enough to prevent or minimize a loss. Liquidity can be thought of as "oxygen for a healthy market"; a market requires it to function properly but most of the time we are not aware of its presence. Its absence, however, is recognized immediately, with often disastrous consequences.

In banking, there is also the concept of *funding liquidity risk*, which refers to the ease with which institutions can raise funding to make payments and meet withdrawals as they arise. The management of funding liquidity risk tends to be

a specialist activity of bank treasuries (see, for example, Choudhry 2012) rather than trading-desk risk managers and is not a subject of this book. However, funding liquidity and market liquidity can interact profoundly in periods of financial stress. Firms that have problems obtaining funding may sell assets in fire sales to raise cash, and this in turn can contribute to market illiquidity, depressing prices, distorting the valuation of assets on balance sheets and, in turn, making funding even more difficult to obtain; this phenomenon has been described as a liquidity spiral (Brunnermeier and Pedersen 2009).

In insurance, a further risk category is *underwriting risk*: the risk inherent in insurance policies sold. Examples of risk factors that play a role here are changing patterns of natural catastrophes, changes in demographic tables underlying (long-dated) life products, political or legal interventions, or customer behaviour (such as lapsation).

1.1.3 Measurement and Management

Much of this book is concerned with techniques for the statistical measurement of risk, an activity which is part of the process of managing risk, as we attempt to clarify in this section.

Risk measurement. Suppose we hold a portfolio consisting of d underlying investments with respective weights $w_1, \ldots, w_d$, so that the change in value of the portfolio over a given holding period (the so-called profit and loss, or P&L) can be written as $X = \sum_{i=1}^{d} w_i X_i$, where X_i denotes the change in value of the ith investment. Measuring the risk of this portfolio essentially consists of determining its distribution function $F_X(x) = P(X \leqslant x)$, or functionals describing this distribution function such as its mean, variance or 99th percentile.

In order to achieve this, we need a properly calibrated *joint* model for the underlying random vector of investments $(X_1, \ldots, X_d)$, so statistical methodology has an important role to play in risk measurement; based on historical observations and given a specific model, a statistical estimate of the distribution of the change in value of a position, or one of its functionals, is calculated. In Chapter 2 we develop a detailed framework framework for risk measurement. As we shall see—and this is indeed a main theme throughout the book—this is by no means an easy task with a unique solution.

It should be clear from the outset that good risk measurement is essential. Increasingly, the clients of financial institutions demand objective and detailed information on the products that they buy, and firms can face legal action when this information is found wanting. For any product sold, a proper quantification of the underlying risks needs to be explicitly made, allowing the client to decide whether or not the product on offer corresponds to his or her risk appetite; the 2007–9 crisis saw numerous violations of this basic principle. For more discussion of the importance of the quantitative approach to risk, see Section 1.5.

Risk management. In a very general answer to the question of what risk management is about, Kloman (1990) writes:

> To many analysts, politicians, and academics it is the management of environmental and nuclear risks, those technology-generated macro-risks that appear to threaten our existence. To bankers and financial officers it is the sophisticated use of such techniques as currency hedging and interest-rate swaps. To insurance buyers or sellers it is coordination of insurable risks and the reduction of insurance costs. To hospital administrators it may mean "quality assurance". To safety professionals it is reducing accidents and injuries. In summary, risk management is *a discipline for living with the possibility that future events may cause adverse effects*.

The last phrase in particular (the emphasis is ours) captures the general essence of risk management: it is about ensuring *resilience* to future events. For a financial institution one can perhaps go further. A financial firm's attitude to risk is not passive and defensive; a bank or insurer actively and willingly takes on risk, because it seeks a return and this does not come without risk. Indeed, risk management can be seen as the core competence of an insurance company or a bank. By using its expertise, market position and capital structure, a financial institution can manage risks by repackaging or bundling them and transferring them to markets in customized ways.

The management of risk at financial institutions involves a range of tasks. To begin with, an enterprise needs to determine the capital it should hold to absorb losses, both for regulatory and economic capital purposes. It also needs to manage the risk on its books. This involves ensuring that portfolios are well diversified and optimizing portfolios according to risk–return considerations. The risk profile of the portfolio can be altered by hedging exposures to certain risks, such as interest-rate or foreign-exchange risk, using derivatives. Alternatively, some risks can be repackaged and sold to investors in a process known as securitization; this has been applied to both insurance risks (weather derivatives and longevity derivatives) and credit risks (mortgage-backed securities, collateralized debt obligations). Firms that use derivatives need to manage their derivatives books, which involves the tasks of pricing, hedging and managing collateral for such trades. Finally, financial institutions need to manage their counterparty credit risk exposures to important trading partners; these arise from bilateral, over-the-counter derivatives trades, but they are also present, for example, in reinsurance treaties.

We also note that the discipline of risk management is very much the core competence of an actuary. Indeed, the Institute and Faculty of Actuaries has used the following definition of the actuarial profession:

> Actuaries are respected professionals whose innovative approach to making business successful is matched by a responsibility to the public interest. Actuaries identify solutions to financial problems. They manage assets and liabilities by analysing past events, assessing the present risk involved and modelling what could happen in the future.

Actuarial organizations around the world have collaborated to create the Chartered Enterprise Risk Actuary qualification to show their commitment to establishing best practice in risk management.

1.2 A Brief History of Risk Management

In this section we treat the historical development of risk management by sketching some of the innovations and some of the events that have shaped modern risk management for the financial industry. We also describe the more recent development of regulation in the industry, which has, to some extent, been a process of reaction to a series of incidents and crises.

1.2.1 From Babylon to Wall Street

Although risk management has been described as "one of the most important innovations of the 20th century" by Steinherr (1998), and most of the story we tell is relatively modern, some concepts that are used in modern risk management, and in derivatives in particular, have been around for longer. In our selective account we stress the example of financial derivatives as these have played a role in many of the events that have shaped modern regulation and increased the complexity of the risk-management challenge.

The ancient world to the twentieth century. A derivative is a financial instrument derived from an underlying asset, such as an option, future or swap. For example, a European call option with strike K and maturity T gives the holder the right, but not the obligation, to obtain from the seller at maturity the underlying security for a price K; a European put option gives the holder the right to dispose of the underlying at a price K.

Dunbar (2000) interprets a passage in the Code of Hammurabi from Babylon of 1800 BC as being early evidence of the use of the option concept to provide financial cover in the event of crop failure. A very explicit mention of options appears in Amsterdam towards the end of the seventeenth century and is beautifully narrated by Joseph de la Vega in his 1688 *Confusión de Confusiones*, a discussion between a lawyer, a trader and a philosopher observing the activity on the Beurs of Amsterdam. Their discussion contains what we now recognize as European call and put options and a description of their use for investment as well as for risk management—it even includes the notion of short selling. In an excellent recent translation (de la Vega 1996) we read:

> If I may explain "opsies" [further, I would say that] through the payment of the premiums, one hands over values in order to safeguard one's stock or to obtain a profit. One uses them as sails for a happy voyage during a beneficent conjuncture and as an anchor of security in a storm.

After this, de la Vega continues with some explicit examples that would not be out of place in any modern finance course on the topic.

Financial derivatives in general, and options in particular, are not so new. Moreover, they appear here as instruments to manage risk, "anchors of security in a

storm", rather than as dangerous instruments of speculation, the "wild beasts of finance" (Steinherr 1998), that many believe them to be.

Academic innovation in the twentieth century. While the use of risk-management ideas such as derivatives can be traced further back, it was not until the late twentieth century that a theory of valuation for derivatives was developed. This can be seen as perhaps the most important milestone in an age of academic developments in the general area of quantifying and managing financial risk.

Before the 1950s, the desirability of an investment was mainly equated to its return. In his groundbreaking publication of 1952, Harry Markowitz laid the foundation of the theory of portfolio selection by mapping the desirability of an investment onto a risk–return diagram, where risk was measured using standard deviation (see Markowitz 1952, 1959). Through the notion of an *efficient frontier* the portfolio manager could optimize the return for a given risk level. The following decades saw explosive growth in risk-management methodology, including such ideas as the Sharpe ratio, the Capital Asset Pricing Model (CAPM) and Arbitrage Pricing Theory (APT). Numerous extensions and refinements that are now taught in any MBA course on finance followed.

The famous Black–Scholes–Merton formula for the price of a European call option appeared in 1973 (see Black and Scholes 1973). The importance of this formula was underscored in 1997 when the Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel was awarded to Robert Merton and Myron Scholes (Fischer Black had died some years earlier) "for a new method to determine the value of derivatives".

In the final two decades of the century the mathematical finance literature developed rapidly, and many ideas found their way into practice. Notable contributions include the pioneering papers by Harrison and Kreps (1979) and Harrison and Pliska (1981) clarifying the links between no-arbitrage pricing and martingale theory. A further example is the work on the term structure of interest rates by Heath, Jarrow and Morton (1992). These and other papers elaborated the mathematical foundations of financial mathematics. Textbooks on stochastic integration and Itô calculus became part of the so-called quant's essential reading and were, for a while, as likely to be seen in the hands of a young investment banker as the *Financial Times*.

Growth of markets in the twentieth century. The methodology developed for the rational pricing and hedging of financial derivatives changed finance. The "wizards of Wall Street" (i.e. the mathematical specialists conversant in the new methodology) have had a significant impact on the development of financial markets over the last few decades. Not only did the new option-pricing formula work, it transformed the market. When the Chicago Options Exchange first opened in 1973, fewer than a thousand options were traded on the first day. By 1995, over a million options were changing hands each day, with current nominal values outstanding in the derivatives markets in the tens of trillions. So great was the role played by the Black–Scholes–Merton formula in the growth of the new options market that, when the American stock market crashed in 1987, the influential business magazine *Forbes* attributed

the blame squarely to that one formula. Scholes himself has said that it was not so much the formula that was to blame, but rather that market traders had not become sufficiently sophisticated in using it.

Along with academic innovation, developments in information technology (IT) also helped lay the foundations for an explosive growth in the volume of new risk-management and investment products. This development was further aided by worldwide deregulation in the 1980s. Important additional factors contributing to an increased demand for risk-management skills and products were the oil crises of the 1970s and the 1970 abolition of the Bretton Woods system of fixed exchange rates. Both energy prices and foreign exchange risk became highly volatile risk factors and customers required products to hedge them. The 1933 Glass–Steagall Act—passed in the US in the aftermath of the 1929 Depression to prohibit commercial banks from underwriting insurance and most kinds of securities—indirectly paved the way for the emergence of investment banks, hungry for new business. Glass–Steagall was replaced in 1999 by the Financial Services Act, which repealed many of the former's key provisions, although the 2010 Dodd–Frank Act, passed in the aftermath of the 2007–9 financial crisis, appears to mark an end to the trend of deregulation.

Disasters of the 1990s. In January 1992 the president of the New York Federal Reserve, E. Gerald Corrigan, speaking at the Annual Mid-Winter Meeting of the New York State Bankers Association, said:

> You had all better take a very, very hard look at off-balance-sheet activities. The growth and complexity of [these] activities and the nature of the credit settlement risk they entail should give us cause for concern.... I hope this sounds like a warning, because it is. Off-balance-sheet activities [i.e. derivatives] have a role, but they must be managed and controlled carefully and they must be understood by top management as well as by traders and rocket scientists.

Corrigan was referring to the growing volume of derivatives in banks' trading books and the fact that, in many cases, these did not appear as assets or liabilities on the balance sheet. His words proved prescient.

On 26 February 1995 Barings Bank was forced into administration. A loss of £700 million ruined the oldest merchant banking group in the UK (established in 1761). Besides numerous operational errors (violating every qualitative guideline in the risk-management handbook), the final straw leading to the downfall of Barings was a so-called straddle position on the Nikkei held by the bank's Singapore-based trader Nick Leeson. A straddle is a short position in a call and a put with the same strike—such a position allows for a gain if the underlying (in this case the Nikkei index) does not move too far up or down. There is, however, considerable loss potential if the index moves down (or up) by a large amount, and this is precisely what happened when the Kobe earthquake occurred.

Three years later, Long-Term Capital Management (LTCM) became another prominent casualty of losses due to derivatives trading when it required a $3.5 billion payout to prevent collapse, a case made all the more piquant by the fact that

Myron Scholes and Robert Merton were principals at the hedge fund. Referring to the Black–Scholes formula, an article in the *Observer* newspaper asked: "Is this really the key to future wealth? Win big, lose bigger."

There were other important cases in this era, leading to a widespread discussion of the need for increased regulation, including Metallgesellschaft in 1993 (speculation on oil prices using derivatives) and Orange County in 1994 (speculation on interest rates using derivatives).

In the life insurance industry, Equitable Life, the world's oldest mutual insurer, provided a case study of what can happen when the liabilities arising from insurance products with embedded options are not properly hedged. Prior to 1988, Equitable Life had sold pension products that offered the option of a guaranteed annuity rate at maturity of the policy. The guarantee rate of 7% had been set in the 1970s when inflation and annuity rates were high, but in 1993 the current annuity rate fell below the guarantee rate and policyholders exercised their options. Equitable Life had not been hedging the option and it quickly became evident that they were faced with an enormous increase in their liabilities; the Penrose Report (finally published in March 2004) concluded that Equitable Life was underfunded by around £4.5 billion by 2001. It was the policyholders who suffered when the company reneged on their pension promises, although many of the company's actions were later ruled unlawful and some compensation from the public purse was agreed. However, this case provides a good illustration of the need to regulate the capital adequacy of insurers to protect policyholders.

The turn of the century. The end of the twentieth century proved to be a pivotal moment for the financial system worldwide. From a value of around 1000 in 1996, the Nasdaq index quintupled to a maximum value of 5408.62 on 10 March 2000 (which remains unsurpassed as this book goes to press). The era 1996–2000 is now known as the dot-com bubble because many of the firms that contributed to the rise in the Nasdaq belonged to the new internet sector.

In a speech before the American Enterprise Institute on 5 December 1996, Alan Greenspan, chairman of the Federal Reserve from 1987 to 2006, said, "But how do we know when *irrational exuberance* has unduly escalated assets, which then become subject to prolonged contractions as they have in Japan over the past decade?" The term irrational exuberance seemed to perfectly describe the times. The Dow Jones Industrial Average was also on a historic climb, breaking through the 10 000 barrier on 29 March 1999, and prompting books with titles like *Dow 40 000: Strategies for Profiting from the Greatest Bull Market in History*. It took four years for the bubble to burst, but from its March 2000 maximum the Nasdaq plummeted to half of its value within a year and tested the 1000 barrier in late 2002. Equity indices fell worldwide, although markets recovered and began to surge ahead again from 2004.

The dot-com bubble was in many respects a conventional asset bubble, but it was also during this period that the seeds of the next financial crisis were being sown. Financial engineers had discovered the magic of securitization: the bundling and repackaging of many risks into securities with defined risk profiles that could be

sold to potential investors. While the idea of transferring so-called tranches of a pool of risks to other risk bearers was well known to the insurance world, it was now being applied on a massive scale to credit-risky assets, such as mortgages, bonds, credit card debt and even student loans (see Section 12.1.1 for a description of the tranching concept).

In the US, the subprime lending boom to borrowers with low credit ratings fuelled the supply of assets to securitize and a market was created in mortgage-backed securities (MBSs). These in turn belonged to the larger pool of assets that were available to be transformed into collateralized debt obligations (CDOs). The banks originating these credit derivative products had found a profitable business turning poor credit risks into securities. The volume of credit derivatives ballooned over a very short period; the CDO market accounted for almost $3 trillion in nominal terms by 2008 but this was dwarfed by the nominal value of the credit default swap (CDS) market, which stood at about $30 trillion.

Credit default swaps, another variety of credit derivative, were originally used as instruments for hedging large corporate bond exposures, but they were now increasingly being used by investors to speculate on the changing credit outlook of companies by adopting so-called naked positions (see Section 10.1.4 for more explanation). Although the actual economic value of CDS and CDO markets was actually smaller (when the netting of cash flows is considered), these are still huge figures when compared with world gross domestic product (GDP), which was of the order of $60 trillion at that time.

The consensus was that all this activity was a good thing. Consider the following remarks made by the then chairman of the Federal Reserve, Alan Greenspan, before the Council on Foreign Relations in Washington DC on 19 November 2002 (Greenspan 2002):

> More recently, instruments ... such as credit default swaps, collateralized debt obligations and credit-linked notes have been developed and their use has grown rapidly in recent years. The result? Improved credit risk management together with more and better risk-management tools appear to have significantly reduced loan concentrations in telecommunications and, indeed, other areas and the associated stress on banks and other financial institutions.... It is noteworthy that payouts in the still relatively small but rapidly growing market in credit derivatives have been proceeding smoothly for the most part. Obviously this market is still too new to have been tested in a widespread down-cycle for credit, but, to date, it appears to have functioned well.

As late as April 2006 the International Monetary Fund (IMF) wrote in its Global Financial Stability Report that:

> There is a growing recognition that the dispersion of credit risk by banks to a broader and more diverse group of investors, rather than warehousing such risks on their balance sheets, has helped to make the banking and overall financial system more resilient.... The improved

> resilience may be seen in fewer bank failures and more consistent credit provision. Consequently, the commercial banks, a core system of the financial system, may be less vulnerable today to credit or economic shocks.

It has to be said that the same IMF report also warned about possible vulnerabilities, and the potential for market disruption, if these credit instruments were not fully understood.

One of the problems was that not all of the risk from CDOs was being dispersed to outside investors as the IMF envisaged. As reported in Acharya et al. (2009), large banks were holding on to a lot of it themselves:

> These large, complex financial institutions ignored their own business models of securitization and chose not to transfer credit risk to other investors. Instead they employed securitization to manufacture and retain tail risk that was systemic in nature and inadequately capitalized.... Starting in 2006, the CDO group at UBS noticed that their risk-management systems treated AAA securities as essentially risk-free even though they yielded a premium (the proverbial free lunch). So they decided to hold onto them rather than sell them! After holding less than $5 billion of them in 02/06, the CDO desk was warehousing a staggering $50 billion in 09/07.... Similarly, by late summer of 2007, Citigroup had accumulated over $55 billion of AAA-rated CDOs.

On the eve of the crisis many in the financial industry seemed unconcerned. AIG, the US insurance giant, had become heavily involved in underwriting MBS and CDO risk by selling CDS protection through its AIG Financial Products arm. In August 2007 the chief executive officer of AIG Financial Products is quoted as saying:

> It is hard for us, without being flippant, to even see a scenario within any kind of realm of reason that would see us losing one dollar in any of these transactions.

The financial crisis of 2007–9. After a peak in early 2006, US house prices began to decline in 2006 and 2007. Subprime mortgage holders, experiencing difficulties in refinancing their loans at higher interest rates, defaulted on their payments in increasing numbers. Starting in late 2007 this led to a rapid reassessment of the riskiness of securitizations and to losses in the value of CDO securities. Banks were forced into a series of dramatic *write-downs* of the value of these assets on their balance sheets, and the severity of the impending crisis became apparent.

Reflecting on the crisis in his article "It doesn't take Nostradamus" in the 2008 issue of *Economists' Voice*, Nobel laureate Joseph E. Stiglitz recalled the views he expressed in 1992 on securitization and the housing market:

> The question is, has the growth of securitization been a result of more efficient transaction technologies or an unfounded reduction in concern about the importance of screening loan applicants? It is perhaps too

> early to tell, but we should at least entertain the possibility that it is the latter rather than the former.

He also wrote:

> At the very least, the banks have demonstrated ignorance of two very basic aspects of risk: (a) the importance of correlation ... [and] (b) the possibility of price declines.

These "basic aspects of risk", which would appear to belong in a Banking 101 class, plunged the world's economy into its most serious crisis since the late 1920s. Salient events included the demise of such illustrious names as Bear Stearns (which collapsed and was sold to JPMorgan Chase in March 2008) and Lehman Brothers (which filed for Chapter 11 bankruptcy on 15 September 2008). The latter event in particular led to worldwide panic. As markets tumbled and liquidity vanished it was clear that many banks were on the point of collapse. Governments had to bail them out by injecting capital or by acquiring their distressed assets in arrangements such as the US Troubled Asset Relief Program.

AIG, which had effectively been insuring the default risk in securitized products by selling CDS protection, got into difficulty when many of the underlying securities defaulted; the company that could not foresee itself "losing one dollar in any of these transactions" required an emergency loan facility of $85 billion from the Federal Reserve Bank of New York on 16 September 2008. In the view of George Soros (2009), CDSs were "instruments of destruction" that should be outlawed:

> Some derivatives ought not to be allowed to be traded at all. I have in mind credit default swaps. The more I've heard about them, the more I've realised they're truly toxic.

Much has been written about these events, and this chapter's Notes and Comments section contains a number of references. One strand of the commentary that is relevant for this book is the apportioning of a part of the blame to mathematicians (or financial engineers); the failure of valuation models for complex securitized products made them an easy target. Perhaps the most publicized attack came in a blog by Felix Salmon (*Wired Magazine*, 23 February 2009) under the telling title "Recipe for disaster: the formula that killed Wall Street". The formula in question was the *Gauss copula*, and its application to credit risk was attributed to David Li. Inspired by what he had learned on an actuarial degree, Li proposed that a tool for modelling dependent lifetimes in life insurance could be used to model correlated default times in bond portfolios, thus providing a framework for the valuation and risk management of CDOs, as we describe in Chapter 12.

While an obscure formula with a strange name was a gift for bloggers and newspaper headline writers, even serious regulators joined in the chorus of criticism of mathematics. The Turner Review of the global banking crisis (Lord Turner 2009) has a section entitled "Misplaced reliance on sophisticated mathematics" (see Section 1.3.3 for more on this theme). But this reliance on mathematics was only one factor in the crisis, and certainly not the most important. Mathematicians had also

warned well beforehand that the world of securitization was being built on shaky model foundations that were difficult to calibrate (see, for example, Frey, McNeil and Nyfeler 2001). It was also abundantly clear that political shortsightedness, the greed of market participants and the slow reaction of regulators had all contributed in very large measure to the scale of the eventual calamity.

Recent developments and concerns. New threats to the financial system emerge all the time. The financial crisis of 2007–9 led to recession and sovereign debt crises. After the wave of bank bailouts, concerns about the solvency of banks were transformed into concerns about the abilities of countries to service their own debts. For a while doubts were cast on the viability of the eurozone, as it seemed that countries might elect to, or be forced to, exit the single currency.

On the more technical side, the world of high-frequency trading has raised concerns among regulators, triggered by such events as the Flash Crash of 6 May 2010. In this episode, due to "computer trading gone wild", the Dow Jones lost around 1000 points in a couple of minutes, only to be rapidly corrected. High-frequency trading is a form of algorithmic trading in which trades are executed by computers according to algorithms in fractions of a second. One notable casualty of algorithmic trading was Knight Capital, which lost $460 million due to trading errors on 1 August 2012. Going forward, it is clear that vigilance is required concerning the risks arising from the deployment of new technologies and their systemic implications.

Indeed, *systemic risk* is an ongoing concern to which we have been sensitized by the financial crisis. This is the risk of the collapse of the entire financial system due to the propagation of financial stress through a network of participants. When Lehman Brothers failed there was a moment when it seemed possible that there could be a catastrophic cascade of defaults of banks and other firms. The interbank lending market had become dysfunctional, asset prices had plummeted and the market for any form of debt was highly illiquid. Moreover, the complex chains of relationships in the CDS markets, in which the same credit-risky assets were referenced in a large volume of bilateral payment agreements, led to the fear that the default of a further large player could cause other banks to topple like dominoes.

The concerted efforts of many governments were successful in forestalling the Armageddon scenario. However, since the crisis, research into financial networks and their embedded systemic risks has been an important research topic. These networks are complex, and as well as banks and insurance companies they contain members of a "shadow banking system" of hedge funds and structured investment vehicles, which are largely unregulated. One important theme is the identification of so-called systemically important financial institutions (SIFI) whose failure might cause a systemic crisis.

1.2.2 The Road to Regulation

There is no doubt that regulation goes back a long way, at least to the time of the Venetian banks and the early insurance enterprises sprouting in London's coffee shops in the eighteenth century. In those days there was more reliance on self-regulation or local regulation, but rules were there. However, the key developments

that led to the present prudential regulatory framework in financial services are a very much more recent story.

The main aim of modern prudential regulation has been to ensure that financial institutions have enough capital to withstand financial shocks and remain solvent. Robert Jenkins, a member of the Financial Policy Committee of the Bank of England, was quoted in the *Independent* on 27 April 2012 as saying:

> Capital is there to absorb losses from risks we understand and risks we may not understand. Evidence suggests that neither risk-takers nor their regulators fully understand the risks that banks sometimes take. That's why banks need an appropriate level of loss absorbing equity.

Much of the regulatory drive originated from the Basel Committee of Banking Supervision. This committee was established by the central-bank governors of the Group of Ten at the end of 1974. The Group of Ten is made up of (oddly) eleven industrial countries that consult and cooperate on economic, monetary and financial matters. The Basel Committee does not possess any formal supranational supervising authority, and hence its conclusions do not have legal force. Rather, it formulates broad supervisory standards and guidelines and recommends statements of best practice in the expectation that individual authorities will take steps to implement them through detailed arrangements—statutory or otherwise—that are best suited to their own national system. The summary below is brief. Interested readers can consult, for example, Tarullo (2008) for further details, and should also see this chapter's Notes and Comments section.

The first Basel Accord. The first Basel Accord on Banking Supervision (Basel I, from 1988) took an important step towards an international minimum capital standard. Its main emphasis was on credit risk, by then clearly the most important source of risk in the banking industry. In hindsight, however, Basel I took an approach that was fairly coarse and measured risk in an insufficiently differentiated way. In measuring credit risk, claims were divided into three crude categories according to whether the counterparties were governments, regulated banks or others. For instance, the risk capital charge for a loan to a corporate borrower was five times higher than for a loan to an Organisation for Economic Co-operation and Development (OECD) bank. The risk weighting for all corporate borrowers was identical, independent of their credit rating. The treatment of derivatives was also considered unsatisfactory.

The birth of VaR. In 1993 the G-30 (an influential international body consisting of senior representatives of the private and public sectors and academia) published a seminal report addressing, for the first time, so-called off-balance-sheet products, like derivatives, in a systematic way. Around the same time, the banking industry clearly saw the need for proper measurement of the risks stemming from these new products. At JPMorgan, for instance, the famous Weatherstone 4.15 report asked for a one-day, one-page summary of the bank's market risk to be delivered to the chief executive officer in the late afternoon (hence "4.15"). Value-at-risk (VaR) as a market risk measure was born and the JPMorgan methodology, which became known as RiskMetrics, set an industry-wide standard.

In a highly dynamic world with round-the-clock market activity, the need for instant market valuation of trading positions (known as *marking-to-market*) became a necessity. Moreover, in markets where so many positions (both long and short) were written on the same underlyings, managing risks based on simple aggregation of nominal positions became unsatisfactory. Banks pushed to be allowed to consider *netting* effects, i.e. the compensation of long versus short positions on the same underlying.

In 1996 an important amendment to Basel I prescribed a so-called *standardized* model for market risk, but at the same time allowed the bigger (more sophisticated) banks to opt for an *internal* VaR-based model (i.e. a model developed in house). Legal implementation was to be achieved by the year 2000. The coarseness problem for credit risk remained unresolved and banks continued to claim that they were not given enough incentives to diversify credit portfolios and that the regulatory capital rules currently in place were far too risk insensitive. Because of overcharging on the regulatory capital side of certain credit positions, banks started shifting business away from certain market segments that they perceived as offering a less attractive risk–return profile.

The second Basel Accord. By 2001 a consultative process for a new Basel Accord (Basel II) had been initiated; the basic document was published in June 2004. An important aspect was the establishment of the three-pillar system of regulation: Pillar 1 concerns the quantification of regulatory capital; Pillar 2 imposes regulatory oversight of the modelling process, including risks not considered in Pillar 1; and Pillar 3 defines a comprehensive set of disclosure requirements.

Under Pillar 1 the main theme of Basel II was credit risk, where the aim was to allow banks to use a finer, more risk-sensitive approach to assessing the risk of their credit portfolios. Banks could opt for an *internal-ratings-based* approach, which permitted the use of internal or external credit-rating systems wherever appropriate.

The second important theme of Basel II at the level of Pillar 1 was the consideration of operational risk as a new risk class. A basic premise of Basel II was that the overall size of regulatory capital throughout the industry should stay unchanged under the new rules. Since the new rules for credit risk were likely to reduce the credit risk charge, this opened the door for operational risk, defined as the risk of losses resulting from inadequate or failed internal processes, people and systems or from external events; this definition included legal risk but excluded reputational and strategic risk.

Mainly due to the financial crisis of 2007–9, implementation of the Basel II guidelines across the globe met with delays and was rather spread out in time. Various further amendments and additions to the content of the original 2004 document were made. One important criticism of Basel II that emerged from the crisis was that it was inherently *procyclical*, in that it forced firms to take action to increase their capital ratios at exactly the wrong point in the business cycle, when their actions had a negative impact on the availability of liquidity and made the situation worse (see Section 1.3.3 for more discussion on this).

Basel 2.5. One clear lesson from the crisis was that modern products like CDOs had opened up opportunities for regulatory arbitrage by transferring credit risk from the capital-intensive banking book (or loan book) to the less-capitalized trading book. Some enhancements to Basel II were proposed in 2009 with the aim of addressing the build-up of risk in the trading book that was evident during the crisis. These enhancements, which have come to be known as Basel 2.5, include a *stressed VaR* charge, based on calculating VaR from data for a twelve-month period of market turmoil, and the so-called *incremental risk charge*, which seeks to capture some of the default risk in trading book positions; there were also specific new rules for certain securitizations.

The third Basel Accord. In view of the failure of the Basel rules to prevent the 2007–9 crisis, the recognized deficiencies of Basel II mentioned above, and the clamour from the public and from politicians for regulatory action to make banks and the banking system safer, it is no surprise that attention quickly shifted to Basel III.

In 2011 a series of measures was proposed that would extend Basel II (and 2.5) in five main areas:

(1) measures to increase the quality and amount of bank capital by changing the definition of key capital ratios and allowing countercyclical adjustments to these ratios in crises;

(2) a strengthening of the framework for counterparty credit risk in derivatives trading, with incentives to use central counterparties (exchanges);

(3) the introduction of a leverage ratio to prevent excessive leverage;

(4) the introduction of various ratios that ensure that banks have sufficient funding liquidity;

(5) measures to force systemically important banks to have even higher capacity to absorb losses.

Most of the new rules will be phased in progressively, with a target end date of 2019, although individual countries may impose stricter guidelines with respect to both schedule and content.

Parallel developments in insurance regulation. The insurance industry worldwide has also been subject to increasing risk regulation in recent times. However, here the story is more fragmented and there has been much less international coordination of efforts. The major exception has been the development of the Solvency II framework in the European Union, a process described in more detail below. As the most detailed and model intensive of the regulatory frameworks proposed, it serves as our main reference point for insurance regulation in this book. The development of the Solvency II framework is overseen by the European Insurance and Occupational Pensions Authority (EIOPA; formerly the Committee of European Insurance and Occupational Pensions Supervisors (CEIOPS)), but the implementation in individual

countries is a matter for national regulators, e.g. the Prudential Regulatory Authority in the UK.

In the US, insurance regulation has traditionally been a matter for state governments. The National Association of Insurance Commissioners (NAIC) provides support to insurance regulators from the individual states, and helps to promote the development of accepted regulatory standards and best practices; it is up to the individual states whether these are passed into law, and if so in what form. In the early 1990s the NAIC promoted the concept of risk-based capital for insurance companies as a response to a number of insolvencies in the preceding years; the NAIC describes risk-based capital as "a method of measuring the minimum amount of capital appropriate for a reporting entity to support its overall business operations in consideration of its size and profile". The method, which is a rules-based approach rather than a model-based approach, has become the main plank of insurance regulation in the US.

Federal encroachment on insurance supervision has generally been resisted, although this may change due to a number of measures enacted after the 2007–9 crisis in the wide-ranging 2010 Dodd–Frank Act. These include the creation of both the Federal Insurance Office, to "monitor all aspects of the insurance sector", and the Financial Stability Oversight Council, which is "charged with identifying risks to the financial stability of the United States" wherever they may arise in the world of financial services.

The International Association of Insurance Supervisors has been working to foster some degree of international convergence in the processes for regulating the capital adequacy of insurers. They have promoted the idea of the Own Risk and Solvency Assessment (ORSA). This has been incorporated into the Solvency II framework and has also been embraced by the NAIC in the US.

There are also ongoing initiatives that aim to bring about convergence of banking and insurance regulation, particularly with respect to financial conglomerates engaged in both banking and insurance business. The Joint Forum on Financial Conglomerates was established in early 1996 under the aegis of the Basel Committee, the International Association of Insurance Supervisors and the International Organization of Securities Commissions to take forward this work.

From Solvency I to Solvency II. Mirroring the progress in the banking sector, Solvency II is the latest stage in a process of regulatory evolution from simple and crude rules to a more risk-sensitive treatment of the capital requirements of insurance companies.

The first European Union non-life and life directives on solvency margins appeared around 1970. The solvency margin was defined as an extra capital buffer against unforeseen events such as higher than expected claims levels or unfavourable investment results. However, there were differences in the way that regulation was applied across Europe and there was a desire for more harmonization of regulation and mutual recognition.

Solvency I, which came into force in 2004, is a rather coarse rules-based framework calling for companies to have a minimum guarantee fund (minimal capital)

of €3 million, and a solvency margin consisting of 16–18% of non-life premiums together with 4% of the technical provisions for life. This has led to a single robust system that is easy to understand and inexpensive to monitor. However, on the negative side, it is mainly volume based, not explicitly risk based; issues like guarantees, embedded options and the proper matching of assets and liabilities are largely neglected in many countries.

To address these shortcomings, Solvency II was initiated in 2001 with the publication of the influential Sharma Report. While the Solvency II directive was adopted by the Council of the European Union and the European Parliament in November 2009, implementation of the framework is not expected until 1 January 2016. The process of refinement of the framework is managed by EIOPA, and one of the features of this process has been a series of quantitative impact studies in which companies have effectively tried out aspects of the proposals and information has been gathered with respect to the impact and practicability of the new regulations.

The goal of the Solvency II process is that the new framework should strengthen the capital adequacy regime by reducing the possibilities of consumer loss or market disruption in insurance; Solvency II therefore has both policyholder-protection and financial-stability motives. Moreover, it is also an aim that the harmonization of regulation in Europe should promote deeper integration of the European Union insurance market and the increased competitiveness of European insurers. A high-level description of the Solvency II framework is given in Section 1.3.2.

The Swiss Solvency Test (SST). Special mention should be made of Switzerland, which has already developed and implemented its own principles-based risk capital regulation for the insurance industry. The SST has been in force since 1 January 2011. It follows similar principles to Solvency II but differs in some details of its treatment of different types of risk; it also places more emphasis on the development of internal models. The implementation of the SST falls under the remit of the Swiss Financial Markets Supervisory Authority, a body formed in 2007 from the merger of the banking and insurance supervisors, which has statutory authority over banks, insurers, stock exchanges, collective investment schemes and other entities.

1.3 The Regulatory Framework

This section describes in more detail the framework that has emerged from the Basel process and the European Union solvency process.

1.3.1 The Basel Framework

As indicated in Section 1.2.2, the Basel framework should be regarded as the product of an evolutionary process. As this book goes to press, the Basel II and Basel 2.5 proposals have been implemented in many developed countries (with some variations in detail), while the proposals of Basel III are still being debated and refined. We sketch the framework as currently implemented, before indicating some of the proposed changes and additions to the framework in Basel III.

The three-pillar concept. A key feature of the Basel framework is the three-pillar concept, as is apparent from the following statement summarizing the Basel philosophy, which accompanied the original Basel II publication (Basel Committee on Banking Supervision 2004):

> The Basel II Framework sets out the details for adopting more risk-sensitive minimum capital requirements [Pillar 1] for banking organizations. The new framework reinforces these risk-sensitive requirements by laying out principles for banks to assess the adequacy of their capital and for supervisors to review such assessments to ensure banks have adequate capital to support their risks [Pillar 2]. It also seeks to strengthen market discipline by enhancing transparency in banks' financial reporting [Pillar 3]. The text that has been released today reflects the results of extensive consultations with supervisors and bankers worldwide. It will serve as the basis for national rule-making and approval processes to continue and for banking organizations to complete their preparations for the new Framework's implementation.

Under *Pillar 1*, banks are required to calculate a *minimum capital charge*, referred to as regulatory capital. There are separate Pillar 1 capital charges for credit risk in the banking book, market risk in the trading book and operational risk, which are considered to be the main quantifiable risks. Most banks use internal models based on VaR methodology to compute the capital charge for market risk. For credit risk and operational risk banks may choose between several approaches of increasing risk sensitivity and complexity, some details of which are discussed below.

Pillar 2 recognizes that any quantitative approach to risk management should be embedded in a properly functioning corporate governance structure. Best-practice risk management imposes constraints on the organization of the institution, i.e. the board of directors, management, employees, and internal and external audit processes. In particular, the board of directors assumes the ultimate responsibility for oversight of the risk landscape and the formulation of the company's risk appetite. Through Pillar 2, also referred to as the *supervisory review process*, local regulators review the various checks and balances that have been put in place. Under Pillar 2, residual quantifiable risks that are not included in Pillar 1, such as interest-rate risk in the banking book, must be considered and *stress tests* of a bank's capital adequacy must be performed. The aim is to ensure that the bank holds capital in line with its true economic loss potential, a concept known as *economic capital*.

Finally, in order to fulfil its promise that increased regulation will increase transparency and diminish systemic risk, clear reporting guidelines on the risks carried by financial institutions are called for. *Pillar 3* seeks to establish *market discipline* through a better public disclosure of risk measures and other information relevant to risk management. In particular, banks will have to offer greater insight into the adequacy of their capitalization.

Credit and market risk; the banking and trading books. Historically, banking activities have been organized around the banking book and the trading book, a split that

reflects different accounting practices for different kinds of assets. The banking book contains assets that are *held to maturity*, such as loans; these are typically valued at book value, based on the original cost of the asset. The trading book contains assets and instruments that are *available to trade*; these are generally valued by *marking-to-market* (i.e. using quoted market prices). From a regulatory point of view, credit risk is mainly identified with the banking book and market risk is mainly identified with the trading book.

We have already noted that there are problems with this simple dichotomy and that the Basel 2.5 rules were introduced (partly) to account for the neglect of credit risk (default and rating-migration risk) in the trading book. There are also forms of market risk in the banking book, such as interest-rate risk and foreign-exchange risk. However, the Basel framework continues to observe the distinction between banking book and trading book and we will describe the capital charges in terms of the two books. It is clear that the distinction is somewhat arbitrary and rests on the concept of "available to trade". Moreover, there can be incentives to "switch" or move instruments from one book to the other (particularly from the banking book to the trading book) to benefit from a more favourable capital treatment. This is acknowledged by the Basel Committee in its background discussion of the "Fundamental review of the trading book: a revised market risk framework" (Basel Committee on Banking Supervision 2013a):

> The Committee believes that the definition of the regulatory boundary between the trading book and the banking book has been a source of weakness in the design of the current regime. A key determinant of the boundary has been banks' self-determined intent to trade.... Coupled with large differences in capital requirements against similar types of risk on either side of the boundary, the overall capital framework proved susceptible to arbitrage before and during the crisis.... To reduce the incentives for arbitrage, the Committee is seeking a less permeable boundary with strict limits on switching between books and measures to prevent "capital benefit" in instances where switching is permitted.

The capital charge for the banking book. The credit risk of the banking book portfolio is assessed as the sum of *risk-weighted assets*: that is, the sum of notional exposures weighted by a coefficient reflecting the creditworthiness of the counterparty (the risk weight). To calculate risk weights, banks use either the *standardized* approach or one of the more advanced *internal-ratings-based* (IRB) approaches. The choice of method depends on the size and complexity of the bank, with the larger, international banks having to go for IRB approaches. The capital charge is determined as a fraction of the sum of risk-weighted assets in the portfolio. This fraction, known as the capital ratio, was 8% under Basel II but is already being increased ahead of the planned implementation of Basel III in 2019.

The standardized approach refers to a system that has been in place since Basel I, whereby the risk weights are prescribed by the regulator according to the nature and creditworthiness of the counterparty. For example, there are risk weights for

retail loans secured on property (mortgages) and for unsecured retail loans (such as credit cards and overdrafts); there are also different risk weights for corporate and government bonds with different ratings.

Under the more advanced IRB approaches, banks may dispense with the system of fixed risk weights provided by the regulator. Instead, they may make an *internal* assessment of the riskiness of a credit exposure, expressing this in terms of an estimated annualized *probability of default* and an estimated *loss given default*, which are used as inputs in the calculation of risk-weighted assets. The total sum of risk-weighted assets is calculated using formulas specified by the Basel Committee; the formulas also take into account the fact that there is likely to be positive correlation (sometimes called systematic risk) between the credit risks in the portfolio. The use of internally estimated probabilities of default and losses given default allows for increased risk sensitivity in the IRB capital charges compared with the standardized approach. It should be noted, however, that the IRB approaches do not permit fully internal models of credit risk in the banking book; they only permit internal estimation of inputs to a model that has been specified by the regulator.

The capital charge for the trading book. For market risk in the trading book there is also the option of a standardized approach based on a system of risk weights and specific capital charges for different kinds of instrument. However, most major banks elect to use an *internal VaR model approach*, as permitted by the 1996 amendment to Basel I. In Sections 2.2 and 9.2 of this book we give a detailed description of the VaR approach to trading book risk measurement. The approach is based on the estimation of a P&L distribution for a ten-day holding period and the estimation of a particular percentile of this distribution: the 99th percentile of the losses.

A ten-day VaR at 99% of $20 million therefore means that it is estimated that our market portfolio will incur a loss of $20 million *or more* with probability 1% by the end of a ten-day holding period, if the composition remains fixed over this period. The conversion of VaR numbers into an actual capital charge is accomplished by a formula that we discuss in Section 2.3.3.

The VaR calculation is the main component of risk quantification for the trading book, but the 2009 Basel 2.5 revision added further elements (see Basel Committee on Banking Supervision 2012, p. 10), including the following.

Stressed VaR: banks are required to carry out a VaR calculation essentially using the standard VaR methodology but calibrating their models to a historical twelve-month period of significant financial stress.

Incremental risk charge: Since default and rating-migration risk are not generally considered in the standard VaR calculation, banks must calculate an additional charge based on an estimate of the 99.9th percentile of the one-year distribution of losses due to defaults and rating changes. In making this calculation they may use internal models for credit risk (in contrast to the banking book) but must also take into account the market liquidity of credit-risky instruments.

Securitizations: exposures to securitizations in the trading book are subject to a series of new capital charges that bring them more into line with equivalent exposures in the banking book.

The capital charge for operational risk. There are also options of increasing sophistication for assessing operational risk. Under the *basic-indicator* and *standardized* approaches, banks may calculate their operational risk charge using simple formulas based on gross annual income. Under the *advanced measurement approach*, banks may develop internal models. Basel is not prescriptive about the form of these models provided they capture the tail risk of extreme events; most such models are based on historical loss data (internal and external to the firm) and use techniques that are drawn from the actuarial modelling of general insurance losses. We provide more detail in Chapter 13.

New elements of Basel III. Under Basel III there will be a number of significant changes and additions to the Basel framework. While the detail of the new rules may change before final implementation in 2019, the main developments are now clear.

- Banks will need to hold both *more capital* and *better-quality capital* as a function of the risks taken. The "better quality" is achieved though a more restrictive definition of eligible capital (through more stringent definitions of Tier 1 and Tier 2 capital and the phasing out of Tier 3 capital); see Section 2.1.3 for more explanation of capital tiers. The "more" comes from the addition (on top of the minimum ratio of 8%) of a capital conservation buffer of 2.5% of risk-weighted assets, for building up capital in good times to absorb losses under stress, and a countercyclical buffer within the range 0–2.5%, in order to enhance the shock resilience of banks and limit expansion in periods of excessive credit growth. This leads to a total (Tier 1 plus Tier 2) ratio of up to 13%, compared with Basel II's 8%. There will be a gradual phasing in of all these new ratios, with a target date for full implementation of 1 January 2019.

- A *leverage ratio* will be imposed to put a floor under the build-up of excessive leverage in the banking system. Leverage will essentially be measured through the ratio of Tier 1 capital to total assets. A minimum ratio of 3% is currently being tested but the precise definitions may well change as a result of testing experience and bank lobbying. The leverage limit will restrain the size of bank assets, regardless of their riskiness.

- The risk coverage of the system of capital charges is being extended, in particular to include a charge for *counterparty credit risk*. When counterparty credit risk is taken into account in the valuation of over-the-counter derivatives contract, the default-risk-free value has to be adjusted by an amount known as the credit value adjustment (CVA); see Section 17.2 for more explanation. There will now be a charge for changes in CVA.

- Banks will become subject to *liquidity rules*; this is a completely new direction for the Basel framework, which has previously been concerned only with capital adequacy. A *liquidity coverage ratio* will be introduced to ensure that banks have enough highly liquid assets to withstand a period of net cash outflow lasting thirty days. A *net stable funding ratio* will ensure that sufficient funding is available in order to cover long-term commitments (exceeding one year).

It should also be mentioned that under an ongoing review of the trading book, the principle of risk quantification may change from one based on VaR (a percentile) to one based on *expected shortfall* (ES). For a given holding period, the ES at the 99% level, say, is the expected loss given that the loss is higher than the VaR at the 99% level over the same period. ES is a severity measure that always dominates the frequency measure VaR and gives information about the expected size of tail losses; it is also a measure with superior aggregation properties to VaR, as discussed in Section 2.3.5 and Chapter 8 (particularly Sections 8.1 and 8.4.4).

1.3.2 The Solvency II Framework

Below we give an outline of the Solvency II framework, which will come into force in the countries of the European Union on or before 1 January 2016.

Main features. In common with the Basel Accords, Solvency II adopts a three-pillar system, where the first pillar requires the quantification of regulatory capital requirements, the second pillar is concerned with governance and supervision, and the third pillar requires the disclosure of information to the public to improve market discipline by making it easier to compare the risk profiles of companies.

Under Pillar 1, a company calculates its *solvency capital requirement*, which is the amount of capital it should have to ensure that the probability of insolvency over a one-year period is no more than 0.5%—this is often referred to as a confidence level of 99.5%. The company also calculates a smaller *minimum capital requirement*, which is the minimum capital it should have to continue operating without supervisory intervention.

To calculate the capital requirements, companies may use either an *internal model* or a simpler *standard formula* approach. In either case the intention is that a *total balance sheet* approach is taken in which all risks and their interactions are considered. The insurer should have *own funds* (a surplus of assets over liabilities) that exceed both the solvency capital requirement and the minimum capital requirement. The assets and liabilities of the firm should be valued in a *market-consistent* manner.

The supervisory review of the company takes place under Pillar 2. The company must demonstrate that it has a risk-management system in place and that this system is integrated into decision-making processes, including the setting of risk appetite by the company's board, and the formulation of risk limits for different business units. An internal model must pass the "use test": it must be an integral part of the risk-management system and be actively used in the running of the firm. Moreover, a firm must undertake an ORSA as described below.

Market-consistent valuation. In Solvency II the valuation must be carried out according to market-consistent principles. Where possible it should be based on actual *market values*, in a process known as *marking-to-market*. In a Solvency II glossary provided by the Comité Européen des Assurances and the Groupe Consultatif in 2007, market value is defined as:

> The amount for which an asset could be exchanged or a liability settled, between knowledgeable, willing parties in an arm's length transaction, based on observable prices within an active, deep and liquid market which is available to and generally used by the entity.

The concept of market value is related to the concept of *fair value* in accounting, and the principles adopted in Solvency II valuation have been influenced by International Financial Reporting Standards (IFRS) accounting standards. When no relevant market values exist (or when they do not meet the quality criteria described by the concept of an "active, deep and liquid market"), then market-consistent valuation requires the use of models that are calibrated, as far as possible, to be consistent with financial market information, a process known as *marking-to-model*; we discuss these ideas in more detail in Section 2.2.2.

The market-consistent valuation of the liabilities of an insurer is possible when the cash flows paid to policyholders can be fully replicated by the cash flows generated by the so-called matching assets that are held for that purpose; the value of the liability is then given by the value of the *replicating portfolio* of matching assets. However, it is seldom the case that liabilities can be fully replicated and hedged; mortality risk is a good example of a risk factor that is difficult to hedge.

The valuation of the unhedgeable part of a firm's liabilities is carried out by computing the sum of a *best estimate* of these liabilities (basically an expected value) plus an extra *risk margin* to cover some of the uncertainty in the value of the liability. The idea of the risk margin is that a third party would not be willing to take over the unhedgeable liability for a price set at the best estimate but would have to be further compensated for absorbing the additional uncertainty about the true value of the liability.

Standard formula approach. Under this approach an insurer calculates capital charges for different kinds of risk within a series of *modules*. There are modules, for example, for market risk, counterparty default risk, life underwriting risk, non-life underwriting risk and health insurance risk. The risk charges arising from these modules are aggregated to obtain the solvency capital requirement using a formula that involves a set of prescribed correlation parameters (see Section 8.4.2).

Within each module, the approach drills down to fundamental risk factors; for example, within the market-risk module, there are sub-modules relating to interest-rate risk, equity risk, credit-spread risk and other typical market-risk factors. Capital charges are calculated with respect to each risk factor by considering the effect of a series of defined stress scenarios on the value of net assets (assets minus liabilities). The stress scenarios are intended to represent 1-in-200-year events (i.e. events with an annual probability of 0.5%).

The capital charges for each risk factor are aggregated to obtain the module risk charge using a similar kind of formula to the one used at the highest level. Once again, a set of correlations expresses the regulatory view of dependencies between the effects of the fundamental risk factors. The details are complex and run to many pages, but the approach is simple and highly prescriptive.

Internal-model approach. Under this approach firms can develop an internal model for the financial and underwriting risk factors that affect their business; they may then seek regulatory approval to use this model in place of the standard formula. The model often takes the form of a so-called *economic scenario generator* in which risk-factor scenarios for a one-year period are randomly generated and applied to the assets and liabilities to determine the solvency capital requirement. Economic scenario generators vary greatly in their detail, ranging from simple distributional models to more sophisticated dynamic models in discrete or continuous time.

ORSA. In a 2008 Issues Paper produced by CEIOPS, the ORSA is described as follows:

> The entirety of the processes and procedures employed to identify, assess, monitor, manage, and report the short and long term risks a (re)insurance undertaking faces or may face and to determine the own funds necessary to ensure that the undertaking's overall solvency needs are met at all times.

The concept of an ORSA is not unique to Solvency II and a useful alternative definition has been provided by the NAIC in the US on its website:

> In essence, an ORSA is an internal process undertaken by an insurer or insurance group to assess the adequacy of its risk management and current and prospective solvency positions under normal and severe stress scenarios. An ORSA will require insurers to analyze all reasonably foreseeable and relevant material risks (i.e., underwriting, credit, market, operational, liquidity risks, etc.) that could have an impact on an insurer's ability to meet its policyholder obligations.

The Pillar 2 ORSA is distinguished from the Pillar 1 capital calculations in a number of ways. First, the definition makes clear that the ORSA refers to a process, or set of processes, and not simply an exercise in regulatory compliance. Second, each firm's ORSA is its *own* process and is likely to be *unique*, since it is not bound by a common set of rules. In contrast, the standard-formula approach to Pillar 1 is clearly a uniform process for all companies; moreover, firms that seek internal-model approval for Pillar 1 are subject to very similar constraints.

Finally, the ORSA goes beyond the one-year time horizon (which is a limitation of Pillar 1) and forces firms to assess solvency over their business planning horizon, which can mean many years for typical long-term business lines, such as life insurance.

1.3.3 Criticism of Regulatory Frameworks

The benefits arising from the regulation of financial services are not generally in doubt. Customer-protection acts, responsible corporate governance, fair and comparable accounting rules, transparent information on risk, capital and solvency for shareholders and clients are all viewed as positive developments.

Very few would argue the extreme position that the prudential regulatory frameworks we have discussed are not needed; in general, after a crisis, the demand (at least from the public and politicians) is for more regulation. Nevertheless, there are aspects of the regulatory frameworks that have elicited criticism, as we now discuss.

Cost and complexity. The cost factor of setting up a well-functioning risk-management system compliant with the present regulatory framework is significant, especially (in relative terms) for smaller institutions. On 27 March 2013, the *Financial Times* quoted Andrew Bailey (head of the Prudential Regulatory Authority in the UK) as saying that Solvency II compliance was set to cost UK companies at least £3 billion, a "frankly indefensible" amount. Related to the issue of cost is the belief that regulation, in its attempt to become more risk sensitive, is becoming too complex; this theme is taken up by the Basel Committee in their 2013 discussion paper entitled "The regulatory framework: balancing risk sensitivity, simplicity and comparability" (Basel Committee on Banking Supervision 2013b).

Endogenous risk. In general terms, this refers to the risk that is generated within a system and amplified by the system due to feedback effects. Regulation, a feature of the system, may be one of the channels by which shocks are amplified.

Regulation can lead to *risk-management herding*, whereby institutions following similar (perhaps VaR-based) rules may all be "running for the same exit" in times of crisis, consequently destabilizing an already precarious situation even further. This herding phenomenon has been suggested in connection with the 1987 stock market crash and the events surrounding the 1998 LTCM crisis (Daníelsson et al. 2001b).

An even more compelling example was observed during the 2007–9 crisis; to comply with regulatory capital ratios in a market where asset values were falling and risks increasing, firms adjusted their balance sheets by selling assets, causing further asset value falls and vanishing market liquidity. This led to criticism of the inherently *procyclical* nature of the Basel II regulation, whereby capital requirements may rise in times of stress and fall in times of expansion; the Basel III proposals attempt to address this issue with a countercyclical capital buffer.

Consequences of fair-value accounting and market-consistent valuation. The issue of procyclicality is also related to the widespread use of fair-value accounting and market-consistent valuation, which are at the heart of both the Basel rules for the trading book and the Solvency II framework. The fact that capital requirements are so closely coupled to volatile financial markets has been another focus of criticism.

An example of this is the debate around the valuation of insurance liabilities in periods of market stress. A credit crisis, of the kind experienced in 2007–9, can impact the high-quality corporate bonds that insurance companies hold on the asset

side of their balance sheets. The relative value of corporate bonds compared with safe government bonds can fall sharply as investors demand more compensation for taking on both the credit risk and, in particular, the liquidity risk of corporate bonds.

The effect for insurers is that the value of their assets falls relative to the value of their liabilities, since the latter are valued by comparing cash flows with safe government bonds. At a particular point in time, an insurer may appear to have insufficient capital to meet solvency capital requirements. However, if an insurer has matched its asset and liability cash flows and can continue to meet its contractual obligations to policyholders, the apparent depletion of capital may not be a problem; insurance is a long-term business and the insurer has no short-term need to sell assets or offload liabilities, so a loss of capital need not be realized unless some of the bonds actually default.

Regulation that paints an unflattering picture of an insurer's solvency position is not popular with regulated firms. Firms have argued that they should be able to value liabilities at a lower level, by comparing the cash flows not with expensive government bonds but instead with the corporate bonds that are actually used as matching assets, making allowance only for the credit risk in corporate bonds. This has given rise to the idea of discounting with an extra *illiquidity premium*, or *matching premium*, above a risk-free rate. There has been much debate about this issue between those who feel that such proposals undermine market-consistent valuation and those who believe that strict adherence to market-consistent valuation overstates risk and has potential systemic consequences (see, for example, Wüthrich 2011).

Limits to quantification. Further criticism has been levelled at the highly quantitative nature of regulation and the extensive use of mathematical and statistical methods. The section on "Misplaced reliance on sophisticated mathematics" in the Turner Review of the global banking crisis (Lord Turner 2009) states that:

> The very complexity of the mathematics used to measure and manage risk, moreover, made it increasingly difficult for top management and boards to assess and exercise judgement over the risk being taken. Mathematical sophistication ended up not containing risk, but providing false assurances that other prima facie indicators of increasing risk (e.g. rapid credit extension and balance sheet growth) could be safely ignored.

This idea that regulation can lead to overconfidence in the quality of statistical risk measures is related to the view that the essentially backward-looking nature of estimates derived from historical data is a weakness. The use of conventional VaR-based methods has been likened to driving a car while looking in the rear-view mirror, the idea being that this is of limited use in preparing for the shocks that lie ahead.

The extension of the quantitative approach to operational risk has been controversial. Whereas everyone agrees that risks such as people risk (e.g. incompetence,

fraud), process risk (e.g. model, transaction and operational control risk), technology risk (e.g. system failure, programming error) and legal risk are important, there is much disagreement on the extent to which these risks can be measured.

Limits to the efficacy of regulation. Finally, there is some debate about whether or not tighter regulation can ever prevent the occurrence of crises like that of 2007–9. The sceptical views of central bankers and regulatory figures were reported in the *Economist* in an article entitled "The inevitability of instability" (25 January 2014) (see also Prates 2013). The article suggests that "rules are constantly overtaken by financial innovation" and refers to the economist J. K. Galbraith (1993), who wrote:

> All financial innovation involves, in one form or another, the creation of debt secured in greater or lesser adequacy by real assets.... All crises have involved debt that, in one fashion or another, has become dangerously out of scale in relation to the underlying means of payment.

Tightening up the capital treatment of securitizations may prevent a recurrence of the events surrounding the 2007–9 crisis, but, according to the sceptical view, it will not prevent different forms of debt-fuelled crisis in the future.

1.4 Why Manage Financial Risk?

An important issue that we have barely touched upon is the reason for investing in risk management in the first place. This question can be addressed from various perspectives, including those of the customer of a financial institution, its shareholders, its management, its board of directors, regulators, politicians, or the general public; each of these stakeholders may have a different view. In the selective account we give here, we focus on two viewpoints: that of society as a whole, and that of the shareholders (owners) of a firm.

1.4.1 A Societal View

Modern society relies on the smooth functioning of banking and insurance systems, and it has a collective interest in the stability of such systems. The regulatory process that has given us the Basel and Solvency II frameworks was initially motivated by the desire to prevent the insolvency of individual institutions, thus protecting customers and policyholders; this is sometimes referred to as a *microprudential* approach. However, the reduction of systemic risk—the danger that problems in a single financial institution may spill over and, in extreme situations, disrupt the normal functioning of the entire financial system—has become an important secondary focus, particularly since the 2007–9 crisis. Regulation therefore now also takes a *macroprudential* perspective.

Most members of society would probably agree that protection of customers against the failure of an individual firm is an important aim, and there would be widespread agreement that the promotion of financial stability is vital. However, it is not always clear that the two aims are well aligned. While there are clearly situations where the failure of one company may lead to spillover effects that result

in a systemic crisis, there may also be situations where the long-term interests of financial stability are better served by allowing a company to fail: it may provide a lesson in the importance of better risk management for other companies. This issue is clearly related to the *systemic importance* of the company in question: in other words, to its size and the extent of its connectivity to other firms. But the recognition that there may be firms that are too important or are *too big to fail* creates a *moral hazard*, since the management of such a firm may take more risk in the knowledge that the company would be bailed out in a crisis. Of course, it may be the case that in some countries some institutions are also *too big to save*.

The 2007–9 crisis provided a case study that brought many of these issues to the fore. As we noted in our account of the crisis in Section 1.2, it was initially believed that the growth in securitization was dispersing credit risk throughout the system and was beneficial to financial stability. But the warehousing of vast amounts of inadequately capitalized credit risk (in the form of CDOs) in trading books, combined with the interconnectedness of banks through derivatives and interbank lending activities, meant that quite the opposite was true. The extent of the systemic risk that had been accumulating became apparent when Lehman Brothers filed for bankruptcy on 15 September 2008 and governments intervened to bail out the banks.

It was the following phase of the crisis during which society suffered. The world economy went into recession, households defaulted on their debts, and savings and pensions were hit hard. The crisis moved "from Wall Street to Main Street". Naturally, this led to resentment as banking remained a highly rewarded profession and it seemed that the government-sponsored bailouts had allowed banks "to privatize their gains and socialize their losses".

There has been much debate since the crisis on whether the US government could have intervened to save Lehman, as it did for other firms such as AIG. In the *Financial Times* on 14 September 2009, the historian Niall Ferguson wrote:

> Like the executed British admiral in Voltaire's famous phrase, Lehman had to die *pour encourager les autres*—to convince the other banks that they needed injections of public capital, and to convince the legislature to approve them. Not everything in history is inevitable; contingencies abound. Sometimes it is therefore right to say "if only". But an imagined rescue of Lehman Brothers is the wrong counterfactual. The right one goes like this. If only Lehman's failure and the passage of TARP had been followed—not immediately, but after six months—by a clear statement to the surviving banks that none of them was henceforth too big to fail, then we might actually have learnt something from this crisis.

While it is difficult to speak with authority for "society", the following conclusions do not seem unreasonable. The interests of society are served by enforcing the discipline of risk management in financial firms, through the use of regulation. Better risk management can reduce the risk of company failure and protect customers and policyholders who stand in a very unequal financial relationship with large firms. However, the regulation employed must be designed with care and should not promote herding, procyclical behaviour or other forms of endogenous risk that could

result in a systemic crisis with far worse implications for society than the failure of a single firm. Individual firms need to be allowed to fail on occasion, provided customers can be shielded from the worst consequences through appropriate compensation schemes. A system that allows firms to become too big to fail creates moral hazard and should be avoided.

1.4.2 The Shareholder's View

It is widely believed that proper financial risk management can increase the value of a corporation and hence shareholder value. In fact, this is the main reason why corporations that are not subject to regulation by financial supervisory authorities engage in risk-management activities. Understanding the relationship between shareholder value and financial risk management also has important implications for the design of risk-management systems. Questions to be answered include the following.

- When does risk management increase the value of a firm, and which risks should be managed?
- How should risk-management concerns factor into investment policy and capital budgeting?

There is a rather extensive corporate-finance literature on the issue of "corporate risk management and shareholder value". We briefly discuss some of the main arguments. In this way we hope to alert the reader to the fact that there is more to risk management than the mainly technical questions related to the implementation of risk-management strategies dealt with in the core of this book.

The first thing to note is that from a corporate-finance perspective it is by no means obvious that in a world with perfect capital markets risk management enhances shareholder value: while *individual* investors are typically risk averse and should therefore manage the risk in their portfolios, it is not clear that risk management or risk reduction at the *corporate level*, such as hedging a foreign-currency exposure or holding a certain amount of risk capital, increases the value of a corporation. The rationale for this (at first surprising) observation is simple: if investors have access to perfect capital markets, they can do the risk-management transactions they deem necessary via their own trading and diversification. The following statement from the chief investment officer of an insurance company exemplifies this line of reasoning: "If our shareholders believe that our investment portfolio is too risky, they should short futures on major stock market indices."

The potential irrelevance of corporate risk management for the value of a corporation is an immediate consequence of the famous *Modigliani–Miller Theorem* (Modigliani and Miller 1958). This result, which marks the beginning of modern corporate-finance theory, states that, in an ideal world without taxes, bankruptcy costs and informational asymmetries, and with frictionless and arbitrage-free capital markets, the financial structure of a firm, and hence also its risk-management decisions, are irrelevant when assessing the firm's value. Hence, in order to find reasons for corporate risk management, one has to "turn the Modigliani–Miller Theorem upside down" and identify situations where risk management enhances

the value of a firm by deviating from the unrealistically strong assumptions of the theorem. This leads to the following rationales for risk management.

- Risk management can reduce *tax costs*. Under a typical tax regime the amount of tax to be paid by a corporation is a *convex* function of its profits; by reducing the variability in a firm's cash flow, risk management can therefore lead to a higher expected after-tax profit.
- Risk management can be beneficial, since a company may (and usually will) have better access to capital markets than individual investors.
- Risk management can increase firm value in the presence of *bankruptcy costs*, as it makes bankruptcy less likely.
- Risk management can reduce the impact of *costly external financing* on the firm value, as it facilitates the achievement of optimal investment.

The last two points merit a more detailed discussion. Bankruptcy costs consist of direct bankruptcy costs, such as the cost of lawsuits, and the more important indirect bankruptcy costs. The latter may include liquidation costs, which can be substantial in the case of intangibles like research and development and knowhow. This is why high research and development spending appears to be positively correlated with the use of risk-management techniques. Moreover, increased likelihood of bankruptcy often has a negative effect on key employees, management and customer relations, in particular in areas where a client wants a long-term business relationship. For instance, few customers would want to enter into a life insurance contract with an insurance company that is known to be close to bankruptcy. On a related note, banks that are close to bankruptcy might be faced with the unpalatable prospect of a bank run, where depositors try to withdraw their money simultaneously. A further discussion of these issues is given in Altman (1993).

It is a "stylized fact" of corporate finance that for a corporation, external funds are more costly to obtain than internal funds, an observation which is usually attributed to problems of asymmetric information between the management of a corporation and bond and equity investors. For instance, raising external capital from outsiders by issuing new shares might be costly if the new investors, who have incomplete information about the economic prospects of a firm, interpret the share issue as a sign that the firm is overvalued. This can generate a rationale for risk management for the following reason: without risk management the increased variability of a company's cash flow will be translated either into an increased variability of the funds that need to be raised externally or to an increased variability in the amount of investment. With increasing marginal costs of raising external capital and decreasing marginal profits from new investment, we are left with a decrease in (expected) profits. Proper risk management, which amounts to a smoothing of the cash flow generated by a corporation, can therefore be beneficial. For references to the literature see Notes and Comments below.

1.5 Quantitative Risk Management

The aim of this chapter has been to place QRM in a larger historical, regulatory and even societal framework, since a study of QRM without a discussion of its proper setting and motivation makes little sense. In the remainder of the book we adopt a somewhat narrower view and treat QRM as a quantitative science that uses the language of mathematics in general, and of probability and statistics in particular.

In this section we discuss the relevance of the Q in QRM, describe the quantitative modelling challenge that we have attempted to meet in this book, and end with thoughts on where QRM may lead in the future.

1.5.1 The Q in QRM

In Section 1.2.1 we discussed the view that the use of advanced mathematical modelling and valuation techniques has been a contributory factor in financial crises, particularly those attributed to derivative products, such as CDOs in the 2007–9 crisis. We have also referred to criticism of the quantitative, statistical emphasis of the modern regulatory framework in Section 1.3.3. These arguments must be taken seriously, but we believe that it is neither possible nor desirable to remove the quantitative element from risk management.

Mathematics and statistics provide us with a suitable language and appropriate concepts for describing financial risk. This is clear for complex financial products such as derivatives, which cannot be valued and handled without mathematical models. But the need for quantitative modelling also arises for simpler products, such as a book of mortgages for retail clients. The main risk in managing such a book is the occurrence of disproportionately many defaults: a risk that is directly related to the dependence between defaults (see Chapter 11 for details). In order to describe this dependence, we need mathematical concepts from multivariate statistics, such as correlations or copulas; if we want to carry out a simulation study of the behaviour of the portfolio under different economic scenarios, we need a mathematical model that describes the joint distribution of default events; if the portfolio is large, we will also need advanced simulation techniques to generate the relevant scenarios efficiently.

Moreover, mathematical and statistical methods can do better than they did in the 2007–9 crisis. In fact, providing concepts, techniques and tools that address some of the weaker points of current methodology is a main theme of our text and we come back to this point in the next section.

There is a view that, instead of using mathematical models, there is more to be learned about risk management through a *qualitative* analysis of historical case studies and the formulation of narratives. What is often overlooked by the non-specialist is that mathematical models are themselves nothing more than narratives, albeit narratives couched in a precise symbolic language. Addressing the question "What is mathematics?", Gale and Shapley (1962) wrote: "Any argument which is carried out with sufficient precision is mathematical." Lloyd Shapley went on to win the 2012 Nobel Memorial Prize in Economic Science.

It is certainly true that mathematical methods can be misused. Mathematicians are very well aware that a mathematical result has not only a conclusion but, equally importantly, certain conditions under which it holds. Statisticians are well aware that inductive reasoning on the basis of models relies on the assumption that these conditions hold in the real world. This is especially true in economics, which as a social science is concerned with phenomena that are not easily described by clear mathematical or physical laws. By starting with questionable assumptions, models can be used (or manipulated) to deliver bad answers. In a talk on 20 March 2009, the economist Roger Guesnerie said, "For this crisis, mathematicians are innocent . . . and this in both meanings of the word." The implication is that quantitative risk managers must become more worldly about the ways in which models are used. But equally, the regulatory system needs to be more vigilant about the ways in which models can be gamed and the institutional pressures that can circumvent the best intentions of prudent quantitative risk managers.

We are firmly of the opinion—an opinion that has only been reinforced by our study of financial crises—that the Q in QRM is an essential part of the process. We reject the idea that the Q is part of the problem, and we believe that it remains (if applied correctly and honestly) a part of the solution to managing risk. In summary, we strongly agree with Shreve (2008), who said:

> Don't blame the quants. Hire good ones instead and listen to them.

1.5.2 The Nature of the Challenge

When we began this book project we set ourselves the task of defining a new discipline of QRM. Our approach to this task has had two main strands. On the one hand, we have attempted to put current practice onto a firmer mathematical footing, where, for example, concepts like P&L distributions, risk factors, risk measures, capital allocation and risk aggregation are given formal definitions and a consistent notation. In doing this we have been guided by the consideration of what topics should form the core of a course on QRM for a wide audience of students interested in risk-management issues; nonetheless, the list is far from complete and will continue to evolve as the discipline matures. On the other hand, the second strand of our endeavour has been to put together material on techniques and tools that go beyond current practice and address some of the deficiencies that have been repeatedly raised by critics. In the following paragraphs we elaborate on some of these issues.

Extremes matter. A very important challenge in QRM, and one that makes it particularly interesting as a field for probability and statistics, is the need to address unexpected, abnormal or extreme outcomes, rather than the expected, normal or average outcomes that are the focus of many classical applications. This is in tune with the regulatory view expressed by Alan Greenspan in 1995 at the Joint Central Bank Research Conference:

> From the point of view of the risk manager, inappropriate use of the normal distribution can lead to an understatement of risk, which must be

> balanced against the significant advantage of simplification. From the central bank's corner, the consequences are even more serious because we often need to concentrate on the left tail of the distribution in formulating lender-of-last-resort policies. Improving the characterization of the distribution of extreme values is of paramount importance.

While the quote is older, the same concern about underestimation of extremes is raised in a passage in the Turner Review (Lord Turner 2009):

> Price movements during the crisis have often been of a size whose probability was calculated by models (even using longer-term inputs) to be almost infinitesimally small. This suggests that the models systematically underestimated the chances of small probability high impact events.... It is possible that financial market movements are inherently characterized by fat-tail distributions. VaR models need to be buttressed by the application of stress test techniques which consider the impact of extreme movements beyond those which the model suggests are at all probable.

Much space in our book is devoted to models for financial risk factors that go beyond the normal (or Gaussian) model and attempt to capture the related phenomena of heavy or fat tails, excess volatility and extreme values.

The interdependence and concentration of risks. A further important challenge is presented by the multivariate nature of risk. Whether we look at market risk or credit risk, or overall enterprise-wide risk, we are generally interested in some form of aggregate risk that depends on high-dimensional vectors of underlying risk factors, such as individual asset values in market risk or credit spreads and counterparty default indicators in credit risk.

A particular concern in our multivariate modelling is the phenomenon of dependence between extreme outcomes, when many risk factors move against us simultaneously. In connection with the LTCM case (see Section 1.2.1) we find the following quote in *Business Week* (September 1998):

> Extreme, synchronized rises and falls in financial markets occur infrequently but they do occur. The problem with the models is that they did not assign a high enough chance of occurrence to the scenario in which many things go wrong at the same time—the "perfect storm" scenario.

In a perfect storm scenario the risk manager discovers that portfolio diversification arguments break down and there is much more of a concentration of risk than had been imagined. This was very much the case with the 2007–9 crisis: when borrowing rates rose, bond markets fell sharply, liquidity disappeared and many other asset classes declined in value, with only a few exceptions (such as precious metals and agricultural land), a perfect storm was created.

We have mentioned (see Section 1.2.1) the notorious role of the Gauss copula in the 2007–9 financial crisis. An April 2009 article in the *Economist*, with the title

"In defence of the Gaussian copula", evokes the environment at the time of the securitization boom:

> By 2001, correlation was a big deal. A new fervour was gripping Wall Street—one almost as revolutionary as that which had struck when the Black–Scholes model brought about the explosion in stock options and derivatives in the early 1980s. This was structured finance, the culmination of two decades of quants on Wall Street.... The problem, however, was correlation. The one thing any off-balance-sheet securitisation could not properly capture was the interrelatedness of all the hundreds of thousands of different mortgage loans they owned.

The Gauss copula appeared to solve this problem by offering a model for the correlated times of default of the loans or other credit-risky assets; the perils of this approach later became clear. In fact, the Gauss copula is not an example of the use of oversophisticated mathematics; it is a relatively simple model that is difficult to calibrate reliably to available market information. The modelling of dependent credit risks, and the issue of model risk in that context, is a subject we look at in some detail in our treatment of credit risk.

The problem of scale. A further challenge in QRM is the typical scale of the portfolios under consideration; in the most general case, a portfolio may represent the entire position in risky assets of a financial institution. Calibration of detailed multivariate models for all risk factors is an almost impossible task, and any sensible strategy must involve dimension reduction; that is to say, the identification of key risk drivers and a concentration on modelling the main features of the overall risk landscape.

In short, we are forced to adopt a fairly broad-brush approach. Where we use econometric tools, such as models for financial return series, we are content with relatively simple descriptions of individual series that capture the main phenomenon of volatility, and which can be used in a parsimonious multivariate factor model. Similarly, in the context of portfolio credit risk, we are more concerned with finding suitable models for the default dependence of counterparties than with accurately describing the mechanism for the default of an individual, since it is our belief that the former is at least as important as the latter in determining the risk of a large diversified portfolio.

Interdisciplinarity. Another aspect of the challenge of QRM is the fact that ideas and techniques from several existing quantitative disciplines are drawn together. When one considers the ideal education for a quantitative risk manager of the future, then a combined quantitative skill set should undoubtedly include concepts, techniques and tools from such fields as mathematical finance, statistics, financial econometrics, financial economics and actuarial mathematics. Our choice of topics is strongly guided by a firm belief that the inclusion of modern statistical and econometric techniques and a well-chosen subset of actuarial methodology are essential for the establishment of best-practice QRM. QRM is certainly not just about financial mathematics and derivative pricing, important though these may be.

Communication and education. Of course, the quantitative risk manager operates in an environment where additional non-quantitative skills are equally important. Communication is certainly an important skill: risk professionals, by the definition of their duties, will have to interact with colleagues with diverse training and backgrounds, at all levels of their organization. Moreover, a quantitative risk manager has to familiarize him or herself quickly with all-important market practice and institutional details. A certain degree of humility will also be required to recognize the role of QRM in a much larger picture.

A lesson from the 2007–9 crisis is that improved education in QRM is essential; from the front office to the back office to the boardroom, the users of models and their outputs need to be better trained to understand model assumptions and limitations. This task of educating users is part of the role of a quantitative risk manager, who should ideally have (or develop) the pedagogical skills to explain methods and conclusions to audiences at different levels of mathematical sophistication.

1.5.3 QRM Beyond Finance

The use of QRM technology is not restricted to the financial services industry, and similar developments have taken place, or are taking place, in other sectors of industry. Some of the earliest applications of QRM are to be found in the manufacturing industry, where similar concepts and tools exist under names like reliability or total quality control. Industrial companies have long recognized the risks associated with bringing faulty products to the market. The car manufacturing industry in Japan, in particular, was an early driving force in this respect.

More recently, QRM techniques have been adopted in the transport and energy industries, to name but two. In the case of energy, there are obvious similarities with financial markets: electrical power is traded on energy exchanges; derivatives contracts are used to hedge future price uncertainty; companies optimize investment portfolios combining energy products with financial products; some Basel methodology can be applied to modelling risk in the energy sector. However, there are also important dissimilarities due to the specific nature of the industry; most importantly, there are the issues of the cost of storage and transport of electricity as an underlying commodity, and the necessity of modelling physical networks including the constraints imposed by the existence of national boundaries and quasi-monopolies.

There are also markets for environmental emission allowances. For example, the Chicago Climate Futures Exchange offers futures contracts on sulphur dioxide emissions. These are traded by industrial companies producing the pollutant in their manufacturing process, and they force such companies to consider the cost of pollution as a further risk in their risk landscape.

A natural consequence of the evolution of QRM thinking in different industries is an interest in the transfer of risks between industries; this process is known as alternative risk transfer. To date the best examples of risk transfer are between the insurance and banking industries, as illustrated by the establishment of catastrophe futures by the Chicago Board of Trade in 1992. These came about in the wake of Hurricane Andrew, which caused $20 billion of insured losses on the East Coast of

the US. While this was a considerable event for the insurance industry in relation to overall reinsurance capacity, it represented only a drop in the ocean compared with the daily volumes traded worldwide on financial exchanges. This led to the recognition that losses could be covered in future by the issuance of appropriately structured bonds with coupon streams and principal repayments dependent on the occurrence or non-occurrence of well-defined natural catastrophe events, such as storms and earthquakes.

A speculative view of where these developments may lead is given by Shiller (2003), who argues that the proliferation of risk-management thinking coupled with the technological sophistication of the twenty-first century will allow any agent in society, from a company to a country to an individual, to apply QRM methodology to the risks they face. In the case of an individual this may be the risk of unemployment, depreciation in the housing market or investment in the education of children.

Notes and Comments

The language of probability and statistics plays a fundamental role throughout this book, and readers are expected to have a good knowledge of these subjects. At the elementary level, Rice (1995) gives a good first introduction to both. More advanced texts in probability and stochastic processes are Williams (1991), Resnick (1992) and Rogers and Williams (1994); the full depth of these texts is certainly not required for the understanding of this book, though they provide excellent reading material for more mathematically sophisticated readers who also have an interest in mathematical finance. Further recommended texts on statistical inference include Casella and Berger (2002), Bickel and Doksum (2001), Davison (2003) and Lindsey (1996).

In our discussion of risk and randomness in Section 1.1.1 we mentioned Knight (1921) and Keynes (1920), whose classic texts are very much worth revisiting. Knightian uncertainty refers to uncertainty that cannot be measured and is sometimes contrasted with risks that can be measured using probability. This relates to the more recent idea of a Black Swan event, a term popularized in Taleb (2007) but introduced in Taleb (2001). Black swans were believed to be imaginary creatures until the European exploration of Australia and the name is applied to unprecedented and unpredictable events that challenge conventional beliefs and models. Donald Rumsfeld, a former US Secretary of Defense, referred to "unknown unknowns" in a 2002 news briefing on the evidence for the presence of weapons of mass destruction in Iraq.

An excellent text on the history of risk and probability with financial applications in mind is Bernstein (1998). We also recommend Shiller (2012) for more on the societal context of financial risk management. A thought-provoking text addressing risk on Wall Street from a historical perspective is Brown (2012).

For the mathematical reader looking to acquire more knowledge about the relevant economics we recommend Mas-Colell, Whinston and Green (1995) for microeconomics, Campbell, Lo and MacKinlay (1997) or Gouriéroux and Jasiak (2001) for econometrics, and Brealey and Myers (2000) for corporate finance. From the

vast literature on options, an entry-level text for the general reader is Hull (2014). At a more mathematical level we like Bingham and Kiesel (2004), Musiela and Rutkowski (1997), Shreve (2004a) and Shreve (2004b). One of the most readable texts on the basic notion of options is Cox and Rubinstein (1985). For a rather extensive list of the kind of animals to be found in the zoological garden of derivatives, see, for example, Haug (1998).

There are several texts on the spectacular losses that occurred as the result of speculative trading and the careless use of derivatives. For a historical overview of financial crises, see Reinhart and Rogoff (2009), as well as the much earlier Galbraith (1993) and Kindleberger (2000). Several texts exist on more recent crises; we list only a few. The LTCM case is well documented in Dunbar (2000), Lowenstein (2000) and Jorion (2000), the latter particularly focusing on the technical risk-measurement issues involved. Boyle and Boyle (2001) give a very readable account of the Orange County, Barings and LTCM stories (see also Jacque 2010). For the Equitable Life case see the original Penrose Report, published by the UK government (Lord Penrose 2004), or an interesting paper by Roberts (2012). Many books have emerged on the 2007–9 crisis; early warnings are well summarized, under Greenspan's memorable "irrational exuberance" phrase, in a pre-crisis book by Shiller (2000), and the post-mortem by the same author is also recommended (Shiller 2008).

An overview of options embedded in life insurance products is given in Dillmann (2002), guarantees are discussed in detail in Hardy (2003), and Briys and de Varenne (2001) contains an excellent account of risk-management issues facing the (life) insurance industry. For risk-management and valuation issues underlying life insurance, see Koller (2011) and Møller and Steffensen (2007). Market-consistent actuarial valuation is discussed in Wüthrich, Bühlmann and Furrer (2010).

The historical development of banking regulation is well described in Crouhy, Galai and Mark (2001) and Steinherr (1998). For details of the current rules and regulations coming from the Basel Committee, see its website at www.bis.org/bcbs. Besides copies of the various accords, one can also find useful working papers, publications and comments written by stakeholders on the various consultative packages. For Solvency II and the Swiss Solvency Test, many documents are to be found on the web. Comprehensive textbook accounts are Sandström (2006) and Sandström (2011), and a more technical treatment is found in Wüthrich and Merz (2013). The complexity of risk-management methodology in the wake of Basel II is critically addressed by Hawke (2003), from his perspective as US Comptroller of the Currency. Among the numerous texts written after the 2007–9 crisis, we found all of Rochet (2008), Shin (2010), Dewatripont, Rochet and Tirole (2010) and Bénéplanc and Rochet (2011) useful. For a discussion of issues related to the use of fair-value accounting during the financial crisis, see Ryan (2008).

For a very detailed overview of relevant practical issues underlying risk management, we again strongly recommend Crouhy, Galai and Mark (2001). A text stressing the use of VaR as a risk measure and containing several worked examples is Jorion (2007), whose author also has a useful teaching manual on the same subject

(Jorion 2002b). Insurance-related issues in risk management are nicely presented in Doherty (2000).

For a comprehensive discussion of the management of bank capital given regulatory constraints, see Matten (2000), Klaassen and van Eeghen (2009) and Admati and Hellwig (2013). Graham and Rogers (2002) contains a discussion of risk management and tax incentives. A formal account of the Modigliani–Miller Theorem and its implications can be found in many textbooks on corporate finance: a standard reference is Brealey and Myers (2000), and de Matos (2001) gives a more theoretical account from the perspective of modern financial economics. Both texts also discuss the implications of informational asymmetries between the various stakeholders in a corporation. Formal models looking at risk management from a corporate-finance angle are to be found in Froot and Stein (1998), Froot, Scharfstein and Stein (1993) and Stulz (1996, 2002). For a specific discussion on corporate-finance issues in insurance, see Froot (2007) and Hancock, Huber and Koch (2001).

There are several studies on the use of risk-management techniques for non-financial firms (see, for example, Bodnar, Hayt and Marston 1998; Geman 2005, 2009). Two references in the area of the reliability of industrial processes are Bedford and Cooke (2001) and Does, Roes and Trip (1999). Interesting edited volumes on alternative risk transfer are Shimpi (2001), Barrieu and Albertini (2009) and Kiesel, Scherer and Zagst (2010); a detailed study of model risk in the alternative risk transfer context is Schmock (1999). An area we have not mentioned so far in our discussion of QRM in the future is that of real options. A real option is the right, but not the obligation, to take an action (e.g. deferring, expanding, contracting or abandoning) at a predetermined cost called the exercise price. The right holds for a predetermined period of time—the life of the option. This definition is taken from Copeland and Antikarov (2001). Examples of real options discussed in the latter are the valuation of an internet project and of a pharmaceutical research and development project. A further useful reference is Brennan and Trigeorgis (1999).

A well-written critical view of the failings of the standard approach to risk management is given in Rebonato (2007). And finally, for an entertaining text on the biology of the much criticized "homo economicus", we like Coates (2012).

2

Basic Concepts in Risk Management

In this chapter we define or explain a number of fundamental concepts used in the measurement and management of financial risk. Beginning in Section 2.1 with the simplified balance sheet of a bank and an insurer, we discuss the risks faced by such firms, the nature of capital, and the need for a firm to have sufficient capital to withstand financial shocks and remain solvent.

In Section 2.2 we establish a mathematical framework for describing changes in the value of portfolios and deriving loss distributions. We provide a number of examples to show how this framework applies to different kinds of asset and liability portfolios. The examples are also used to discuss the meaning of value in more detail with reference to fair-value accounting and risk-neutral valuation.

Section 2.3 is devoted to the subject of using risk measures to determine risk or solvency capital. We present different quantitative approaches to measuring risk, with a particular focus on risk measures that are calculated from loss distributions, like value-at-risk and expected shortfall.

2.1 Risk Management for a Financial Firm

2.1.1 Assets, Liabilities and the Balance Sheet

A good way to understand the risks faced by a modern financial institution is to look at the stylized balance sheet of a typical bank or insurance company. A balance sheet is a financial statement showing *assets and liabilities*; roughly speaking, the assets describe the financial institution's investments, whereas liabilities refer to the way in which funds have been raised and the obligations that ensue from that fundraising.

A typical bank raises funds by taking in customer deposits, by issuing bonds and by borrowing from other banks or from central banks. Collectively these form the *debt capital* of the bank, which is invested in a number of ways. Most importantly, it is used for loans to retail, corporate and sovereign customers, invested in traded securities, lent out to other banks or invested in property or in other companies. A small fraction is also held as cash.

A typical insurance company sells insurance contracts, collecting premiums in return, and raises additional funds by issuing bonds. The liabilities of an insurance company thus consist of its obligations to policyholders, which take the form of a *technical reserve against future claims*, and its obligations to bondholders. The funds raised are then invested in traded securities, particularly bonds, as well as other assets such as real estate.

Table 2.1. The stylized balance sheet of a typical bank.

Bank ABC (31 December 2015)			
Assets		Liabilities	
Cash (and central bank balance)	£10M	Customer deposits	£80M
Securities	£50M	Bonds issued	
– bonds		– senior bond issues	£25M
– stocks		– subordinated bond issues	£15M
– derivatives		Short-term borrowing	£30M
Loans and mortgages	£100M	Reserves (for losses on loans)	£20M
– corporates			
– retail and smaller clients		*Debt (sum of above)*	£170M
– government			
Other assets	£20M		
– property			
– investments in companies		*Equity*	£30M
Short-term lending	£20M		
Total	£200M	Total	£200M

In both cases a small amount of extra funding stems from occasional *share issues*, which form the share capital of the bank or insurer. This form of funding is crucial as it entails no obligation towards outside parties.

These simplified banking and insurance *business models* are reflected in the stylized balance sheets shown in Tables 2.1 and 2.2. In these financial statements, assets and liabilities are valued on a given date. The position marked *equity* on the liability side of the balance sheet is the residual value defined in the balance sheet equation

$$\text{value of assets} = \text{value of liabilities} = \text{debt} + \text{equity}. \tag{2.1}$$

A company is *solvent* at a given point in time if the equity is nonnegative; otherwise it is insolvent. Insolvency should be distinguished from the notion of default, which occurs if a firm misses a payment to its debtholders or other creditors. In particular, an otherwise-solvent company can default because of *liquidity* problems, as discussed in more detail in the next section.

It should be noted that assigning values to the items on the balance sheet of a bank or insurance company is a non-trivial task. Broadly speaking, two different approaches can be distinguished. The practice of *fair-value accounting* attempts to value assets at the prices that would be received if they were sold and to value liabilities at the prices that would have to be paid if they were transferred to another party. Fair-value accounting is relatively straightforward for positions that are close to securities traded on liquid markets, since these are simply valued by (an estimate of) their market price. It is more challenging to apply fair-value principles to non-traded or illiquid assets and liabilities.

The more traditional practice of *amortized cost accounting* is still applied to many kinds of financial asset and liability. Under this practice the position is assigned a

Table 2.2. The stylized balance sheet of a typical insurer.

Insurer XYZ (31 December 2015)			
Assets		Liabilities	
Investments		Reserves for policies written (technical provisions)	£80M
– bonds	£50M		
– stocks	£5M	Bonds issued	£10M
– real estate	£5M		
Investments for unit-linked contracts	£30M	*Debt (sum of above)*	£90M
Other assets	£10M		
– property			
		Equity	£10M
Total	£100M	Total	£100M

book value at its inception and this is carried forward over time. In some cases the value is progressively reduced or impaired to account for the aging of the position or the effect of adverse events. An example of assets valued at book value are the loans on the balance sheet of the bank. The book value would typically be an estimate of the present value (at the time the loans were made) of promised future interest and principal payments minus a provision for losses due to default.

In the European insurance industry the practice of *market-consistent valuation* has been promoted under the Solvency II framework. As described in Section 1.3.2, the rationale is very similar to that of fair-value accounting: namely, to value positions by "the amount for which an asset could be exchanged or a liability settled, between knowledgeable, willing parties in an arm's length transaction, based on observable prices within an active, deep and liquid market". However, there are some differences between market-consistent valuation and fair-value accounting for specific kinds of position. A European insurer will typically have two versions of the balance sheet in order to comply with accounting rules, on the one hand, and Solvency II rules for capital adequacy, on the other. The accounting balance sheet may mix fair-value and book-value approaches, but the Solvency II balance sheet will apply market-consistent principles throughout.

Overall, across the financial industry, there is a tendency for the accounting standard to move towards fair-value accounting, even if the financial crisis of 2007–9 demonstrated that this approach is not without problems during periods when trading activity and market liquidity suddenly vanish (see Section 1.3.3 for more discussion of this issue). Fair-value accounting for financial products will be discussed in more detail in Section 2.2.2.

2.1.2 Risks Faced by a Financial Firm

An obvious source of risk for a bank is a decrease in the value of its investments on the asset side of the balance sheet. This includes market risk, such as losses from securities trading, and credit risk. Another important risk is related to funding and

so-called *maturity mismatch*: for a typical bank, large parts of the asset side consist of relatively illiquid, long-term investments such as loans or property, whereas a large part of the liabilities side consists of short-term obligations such as funds borrowed from money markets and most customer deposits. This may lead to problems when the cost of short-term refinancing increases due to rising short-term interest rates, because the banks may have difficulties selling long-term assets to raise funds. This can lead to the default of a bank that is technically solvent; in extreme cases there might even be a *bank run*, as was witnessed during the 2007–9 financial crisis. This clearly shows that risk is found on both sides of the balance sheet and that risk managers should not focus exclusively on the asset side.

The primary risk for an insurance company is clearly insolvency, i.e. the risk that the claims of policyholders cannot be met. This can happen due to adverse events affecting the asset side or the liability side of the balance sheet. On the asset side, the risks are similar to those for a bank. On the liability side, the main risk is that reserves are insufficient to cover future claim payments. It is important to bear in mind that the liabilities of a life insurer are of a long-term nature (due to the sale of products such as annuities) and are subject to many categories of risk including interest-rate risk, inflation risk and longevity risk, some of which also affect the asset side. An important aspect of the risk-management strategy of an insurance company is, therefore, to hedge parts of these risks by proper investment of the premium income (so-called liability-driven investment).

It should be clear from this discussion that a sound approach to risk management cannot look at one side of the balance sheet in isolation from the other.

2.1.3 Capital

There are many different notions of bank *capital*, and three broad concepts can be distinguished: *equity (or book) capital*, *regulatory capital* and *economic capital*. All of these notions of capital refer to items on the liability side of the balance sheet that entail no (or very limited) obligations to outside creditors and that can thus serve as a buffer against losses.

The equity capital can be read from the balance sheet according to the balance sheet equation in (2.1). It is therefore a measure of the value of the company to the shareholders. The balance sheet usually gives a more detailed breakdown of the equity capital by listing separate positions for *shareholder capital*, *retained earnings* and other items of lesser importance. Shareholder capital is the initial capital invested in the company by purchasers of equity. For companies financed by a single share issue, this is given by the numbers of shares issued multiplied by their price at the issuance date. Shareholder capital is therefore different from market capitalization, which is given by the number of shares issued multiplied by their current market price. Retained earnings are the accumulated earnings that have not been paid out in the form of dividends to shareholders; these can in principle be negative if the company has made losses.

Regulatory capital is the amount of capital that a company should have according to regulatory rules. For a bank, the rules are set out in the Basel framework,

as described in more detail in Section 1.3.1. For European insurance companies, regulatory capital takes the form of a minimum capital requirement and a solvency capital requirement as set out in the Solvency II framework (see Section 1.3.2).

A regulatory capital framework generally specifies the amount of capital necessary for a financial institution to continue its operations, taking into account the size and the riskiness of its positions. Moreover, it specifies the quality of the capital and hence the form it should take on the balance sheet. In this context one usually distinguishes between different numbered capital *tiers*.

For example, in the Basel framework, Tier 1 capital is the sum of shareholder capital and retained earnings; in other words, the main constituents of the equity capital. This capital can act in full as a buffer against losses as there are no other claims on it. Tier 2 capital includes other positions of the balance sheet, in particular subordinated debt. Holders of this debt would effectively be the last to be paid before the shareholders in the event of the liquidation of the company, so subordinated debt can be viewed as an extra layer of protection for depositors and other senior debtholders. For illustration, the bank in Table 2.1 has Tier 1 capital of £30 million (assuming the equity capital consists of shareholder capital and retained earnings only) and Tier 2 capital of £45 million.

Economic capital is an estimate of the amount of capital that a financial institution needs in order to control the probability of becoming insolvent, typically over a one-year horizon. It is an internal assessment of risk capital that is guided by economic modelling principles. In particular, an economic capital framework attempts to take a holistic view that looks at assets and liabilities simultaneously, and works, where possible, with fair or market-consistent values of balance sheet items. Although, historically, regulatory capital frameworks have been based more on relatively simple rules and on book values for balance sheet items, there is increasing convergence between the economic and regulatory capital concepts, particularly in the insurance world, where Solvency II emphasizes market-consistent valuation of liabilities.

Note that the various notions of capital refer to the way in which a financial firm finances itself and not to the assets it invests in. In particular, capital requirements do not require the setting aside of funds that cannot be invested productively, e.g. by issuing new loans. There are other forms of financial regulation that refer to the asset side of the balance sheet and restrict the investment possibilities, such as obligatory cash reserves for banks and constraints on the proportion of insurance assets that may be invested in stocks.

Notes and Comments

A good introduction to the business of banking and the risks affecting banks is Choudhry (2012), while Thoyts (2010) provides a very readable overview of theory and practice in the insurance industry, with a focus on the UK. Readers wanting to go deeper into the subject of balance sheets have many financial accounting textbooks to choose from, a popular one being Elliott and Elliott (2013). A paper that gives more explanation of fair-value accounting and also discusses issues raised by the financial crisis is Ryan (2008).

Regulatory capital in the banking industry is covered in many of the documents produced by the Basel Committee, in particular the papers covering the Basel II and Basel III capital frameworks (Basel Committee on Banking Supervision 2006, 2011). For regulatory capital under Solvency II, see Sandström (2011). Textbook treatments of the management of bank capital given regulatory constraints are found in Matten (2000) and Klaassen and van Eeghen (2009), while Admati et al. (2013) provides a strong argument for capital regulation that ensures banks have a high level of equity capital. This issue is discussed at a slightly less technical level in the book by Admati and Hellwig (2013). A good explanation of the concept of economic capital may be found in the relevant entry in the *Encyclopedia of Quantitative Finance* (Rosen and Saunders 2010).

2.2 Modelling Value and Value Change

We have seen in Section 2.1.1 that an analysis of the risks faced by a financial institution requires us to consider the change in the value of its assets and liabilities. In Section 2.2.1 we set up a formal framework for modelling value and value change and illustrate this framework with stylized asset and liability portfolios. With the help of these examples we take a closer look at valuation methods in Section 2.2.2. Finally, in Section 2.2.3 we discuss the different approaches that are used to construct loss distributions for portfolios over given time horizons.

2.2.1 Mapping Risks

In our general mathematical model for describing financial risks we represent the uncertainty about future states of the world by a probability space $(\Omega, \mathcal{F}, P)$, which is the domain of all random variables (rvs) we introduce below.

We consider a given portfolio of assets and, in some cases, liabilities. At the simplest level, this could be a collection of stocks or bonds, a book of derivatives or a collection of risky loans. More generally, it could be a portfolio of life insurance contracts (liabilities) backed by investments in securities such as bonds, or even a financial institution's overall balance sheet. We denote the *value* of the portfolio at time t by V_t and assume that the rv V_t is known, or can be determined from information available, at time t. Of course, the valuation of many positions on a financial firm's balance sheet is a challenging task; we return to this issue in more detail in Section 2.2.2.

We consider a given risk-management time horizon Δt, which might be one day or ten days in market risk, or one year in credit, insurance or enterprise-wide risk management. To develop a simple formalism for talking about value, value change and the role of risk factors, we will make two simplifying assumptions:

- the portfolio composition remains fixed over the time horizon; and
- there are no intermediate payments of income during the time period.

While these assumptions may hold approximately for a one-day or ten-day horizon, they are unlikely to hold over one year, where items in the portfolio may mature

and be replaced by other investments and where dividend or interest income may accumulate. In specific situations it would be possible to relax these assumptions, e.g. by specifying simple rebalancing rules for portfolios or by taking intermediate income into account.

Using a time-series notation (with time recorded in multiples of the time horizon Δt) we write the value of the portfolio at the end of the time period as V_{t+1} and the change in value of the portfolio as $\Delta V_{t+1} = V_{t+1} - V_t$. We define the *loss* to be $L_{t+1} := -\Delta V_{t+1}$, which is natural for short time intervals. For longer time intervals, on the other hand, this definition neglects the time value of money, and an alternative would be to define the loss to be $V_t - V_{t+1}/(1 + r_{t,1})$, where $r_{t,1}$ is the simple risk-free interest rate that applies between times t and $t + 1$; this measures the loss in units of money at time t. The rv L_{t+1} is typically random from the viewpoint of time t, and its distribution is termed the *loss distribution*. Practitioners in risk management are often concerned with the so-called P&L distribution. This is the distribution of the change in portfolio value ΔV_{t+1}. In this text we will often focus on L_{t+1} as this simplifies the application of many statistical methods and is in keeping with conventions in actuarial risk theory.

The value V_t is typically modelled as a function of time and a d-dimensional random vector $\boldsymbol{Z}_t = (Z_{t,1}, \ldots, Z_{t,d})'$ of *risk factors*, i.e. we have the representation

$$V_t = f(t, \boldsymbol{Z}_t) \tag{2.2}$$

for some measurable function $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$. Risk factors are usually assumed to be observable, so the random vector $\boldsymbol{Z}_t$ takes some known realized value $\boldsymbol{z}_t$ at time t and the portfolio value V_t has realized value $f(t, \boldsymbol{z}_t)$. The choice of the risk factors and of f is of course a modelling issue and depends on the portfolio at hand, on the data available and on the desired level of precision (see also Section 2.2.2). A representation of the portfolio value in the form (2.2) is termed a *mapping* of risks. Some examples of the mapping procedure are provided below.

We define the random vector of *risk-factor changes* over the time horizon to be $\boldsymbol{X}_{t+1} := \boldsymbol{Z}_{t+1} - \boldsymbol{Z}_t$. Assuming that the current time is t and using the mapping (2.2), the portfolio loss is given by

$$L_{t+1} = -(f(t + 1, \boldsymbol{z}_t + \boldsymbol{X}_{t+1}) - f(t, \boldsymbol{z}_t)), \tag{2.3}$$

which shows that the loss distribution is determined by the distribution of the risk-factor change $\boldsymbol{X}_{t+1}$.

If f is differentiable, we may also use a first-order approximation L^{Δ}_{t+1} of the loss in (2.3) of the form

$$L^{\Delta}_{t+1} := -\left(f_t(t, \boldsymbol{z}_t) + \sum_{i=1}^{d} f_{z_i}(t, \boldsymbol{z}_t) X_{t+1,i} \right), \tag{2.4}$$

where the subscripts on f denote partial derivatives. The notation L^{Δ} stems from the standard *delta* terminology in the hedging of derivatives (see Example 2.2 below). The first-order approximation is convenient as it allows us to represent the loss as

a *linear* function of the risk-factor changes. The quality of the approximation (2.4) is obviously best if the risk-factor changes are likely to be small (i.e. if we are measuring risk over a short horizon) and if the portfolio value is almost linear in the risk factors (i.e. if the function f has small second derivatives).

We now consider a number of examples from the areas of market, credit and insurance risk, illustrating how typical risk-management problems fit into this framework.

Example 2.1 (stock portfolio). Consider a fixed portfolio of d stocks and denote by λ_i the number of shares of stock i in the portfolio at time t. The price process of stock i is denoted by $(S_{t,i})_{t\in\mathbb{N}}$. Following standard practice in finance and risk management we use logarithmic prices as risk factors, i.e. we take $Z_{t,i} := \ln S_{t,i}$, $1 \leqslant i \leqslant d$, and we get $V_t = \sum_{i=1}^{d} \lambda_i \mathrm{e}^{Z_{t,i}}$. The risk-factor changes $X_{t+1,i} = \ln S_{t+1,i} - \ln S_{t,i}$ then correspond to the log-returns of the stocks in the portfolio. The portfolio loss from time t to $t+1$ is given by

$$L_{t+1} = -(V_{t+1} - V_t) = -\sum_{i=1}^{d} \lambda_i S_{t,i}(\mathrm{e}^{X_{t+1,i}} - 1),$$

and the linearized loss L^{Δ}_{t+1} is given by

$$L^{\Delta}_{t+1} = -\sum_{i=1}^{d} \lambda_i S_{t,i} X_{t+1,i} = -V_t \sum_{i=1}^{d} w_{t,i} X_{t+1,i}, \tag{2.5}$$

where the weight $w_{t,i} := (\lambda_i S_{t,i})/V_t$ gives the proportion of the portfolio value invested in stock i at time t. Given the mean vector and covariance matrix of the distribution of the risk-factor changes, it is very easy to compute the first two moments of the distribution of the linearized loss L^{Δ}. Suppose that the random vector $\boldsymbol{X}_{t+1}$ has a distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. Using general rules for the mean and variance of linear combinations of a random vector (see also equations (6.7) and (6.8)), we immediately get

$$E(L^{\Delta}_{t+1}) = -V_t \boldsymbol{w}'\boldsymbol{\mu} \quad \text{and} \quad \operatorname{var}(L^{\Delta}_{t+1}) = V_t^2 \boldsymbol{w}'\Sigma\boldsymbol{w}. \tag{2.6}$$

Example 2.2 (European call option). We now consider a simple example of a portfolio of derivative securities: namely, a standard European call on a non-dividend-paying stock with maturity time T and exercise price K. We use the *Black–Scholes option-pricing formula* for the valuation of our portfolio. The value of a call option on a stock with price S at time t is given by

$$C^{\mathrm{BS}}(t, S; r, \sigma, K, T) := S\Phi(d_1) - K\mathrm{e}^{-r(T-t)}\Phi(d_2), \tag{2.7}$$

where Φ denotes the standard normal distribution function (df), r represents the continuously compounded risk-free interest rate, σ denotes the volatility of the underlying stock, and where

$$d_1 = \frac{\ln(S/K) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T-t}. \tag{2.8}$$

For notational simplicity we assume that the time to maturity of the option, $T-t$, is measured in units of the time horizon, and that the parameters r and σ are expressed in terms of those units; for example, if the time horizon is one day, then r and σ are the daily interest rate and volatility. This differs from standard market practice where time is measured in years and r and σ are expressed in annualized terms.

To map the portfolio at time t, let S_t denote the stock price at time t and let r_t and σ_t denote the values that a practitioner chooses to use at that time for the interest rate and volatility. The log-price of the stock ($\ln S_t$) is an obvious risk factor for changes in value of the portfolio. While in the Black–Scholes option-pricing model the interest rate and volatility are assumed to be constant, in real markets interest rates change constantly, as do the *implied volatilities* that practitioners tend to use as inputs for the volatility parameter. Hence, we take $\boldsymbol{Z}_t = (\ln S_t, r_t, \sigma_t)'$ as the vector of risk factors.

According to the Black–Scholes formula the value of the call option at time t equals $C^{\mathrm{BS}}(t, S_t; r_t, \sigma_t, K, T)$, which is of the form (2.2). The risk-factor changes are given by

$$\boldsymbol{X}_{t+1} = (\ln S_{t+1} - \ln S_t, r_{t+1} - r_t, \sigma_{t+1} - \sigma_t)',$$

and the linearized loss can be calculated to be

$$L_{t+1}^{\Delta} = -(C_t^{\mathrm{BS}} + C_S^{\mathrm{BS}} S_t X_{t+1,1} + C_r^{\mathrm{BS}} X_{t+1,2} + C_\sigma^{\mathrm{BS}} X_{t+1,3}), \tag{2.9}$$

where the subscripts denote partial derivatives of the Black–Scholes formula (2.7). Note that we have omitted the arguments of C^{BS} to simplify the notation. Note also that an S_t term appears because we take the equity risk factor to be the log-price of the stock rather than the price; applying the chain rule with $S = \mathrm{e}^{z_1}$ we have

$$C_{z_1}^{\mathrm{BS}} = C_S^{\mathrm{BS}} \frac{\mathrm{d}S}{\mathrm{d}z_1}\bigg|_{z_1=\ln S_t} = C_S^{\mathrm{BS}} S_t.$$

In Section 9.1.2 and Example 9.1 we give more detail concerning the derivation of mapping formulas similar to (2.9) and pay more attention to the choice of timescale in the mapping function.

The derivatives of the Black–Scholes option-pricing function are often referred to as the *Greeks*: C_S^{BS} (the partial derivative with respect to the stock price S) is called the *delta* of the option; C_t^{BS} (the partial derivative with respect to time) is called the *theta* of the option; C_r^{BS} (the partial derivative with respect to the interest rate r) is called the *rho* of the option; and, in a slight abuse of the Greek language, C_σ^{BS} (the partial derivative with respect to volatility σ) is called the *vega* of the option. The Greeks play an important role in the risk management of derivative portfolios.

The reader should keep in mind that for portfolios containing derivatives, the linearized loss can be a rather poor approximation of the true loss, since the portfolio value is often a highly nonlinear function of the risk factors. This has led to the development of higher-order approximations such as the *delta–gamma approximation*, where first- and second-order derivatives are used (see Section 9.1.2).

Example 2.3 (stylized loan portfolio). In this example we show how losses from a portfolio of short-term loans fit into our general framework; a detailed discussion of models for loan portfolios will be presented in Chapter 11.

Following standard practice in credit risk management, the risk-management horizon Δt is taken to be one year. We consider a portfolio of loans to m different borrowers or obligors that have been made at time t and valued using a book-value approach. To keep the example simple we assume that all loans have to be repaid at time $t+1$. We denote the amount to be repaid by obligor i by k_i; this term comprises the interest payment at $t+1$ and the repayment of the loan principal. The *exposure* to obligor i is defined to be the present value of the promised interest and principal cash flows, and it is therefore given by $e_i = k_i/(1+r_{t,1})$.

In order to take the possibility of default into account we introduce a series of random variables $(Y_{t,i})_{t\in\mathbb{N}}$ that represent the default state of obligor i at t, and we let $Y_{t,i} = 1$ if obligor i has defaulted by time t, with $Y_{t,i} = 0$ otherwise. These variables are known as *default indicators*. For simplicity we assume that all obligors are in a non-default state at time t, so $Y_{t,i} = 0$ for all $1 \leqslant i \leqslant m$.

In keeping with valuation conventions, in practice we define the book value of a loan to be the exposure of the loan reduced by the discounted expected loss due to default; in this way the valuation includes a provision for default risk. We assume that in the case of a default of borrower i, the lender can recover an amount $(1-\delta_i)k_i$ at the maturity date $t+1$, where $\delta_i \in (0, 1]$ describes the so-called *loss given default* of the loan, which is the percentage of the exposure that is lost in the event of default. Moreover, we denote by p_i the probability that obligor i defaults in the period $(t, t+1]$. In this introductory example we suppose that δ_i and p_i are known constants. In practice, p_i could be estimated using a credit scoring model (see Section 10.2 for more discussion). The discounted expected loss due to a default of obligor i is thus given by

$$\frac{1}{1+r_{t,1}}\delta_i p_i k_i = \delta_i p_i e_i.$$

The book value of loan i is therefore equal to $e_i(1-\delta_i p_i)$, the discounted expected pay-off of the loan. Note that in practice, one would make further provisions for administrative, refinancing and capital costs, but we ignore these issues for the sake of simplicity. Moreover, one should keep in mind that the book value is not an estimate for the fair value of the loan (an estimate for the amount for which the loan could be sold in a securitization deal); the latter is usually lower than the discounted expected pay-off of the loan, as investors demand a premium for bearing the default risk (see also our discussion of risk-neutral valuation in the next section). The book value of the loan portfolio at time t is thus given by

$$V_t = \sum_{i=1}^{m} e_i(1-\delta_i p_i).$$

The value of a loan to obligor i at the maturity date $t+1$ equals the size of the repayment and is therefore equal to k_i if $Y_{t+1,i} = 0$ (no default of obligor i) and

equal to $(1-\delta_i)k_i$ if $Y_{t+1,i}=1$ (default of obligor i). Hence, V_{t+1}, the value of the portfolio at time $t+1$, equals

$$V_{t+1}=\sum_{i=1}^{m}((1-Y_{t+1,i})k_i+Y_{t+1,i}(1-\delta_i)k_i)=\sum_{i=1}^{m}k_i(1-\delta_i Y_{t+1,i}).$$

Since we use a relatively long risk-management horizon of one year, it is natural to discount V_{t+1} in computing the portfolio loss. Again using the fact that $e_i = k_i/(1+r_{t,1})$, we obtain

$$\begin{aligned}L_{t+1}=V_t-\frac{V_{t+1}}{1+r_{t,1}}&=\sum_{i=1}^{m}e_i(1-\delta_i p_i)-\sum_{i=1}^{m}e_i(1-\delta_i Y_{t+1,i})\\&=\sum_{i=1}^{m}\delta_i e_i Y_{t+1,i}-\sum_{i=1}^{m}\delta_i e_i p_i,\end{aligned}$$

which gives a simple formula for the portfolio loss involving exposures, default probabilities, losses given default and default indicators.

Finally, we explain how this example fits into the mapping framework given by (2.2) and (2.3). In this case the risk factors are the default indicator variables $\boldsymbol{Z}_t=(Y_{t,1},\dots,Y_{t,m})'$. If we write the mapping formula as

$$f(s,\boldsymbol{Z}_s)=\sum_{i=1}^{m}(1-Y_{s,i})\frac{k_i}{(1+r_{s,1})^{t+1-s}}(1-(t+1-s)\delta_i p_i)+\sum_{i=1}^{m}Y_{s,i}k_i(1-\delta_i),$$

we see that this gives the correct portfolio values at times $s=t$ and $s=t+1$. The issue of finding and calibrating a good model for the joint distribution of the risk-factor changes $\boldsymbol{Z}_{t+1}-\boldsymbol{Z}_t$ is taken up in Chapter 11.

Example 2.4 (insurance example). We consider a simple whole-life annuity product in which a policyholder (known as an annuitant) has purchased the right to receive a series of payments as long as he or she remains alive. Although realistic products would typically make monthly payments, we assume annual payments for simplicity and consider a risk-management horizon of one year.

At time t we assume that an insurer has a portfolio of n annuitants with current ages x_i, $i=1,\dots,n$. Annuitant i receives a fixed annual annuity payment κ_i, and the time of their death is represented by the random variable τ_i. The annuity payments are made in arrears at times $t+1, t+2, \dots$, a form of product known as a whole-life immediate annuity.

At time t there is uncertainty about the value of the cash flow to any individual annuitant stemming from the uncertainty about their time of death. The liability due to a single annuitant takes the form

$$\sum_{h=1}^{\infty}I_{\{\tau_i>t+h\}}\kappa_i D(t,t+h),$$

where $D(t,t+h)$ is a discount factor that gives the time-t value of one unit paid out at time $t+h$. Following standard discrete-time actuarial practice we set $D(t,t+h)=$

$(1+r_{t,h})^{-h}$, where $r_{t,h}$ is the h-year simple spot interest rate at time t. The expected present value of this liability at time t is given by

$$\sum_{h=1}^{\infty} q_t(x_i, h)\kappa_i \frac{1}{(1+r_{t,h})^h},$$

where we assume that $P(\tau_i > t+h) = q_t(x_i, h)$. In other words, the survival probability of annuitant i depends on only the current time t and the age x_i of the annuitant; $q_t(x, h)$ represents the probability that an individual aged x at time t will survive a further h years.

If n is sufficiently large, diversification arguments suggest that individual mortality risk (deviation of the variables $\tau_1, \ldots, \tau_n$ from their expected values) may be neglected, and the overall portfolio liability may be represented by

$$B_t = \sum_{i=1}^{n} \sum_{h=1}^{\infty} q_t(x_i, h)\kappa_i \frac{1}{(1+r_{t,h})^h}.$$

Now consider the liability due to a single annuitant at time $t+1$, which is given by

$$\sum_{h=1}^{\infty} I_{\{\tau_i > t+1+h\}}\kappa_i \frac{1}{(1+r_{t+1,h})^h}.$$

We again use the large-portfolio diversification argument to replace $I_{\{\tau_i > t+1+h\}}$ by its expected value $q_t(x_i, h+1)$, and thus we approximate the portfolio liability at $t+1$ by

$$B_{t+1} = \sum_{i=1}^{n} \sum_{h=1}^{\infty} q_t(x_i, h+1)\kappa_i \frac{1}{(1+r_{t+1,h})^h}.$$

The lump-sum premium payments of the annuitants would typically be invested in a matching portfolio of bonds: that is, a portfolio chosen so that the cash flows from the bonds closely match the cash flows due to the policyholders. We assume that the investments have been made in (default-free) government bonds with d different maturities (all greater than or equal to one year) so that the asset value at time t is

$$A_t = \sum_{j=1}^{d} \frac{\lambda_j}{(1+r_{t,h_j})^{h_j}},$$

where h_j is the maturity of the jth bond and λ_j is the number of such bonds that have been purchased. The net asset value of the portfolio at time t is given by $V_t = A_t - B_t$.

This is a situation in which it would be natural to discount future asset and liability values back to time t, so that the loss (in units of time-t money) would be given by

$$L_{t+1} = -\left(\frac{A_{t+1}}{1+r_{t,1}} - A_t\right) + \left(\frac{B_{t+1}}{1+r_{t,1}} - B_t\right).$$

The risk factors in this example are the spot rates $\mathbf{Z}_t = (r_{t,1}, \ldots, r_{t,m})'$, where m represents the maximum time horizon at which an annuity payment might have to

be made. The mortality risk in the lifetime variables $\tau_1, \ldots, \tau_m$ is eliminated from consideration by using the *life table* of fixed survival probabilities $\{q_t(x, h)\}$.

2.2.2 Valuation Methods

We now take a closer look at valuation principles in the light of the stylized examples of the previous section. While the loan portfolio of Example 2.3 would typically be valued using a book-value approach, as indicated, the stock portfolio (Example 2.1), the European call option (Example 2.2) and the asset-backed annuity portfolio (Example 2.4) would all be valued using a fair-value approach in practice. In this section we elaborate on the different methods used in fair-value accounting and explain how risk-neutral valuation may be understood as a special case of fair-value accounting.

We recall from Section 2.1.1 that the use of fair-value methodology for the assets and liabilities of an insurer is closely related to the concept of market-consistent valuation. The main practical difference is that the fair-value approach is applied to the accounting balance sheet for reporting purposes, whereas market-consistent valuation is applied to the Solvency II balance sheet for capital adequacy purposes. While there are differences in detail between the two rule books, it is sufficient for our purposes to view market-consistent valuation as a variant of fair-value accounting.

Fair-value accounting. In general terms, the fair value of an asset is an estimate of the price that would be received in selling the asset in a transaction on an active market. Similarly, the fair value of a liability is an estimate of the price that would have to be paid to transfer the liability to another party in a market-based transaction; this is sometimes referred to as the exit value.

Only a minority of balance sheet positions are traded directly in an active market. Accountants have therefore developed a three-stage hierarchy of fair-value accounting methods, extending fair-value accounting to non-traded items. This hierarchy, which is codified in the US as Financial Accounting Standard 157 and worldwide in the 2009 amendment to International Financial Reporting Standard 7, has the following levels.

Level 1: the fair value of an instrument is determined from quoted prices in an active market for the same instrument, without modification or repackaging.

Level 2: the fair value of an instrument is determined using quoted prices in active markets for similar (but not identical) instruments or by the use of valuation techniques, such as pricing models for derivatives, for which all significant inputs are based on observable market data.

Level 3: the fair value of an instrument is estimated using a valuation technique (pricing model) for which some key inputs are not observable market data (or otherwise publicly observable quantities).

In risk-management language these levels are sometimes described as mark-to-market, mark-to-model with objective inputs, and mark-to-model with subjective inputs.

The stock portfolio in Example 2.1 is a clear example of Level 1 valuation: the portfolio value is determined by simply looking up current market prices of the stocks.

Now consider the European call option of Example 2.2 and assume that the option is not traded on the market, perhaps because of a non-standard strike price or maturity, but that there is an otherwise active market for options on that stock. This would be an example of Level 2 valuation: a valuation technique (namely, the Black–Scholes option-pricing formula) is used to price the instrument. The inputs to the formula are the stock price, the interest rate and the implied volatility, which are market observables (since we assumed that there is an active market for options on the stock).

The insurance portfolio from Example 2.4 can be viewed as an example of Level 2 or Level 3 valuation, depending on the methods used to determine the input parameters. If the survival probabilities are determined from publicly available sources such as official life tables, the annuity example corresponds to Level 2 valuation since the other risk factors are essentially market observables, with the possible exception of long-term interest rates. If, on the other hand, proprietary data and methods are used to estimate the survival probabilities, then this would be Level 3 valuation.

Risk-neutral valuation. Risk-neutral valuation is a special case of fair-value accounting that is widely used in the pricing of financial products such as derivative securities. In risk-neutral pricing the values of financial instruments are computed as expected discounted values of future cash flows, where expectation is taken with respect to some probability measure Q, called a *risk-neutral pricing measure*. Q is an artificial measure that turns the discounted prices of traded securities into so-called martingales (fair bets), and it is also known as an *equivalent martingale measure*. Calibration procedures are used to ensure that prices obtained in this way are consistent with quoted market prices.

Hitherto, all our probabilities and expectations have been taken with respect to the *physical* or *real-world measure* P. In order to explain the concept of a risk-neutral measure Q and to illustrate the relationship between P and Q, we use a simple one-period model from the field of credit risk, which we refer to as the basic one-period default model. We consider a defaultable zero-coupon bond with maturity T equal to one year and make the following assumptions: the real-world default probability is $p = 1\%$; the recovery rate $1 - \delta$ (the proportion of the notional of the bond that is paid back in the case of a default) is deterministic and is equal to 60%; the risk-free simple interest rate equals 5%; the current ($t = 0$) price of the bond is $p_1(0, 1) = 0.941$; the price of the corresponding default-free bond is $p_0(0, 1) = (1.05)^{-1} = 0.952$. The price evolution of the bond is depicted in Figure 2.1.

The expected discounted value of the bond equals $(1.05)^{-1}(0.99 \cdot 1 + 0.01 \cdot 0.6) = 0.949 > p_1(0, 1)$. We see that in this example the price $p_1(0, 1)$ is smaller than the expected discounted value of the claim. This is the typical situation in real markets for corporate bonds, as investors demand a premium for bearing the default risk of the bond.

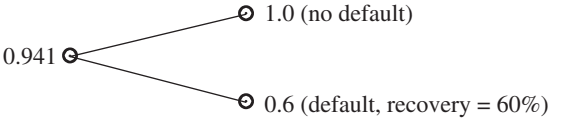

Figure 2.1. Evolution of the price $p_1(\cdot, 1)$ of a defaultable bond in the basic one-period default model; the probabilities of the upper and lower branches are 0.99 and 0.01, respectively.

In a one-period model, an equivalent martingale measure or risk-neutral measure is simply a new probability measure Q such that for every traded security the Q-expectation of the discounted pay-off equals the current price of the security, so that investing in this security becomes a fair bet. In more general situations (for example, in continuous-time models), the idea of a fair bet is formalized by the requirement that the discounted price process of a traded security is a so-called Q-martingale (hence the name martingale measure). In the basic one-period default model, Q is thus given in terms of an artificial default probability q such that

$$p_1(0, 1) = (1.05)^{-1}((1 - q) \cdot 1 + q \cdot 0.6).$$

Clearly, q is uniquely determined by this equation and we get that $q = 0.03$. Note that, in our example, q is bigger than the physical default probability $p = 0.01$; again, this is typical for real markets and reflects the risk premium demanded by buyers of defaultable bonds. The example also shows that different approaches are needed in order to determine the historical default probability p and the risk-neutral default probability q: the former is *estimated* from historical data such as the default history of firms of similar credit quality (see, for example, Sections 10.3.3 and 11.5), whereas q is *calibrated* to market prices of traded securities.

Under the risk-neutral pricing approach, the price of a security is computed as the (conditional) expected value of the discounted future cash flows, where expectation is taken with respect to the risk-neutral measure Q. Denoting the pay-off of the security at $t = 1$ by the rv H and the risk-free simple interest rate between time 0 and time 1 by $r_{0,1} \geqslant 0$, we obtain the following formula for the value V_0^H of the claim H at $t = 0$:

$$V_0^H = E^Q\left(\frac{H}{1 + r_{0,1}}\right). \tag{2.10}$$

For a specific example in the basic one-period default model, consider a *default put option* that pays one unit at $t = 1$ if the bond defaults and zero otherwise; the option can be thought of as a simplified version of a credit default swap. Using risk-neutral pricing, the value of the option at $t = 0$ is given by

$$V_0 = (1.05)^{-1}((1 - q) \cdot 0 + q \cdot 1) = (1.05)^{-1}0.03 = 0.0285.$$

In continuous-time models one usually uses continuous compounding, and (2.10) is therefore replaced by the slightly more general expression

$$V_t^H = E_t^Q(\mathrm{e}^{-r(T-t)}H), \quad t < T. \tag{2.11}$$

Here, T is the maturity date of the security and the subscript t on the expectation operator indicates that the expectation is taken with respect to the information available to investors at time t, as will be explained in more detail in Chapter 10.

Formulas (2.10) and (2.11) are known as *risk-neutral pricing rules*. Risk-neutral pricing applied to non-traded financial products is a typical example of Level 2 valuation: prices of traded securities are used to calibrate model parameters under the risk-neutral measure Q; this measure is then used to price the non-traded products. We give one example that underscores our use of the Black–Scholes pricing rule in Example 2.2.

Example 2.5 (European call option in Black–Scholes model). Consider again the European call option in Example 2.2 and suppose that options with our desired strike K and/or maturity time T are not traded, but that other options on the same stock are traded. We assume that under the real-world probability measure P the stock price (S_t) follows a geometric Brownian motion model (the so-called Black–Scholes model) with dynamics given by

$$\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}W_t$$

for constants $\mu \in \mathbb{R}$ (the drift) and $\sigma > 0$ (the volatility), and a standard Brownian motion (W_t). It is well known that there is an equivalent martingale measure Q under which the discounted stock price $(\mathrm{e}^{-rt} S_t)$ is a martingale; under Q, the stock price follows a geometric Brownian motion model with drift r and volatility σ. The European call option pay-off is $H = (S_T - K)^+$ and the risk-neutral valuation formula in (2.11) may be shown to take the form

$$V_t = E_t^Q(\mathrm{e}^{-r(T-t)}(S_T - K)^+) = C^{\mathrm{BS}}(t, S_t; r, \sigma, K, T), \quad t < T, \tag{2.12}$$

with C^{BS} as in Example 2.2. To assign a risk-neutral value to the call option at time t (knowing the current price of the stock S_t, the interest rate r and the option characteristics K and T), we need to calibrate the model parameter σ. As discussed above, we would typically use quoted prices $C^{\mathrm{BS}}(t, S_t; r, \sigma, K^*, T^*)$ for options on the stock with different characteristics to infer a value for σ and then plug the so-called implied volatility into (2.12).

There are two theoretical justifications for risk-neutral pricing. First, a standard result of mathematical finance (the so-called *first fundamental theorem of asset pricing*) states that a model for security prices is arbitrage free if and only if it admits at least one equivalent martingale measure Q. Hence, if a financial product is to be priced in accordance with no-arbitrage principles, its price must be given by the risk-neutral pricing formula for some risk-neutral measure Q. A second justification refers to hedging: in financial models it is often possible to replicate the pay-off of a financial product by trading in the assets, a practice known as *(dynamic) hedging*, and it is well known that in a frictionless market the cost of carrying out such a hedge is given by the risk-neutral pricing rule. Advantages and limitations of risk-neutral pricing will be discussed in more detail in Section 10.4.2.

2.2.3 Loss Distributions

Having mapped the risks of a portfolio, we now consider how to derive loss distributions with a view to using them in risk-management applications such as capital setting. Assuming the current time is t and recalling formula (2.3) for the loss over the time period $[t, t+1]$,

$$L_{t+1} = -\Delta V_{t+1} = -(f(t+1, \boldsymbol{z}_t + \boldsymbol{X}_{t+1}) - f(t, \boldsymbol{z}_t)),$$

we see that in order to determine the loss distribution (i.e. the distribution of L_{t+1}) we need to do two things: (i) specify a model for the risk-factor changes $\boldsymbol{X}_{t+1}$; and (ii) determine the distribution of the rv $f(t+1, \boldsymbol{z}_t + \boldsymbol{X}_{t+1})$.

Note that effectively two kinds of model enter into this process. The models used in (i) are *projection models* used to forecast the behaviour of risk factors in the real world, and they are generally estimated from empirical data describing past risk-factor changes $(\boldsymbol{X}_s)_{s \leqslant t}$. Depending on the complexity of the positions involved, the mapping function f in (ii) will typically also embody *valuation models*; consider in this context the use of the Black–Scholes model to value a European call option, as described in Examples 2.2 and 2.5.

Broadly speaking, there are three kinds of method that can be used to address these challenges: an analytical method, a method based on the idea of historical simulation, or a simulation approach (also known as a Monte Carlo method).

Analytical method. In an analytical method we attempt to choose a model for $\boldsymbol{X}_{t+1}$ and a mapping function f in such a way that the distribution of L_{t+1} can be determined analytically. A prime example of this approach is the so-called variance–covariance method for market-risk management, which dates back to the early work of the RiskMetrics Group (JPMorgan 1996). In the variance–covariance method the risk-factor changes $\boldsymbol{X}_{t+1}$ are assumed to follow a multivariate normal distribution, denoted by $\boldsymbol{X}_{t+1} \sim N_d(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu}$ is the mean vector and Σ the covariance (or variance–covariance) matrix of the distribution. This would follow, for example, from assuming that the risk factors $\boldsymbol{Z}_t$ evolve in continuous time according to a multivariate Brownian motion. The properties of the multivariate normal distribution are discussed in detail in Section 6.1.3.

We also assume that the linearized loss in terms of the risk factors is a sufficiently accurate approximation of the actual loss and simplify the problem by considering the distribution of L^{Δ}_{t+1} defined in (2.4). The linearized loss will have general structure

$$L^{\Delta}_{t+1} = -(c_t + \boldsymbol{b}_t' \boldsymbol{X}_{t+1}) \tag{2.13}$$

for some constant c_t and constant vector $\boldsymbol{b}_t$, which are known to us at time t. For a concrete example, consider the stock portfolio of Example 2.1, where the loss takes the form $L^{\Delta}_{t+1} = -v_t \boldsymbol{w}_t' \boldsymbol{X}_{t+1}$ and $\boldsymbol{w}_t$ is the vector of portfolio weights at time t.

An important property of the multivariate normal distribution is that a linear function (2.13) of $\boldsymbol{X}_{t+1}$ must have a univariate normal distribution. From general rules for calculating the mean and variance of linear combinations of a random

vector we obtain that

$$L_{t+1}^{\Delta} \sim N(-c_t - \boldsymbol{b}_t'\boldsymbol{\mu}, \boldsymbol{b}_t'\Sigma\boldsymbol{b}_t). \tag{2.14}$$

The variance–covariance method offers a simple solution to the risk-measurement problem, but this convenience is achieved at the cost of two crude simplifying assumptions. First, linearization may not always offer a good approximation of the relationship between the true loss distribution and the risk-factor changes. Second, the assumption of normality is unlikely to be realistic for the distribution of the risk-factor changes, certainly for daily data and probably also for weekly and even monthly data. A stylized fact of empirical finance suggests that the distribution of financial risk-factor returns is leptokurtic and heavier tailed than the Gaussian distribution. In Section 3.1.2 we will present evidence for this observation in an analysis of daily, weekly, monthly and quarterly stock returns. The implication is that an assumption of Gaussian risk factors will tend to underestimate the tail of the loss distribution and thus underestimate the risk of the portfolio.

Remark 2.6. Note that we postpone a detailed discussion of how the model parameters $\boldsymbol{\mu}$ and Σ are estimated from historical risk-factor changes $(\boldsymbol{X}_s)_{s\leqslant t}$ until later chapters. We should, however, point out that when a dynamic model for $\boldsymbol{X}_{t+1}$ is considered, different estimation methods are possible depending on whether we focus on the conditional distribution of $\boldsymbol{X}_{t+1}$ given past values of the process or whether we consider the equilibrium distribution in a stationary model. These different approaches are said to constitute conditional and unconditional methods of computing the loss distribution—an issue we deal with in much more detail in Chapter 9.

Historical simulation. Instead of estimating the distribution of L_{t+1} in some explicit parametric model for $\boldsymbol{X}_{t+1}$, the historical-simulation method can be thought of as estimating the distribution of the loss using the *empirical distribution* of past risk-factor changes. Suppose we collect historical risk-factor change data over n time periods and denote these data by $\boldsymbol{X}_{t-n+1}, \ldots, \boldsymbol{X}_t$. In historical simulation we construct the following univariate data set of imaginary losses:

$$\{\tilde{L}_s = -(f(t+1, \boldsymbol{z}_t + \boldsymbol{X}_s) - f(t, \boldsymbol{z}_t)) \colon s = t-n+1, \ldots, t\}.$$

The values $\tilde{L}_s$ show what would happen to the current portfolio if the risk-factor changes in period s were to recur. If we assume that the process of risk-factor changes is stationary with df $F_{\boldsymbol{X}}$, then (subject to further technical conditions) the empirical df of the historically simulated losses is a consistent estimator of the loss distribution. Estimators for any statistic of the loss distribution—such as the expected loss, the variance of the loss, or the value-at-risk (see Section 2.3.2 for a definition)—can be computed from the empirical df of the historically simulated losses. For instance, the expected loss can be estimated by $E(L_{t+1}) \approx n^{-1} \sum_{s=t-n+1}^{t} \tilde{L}_s$, and techniques like empirical quantile estimation can be used to derive estimates of value-at-risk. Further details can be found in Chapter 9.

The historical-simulation method has obvious attractions: it is easy to implement and it reduces the loss distribution calculation problem to a one-dimensional problem. However, the success of the approach is dependent on our ability to collect sufficient quantities of relevant synchronized historical data for all risk factors. As such, the method is mainly used in market-risk management for banks, where the issue of data availability is less of a problem (due to the relatively short risk-management time horizon).

Monte Carlo method. Any approach to risk measurement that involves the simulation of an explicit parametric model for risk-factor changes is known as a Monte Carlo method. The method does not solve the problem of finding a multivariate model for $\boldsymbol{X}_{t+1}$, and any results that are obtained will only be as good as the model that is used. For large portfolios the computational cost of the Monte Carlo approach can be considerable, as every simulation requires the revaluation of the portfolio. This is particularly problematic if the portfolio contains many derivatives that cannot be priced in closed form. Such derivative positions might have to be valued using Monte Carlo approximation techniques, which are also based on simulations. This leads to situations where Monte Carlo procedures are *nested* and simulations are being generated within simulations, which can be very slow.

Simulation techniques are frequently used in the management of credit portfolios (see, for example, Section 11.4). So-called *economic scenario generation* models, which are used in insurance, also fall under the heading of Monte Carlo methods. These are economically motivated and (typically) dynamic models for the evolution and interaction of different risk factors, and they can be used to generate realizations of $\boldsymbol{X}_{t+1}$.

Notes and Comments

The concept of mapping portfolio values to fundamental risk factors was pioneered by the RiskMetrics Group: see the RiskMetrics Technical Document (JPMorgan 1996) and Mina and Xiao (2001). We explore the topic in more detail, with further examples, in Chapter 9. Other textbooks that treat the mapping of positions include Dowd (1998), Jorion (2007) and Volume III of *Market Risk Analysis* by Alexander (2009). The use of first-order approximations to the portfolio value (the so-called delta approximation) may be found in Duffie and Pan (1997); for second-order approximations, see Section 9.1.2.

More details of the Black–Scholes valuation formula used in Example 2.2 may be found in many texts on options and derivatives, such as Haug (1998), Wilmot (2000) and Hull (2014). For annuity products similar to the one analysed in Example 2.4 and other standard life insurance products, good references are Hardy (2003), Møller and Steffensen (2007), Koller (2011), Dickson, Hardy and Waters (2013) and the classic book by Gerber (1997).

The best resource for more on International Financial Reporting Standard 7 and the fair-value accounting of financial instruments is the International Financial Reporting Standards website at www.ifrs.org. Shaffer (2011) considers the impact of fair-value accounting on financial institutions, while Laux and Leuz (2010) address

the issue of whether fair-value accounting may have contributed to the financial crisis.

Market-consistent actuarial valuation in insurance is the subject of a textbook by Wüthrich, Bühlmann and Furrer (2010) (see also Wüthrich and Merz 2013). The fundamental theorem of asset pricing and the conceptual underpinnings of risk-neutral pricing are discussed in most textbooks on mathematical finance: see, for example, Björk (2004) or Shreve (2004b).

The analytical method for deriving loss distributions based on an assumption of normal risk-factor changes belongs to the original RiskMetrics methodology cited above. For the analysis of the distribution of losses in a bank's trading book, this has largely been supplanted by the use of historical simulation (Pérignon and Smith 2010). Monte Carlo approaches (economic scenario generators) are widely used in internal models for Solvency II in the insurance industry (see Varnell 2011).

2.3 Risk Measurement

In very general terms a risk measure associates a financial position with loss L with a real number that measures the "riskiness of L". In practice, risk measures are used for a variety of purposes. To begin with, they are used to determine the amount of capital a financial institution needs to hold as a buffer against unexpected future losses on its portfolio in order to satisfy a regulator who is concerned with the solvency of the institution. Similarly, they are used to determine appropriate margin requirements for investors trading at an organized exchange. Moreover, risk measures are often used by management as a tool for limiting the amount of risk a business unit within a firm may take. For instance, traders in a bank might be constrained by the rule that the daily 95% value-at-risk of their position should not exceed a given bound.

In Section 2.3.1 we give an overview of some different approaches to measuring risk before focusing on risk measures that are derived from loss distributions. We introduce the widely used value-at-risk measure in Section 2.3.2 and explain how VaR features in risk capital calculations in Section 2.3.3. In Section 2.3.4 alternative risk measures derived from loss distributions are presented, and in Section 2.3.5 an introduction to the subject of desirable risk measure properties is given, in which the notions of coherent and convex risk measures are defined and examples are discussed.

2.3.1 Approaches to Risk Measurement

Existing approaches to measuring the risk of a financial position can be grouped into three categories: the notional-amount approach, risk measures based on loss distributions, and risk measures based on scenarios.

Notional-amount approach. This is the oldest approach to quantifying the risk of a portfolio of risky assets. In the notional-amount approach the risk of a portfolio is defined as the sum of the notional values of the individual securities in the portfolio, where each notional value may be weighted by a factor representing an assessment

of the riskiness of the broad asset class to which the security or instrument belongs. An example of this approach is the so-called *standardized approach* in the Basel regulatory framework (see Section 1.3.1 for a general description and Section 13.1.2 for the standardized approach as it applies to operational risk).

The advantage of the notional-amount approach is its apparent simplicity. However, as noted in Section 1.3.1, the approach is flawed from an economic viewpoint for a number of reasons. To begin with, the approach does not differentiate between long and short positions and there is no netting. For instance, the risk of a long position in corporate bonds hedged by an offsetting position in credit default swaps would be counted as twice the risk of the unhedged bond position. Moreover, the approach does not reflect the benefits of diversification on the overall risk of the portfolio. For example, if we use the notional-amount approach, a well-diversified credit portfolio consisting of loans to many companies appears to have the same risk as a portfolio in which the whole amount is lent to a single company. Finally, the notional-amount approach has problems in dealing with portfolios of derivatives, where the notional amount of the underlying and the economic value of the derivative position can differ widely.

Risk measures based on loss distributions. Most modern measures of the risk in a portfolio are statistical quantities describing the conditional or unconditional loss distribution of the portfolio over some predetermined horizon Δt. Examples include the variance, the VaR and the ES risk measures, which we discuss in more detail later in this chapter. Risk measures based on loss distributions have a number of advantages. The concept of a loss distribution makes sense on all levels of aggregation, from a portfolio consisting of a single instrument to the overall position of a financial institution. Moreover, if estimated properly, the loss distribution reflects netting and diversification effects.

Two issues should be borne in mind when working with loss distributions. First, any estimate of the loss distribution is based on past data. If the laws governing financial markets change, these past data are of limited use in predicting future risk. Second, even in a stationary environment it is difficult to estimate the loss distribution accurately, particularly for large portfolios. Many seemingly sophisticated risk-management systems are based on relatively crude statistical models for the loss distribution (incorporating, for example, untenable assumptions of normality). These issues call for continual improvements in the way that loss distributions are estimated and, of course, for prudence in the practical application of risk-management models based on estimated loss distributions. In particular, risk measures based on the loss distribution should be complemented by information from hypothetical scenarios. Moreover, forward-looking information reflecting the expectations of market participants, such as implied volatilities, should be used in conjunction with statistical estimates (which are necessarily based on past information) in calibrating models of the loss distribution.

Scenario-based risk measures. In the scenario-based approach to measuring the risk of a portfolio, one considers a number of possible future risk-factor changes

(scenarios), such as a 10% rise in key exchange rates, a simultaneous 20% drop in major stock market indices or a simultaneous rise in key interest rates around the globe. The risk of the portfolio is then measured as the maximum loss of the portfolio under all scenarios. The scenarios can also be weighted for plausibility. This approach to risk measurement is the one that is typically adopted in stress testing.

We now give a formal description. Fix a set $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ of risk-factor changes (the scenarios) and a vector $\boldsymbol{w} = (w_1, \ldots, w_n)' \in [0,1]^n$ of weights. Denote by $L(\boldsymbol{x})$ the loss the portfolio would suffer if the hypothetical scenario $\boldsymbol{x}$ were to occur. Using the notation of Section 2.2.1 we get

$$L(\boldsymbol{x}) := -(f(t+1, z_t + \boldsymbol{x}) - f(t, z_t)), \quad \boldsymbol{x} \in \mathbb{R}^d.$$

The risk of the portfolio is then measured by

$$\psi_{[\mathcal{X}, \boldsymbol{w}]} := \max\{w_1 L(\boldsymbol{x}_1), \ldots, w_n L(\boldsymbol{x}_n)\}. \tag{2.15}$$

Many risk measures that are used in practice are of the form (2.15). The following is a simplified description of a system for determining margin requirements developed by the Chicago Mercantile Exchange (see Chicago Mercantile Exchange 2010). To compute the initial margin for a simple portfolio consisting of a position in a futures contract and call and put options on this contract, sixteen different scenarios are considered. The first fourteen consist of an up move or a down move of volatility combined with no move, an up move or a down move of the futures price by $\frac{1}{3}$, $\frac{2}{3}$ or $\frac{3}{3}$ of a unit of a specified range. The weights w_i, $i = 1, \ldots, 14$, of these scenarios are equal to 1. In addition, there are two extreme scenarios with weights $w_{15} = w_{16} = 0.35$. The amount of capital required by the exchange as margin for the portfolio is then computed according to (2.15).

Remark 2.7. We can give a slightly different mathematical interpretation to formula (2.15), which will be useful in Section 2.3.5. Assume for the moment that $L(\mathbf{0}) = 0$, i.e. that the value of the position is unchanged if all risk factors stay the same. This is reasonable, at least for a short risk-management horizon Δt. In that case, the expression $w_i L(\boldsymbol{x}_i)$ can be viewed as the expected value of L under a probability measure on the space of risk-factor changes; this measure associates a mass of $w_i \in [0,1]$ to the point $\boldsymbol{x}_i$ and a mass of $1 - w_i$ to the point $\mathbf{0}$. Denote by $\delta_{\boldsymbol{x}}$ the probability measure associating a mass of one to the point $\boldsymbol{x} \in \mathbb{R}^d$ and by $\mathcal{P}_{[\mathcal{X}, \boldsymbol{w}]}$ the following set of probability measures on $\mathbb{R}^d$:

$$\mathcal{P}_{[\mathcal{X}, \boldsymbol{w}]} = \{w_1 \delta_{\boldsymbol{x}_1} + (1 - w_1)\delta_{\mathbf{0}}, \ldots, w_n \delta_{\boldsymbol{x}_n} + (1 - w_n)\delta_{\mathbf{0}}\}.$$

Then $\psi_{[\mathcal{X}, \boldsymbol{w}]}$ can be written as

$$\psi_{[\mathcal{X}, \boldsymbol{w}]} = \max\{E^P(L(\boldsymbol{X})) \colon P \in \mathcal{P}_{[\mathcal{X}, \boldsymbol{w}]}\}. \tag{2.16}$$

A risk measure of the form (2.16), where $\mathcal{P}_{[\mathcal{X}, \boldsymbol{w}]}$ is replaced by some arbitrary subset $\mathcal{P}$ of the set of all probability measures on the space of risk-factor changes, is termed a *generalized scenario*. Generalized scenarios play an important role in the theory of coherent risk measures (see Section 8.1).

Scenario-based risk measures are a very useful risk-management tool for portfolios exposed to a relatively small set of risk factors, as in the Chicago Mercantile Exchange example. Moreover, they provide useful complementary information to measures based on statistics of the loss distribution. The main problem in setting up a scenario-based risk measure is, of course, determining an appropriate set of scenarios and weighting factors.

2.3.2 Value-at-Risk

VaR is probably the most widely used risk measure in financial institutions. It has a prominent role in the Basel regulatory framework and has also been influential in Solvency II.

Consider a portfolio of risky assets and a fixed time horizon Δt, and denote by $F_L(l) = P(L \leqslant l)$ the df of the corresponding loss distribution. We want to define a statistic based on F_L that measures the severity of the risk of holding our portfolio over the time period Δt. An obvious candidate is the maximum possible loss, given by $\inf\{l \in \mathbb{R}\colon F_L(l) = 1\}$. However, for most distributions of interest, the maximum loss is infinity. Moreover, by using the maximum loss, any probability information in F_L is neglected. The idea in the definition of VaR is to replace "maximum loss" by "maximum loss that is not exceeded with a given high probability".

Definition 2.8 (value-at-risk). Given some confidence level $\alpha \in (0, 1)$, the VaR of a portfolio with loss L at the confidence level α is given by the smallest number l such that the probability that the loss L exceeds l is no larger than $1 - \alpha$. Formally,

$$\mathrm{VaR}_\alpha = \mathrm{VaR}_\alpha(L) = \inf\{l \in \mathbb{R}\colon P(L > l) \leqslant 1 - \alpha\} = \inf\{l \in \mathbb{R}\colon F_L(l) \geqslant \alpha\}. \tag{2.17}$$

In probabilistic terms, VaR is therefore simply a *quantile* of the loss distribution. Typical values for α are $\alpha = 0.95$ or $\alpha = 0.99$; in market-risk management, the time horizon Δt is usually one or ten days, while in credit risk management and operational risk management, Δt is usually one year. Note that by its very definition the VaR at confidence level α does not give any information about the severity of losses that occur with a probability of less than $1 - \alpha$. This is clearly a drawback of VaR as a risk measure. For a small case study that illustrates this problem numerically we refer to Example 2.16 below.

Figure 2.2 illustrates the notion of VaR. The probability density function of a loss distribution is shown with a vertical line at the value of the 95% VaR. Note that the mean loss is negative ($E(L) = -2.6$), indicating that we expect to make a profit, but the right tail of the loss distribution is quite long in comparison with the left tail. The 95% VaR value is approximately 2.2, indicating that there is a 5% chance that we lose at least this amount.

Since quantiles play an important role in risk management, we recall the precise definition.

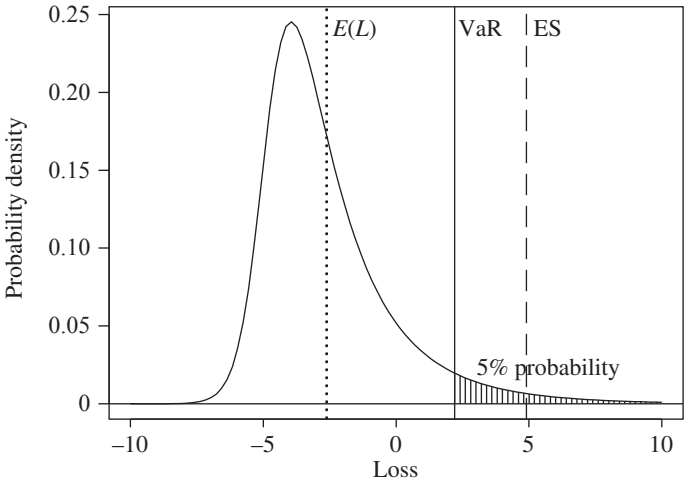

Figure 2.2. An example of a loss distribution with the 95% VaR marked as a vertical line; the mean loss is shown with a dotted line and an alternative risk measure known as the 95% ES (see Section 2.3.4 and Definition 2.12) is marked with a dashed line.

Definition 2.9 (the generalized inverse and the quantile function).

(i) Given some increasing function $T: \mathbb{R} \to \mathbb{R}$, the *generalized inverse* of T is defined by $T^{\leftarrow}(y) := \inf\{x \in \mathbb{R}: T(x) \geqslant y\}$, where we use the convention that the infimum of an empty set is ∞.

(ii) Given some df F, the generalized inverse $F^{\leftarrow}$ is called the *quantile function* of F. For $\alpha \in (0, 1)$ the α-quantile of F is given by

$$q_\alpha(F) := F^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R}: F(x) \geqslant \alpha\}.$$

For an rv X with df F we often use the alternative notation $q_\alpha(X) := q_\alpha(F)$. If F is continuous and strictly increasing, we simply have $q_\alpha(F) = F^{-1}(\alpha)$, where F^{-1} is the ordinary inverse of F. To compute quantiles in more general cases we may use the following simple criterion.

Lemma 2.10. *A point $x_0 \in \mathbb{R}$ is the α-quantile of some df F if and only if the following two conditions are satisfied: $F(x_0) \geqslant \alpha$; and $F(x) < \alpha$ for all $x < x_0$.*

The lemma follows immediately from the definition of the generalized inverse and the right-continuity of F. Examples of the computation of quantiles in certain tricky cases and further properties of generalized inverses are given in Section A.1.2.

Example 2.11 (VaR for normal and t loss distributions). Suppose that the loss distribution F_L is normal with mean μ and variance σ^2. Fix $\alpha \in (0, 1)$. Then

$$\mathrm{VaR}_\alpha = \mu + \sigma\Phi^{-1}(\alpha), \tag{2.18}$$

where Φ denotes the standard normal df and $\Phi^{-1}(\alpha)$ is the α-quantile of Φ. The proof is easy: since F_L is strictly increasing, by Lemma 2.10 we only have to show

that $F_L(\text{VaR}_\alpha) = \alpha$. Now,

$$P(L \leqslant \text{VaR}_\alpha) = P\left(\frac{L-\mu}{\sigma} \leqslant \Phi^{-1}(\alpha)\right) = \Phi(\Phi^{-1}(\alpha)) = \alpha.$$

This result is routinely used in the *variance–covariance* approach (also known as the delta-normal approach) to computing risk measures.

Of course, a similar result is obtained for any location-scale family, and another useful example is the Student t loss distribution. Suppose that our loss L is such that $(L-\mu)/\sigma$ has a standard t distribution with ν degrees of freedom; we denote this loss distribution by $L \sim t(\nu, \mu, \sigma^2)$ and note that the moments are given by $E(L) = \mu$ and $\text{var}(L) = \nu\sigma^2/(\nu-2)$ when $\nu > 2$, so σ is not the standard deviation of the distribution. We get

$$\text{VaR}_\alpha = \mu + \sigma t_\nu^{-1}(\alpha), \tag{2.19}$$

where t_ν denotes the df of a standard t distribution, which is available in most statistical computer packages along with its inverse.

In the remainder of this section we discuss a number of further issues relating to the use of VaR as a risk measure in practice.

Choice of VaR parameters. In working with VaR the parameters Δt and α need to be chosen. There is of course no single optimal value for these parameters, but there are some considerations that might influence the choice of regulators or internal-model builders.

The risk-management horizon Δt should reflect the time period over which a financial institution is committed to hold its portfolio, which will be affected by contractual and legal constraints as well as liquidity considerations. In choosing a horizon for enterprise-wide risk management, a financial institution has little choice but to use the horizon appropriate for the market in which its core business activities lie. For example, insurance companies are typically bound to hold their portfolio of liabilities for one year, during which time they are not able to alter the portfolio or renegotiate the premiums they receive; one year is therefore an appropriate time horizon for measuring the risk in the liability and asset portfolios of an insurer. Moreover, a financial institution can be forced to hold a loss-making position in a risky asset if the market for that asset is not very liquid, so a relatively long horizon may be appropriate for illiquid assets.

There are other, more practical, considerations that suggest that Δt should be relatively small. The assumption that the composition of the portfolio remains unchanged is tenable only for a short holding period. Moreover, the calibration and testing of statistical models for historical risk-factor changes $(\boldsymbol{X}_t)$ are easier if Δt is small, since this typically means that we have more data at our disposal.

For the confidence level α, different values are also appropriate for different purposes. In order to set limits for traders, a bank would typically take α to be 95% and Δt to be one day. For capital adequacy purposes higher confidence levels are generally used. For instance, the Basel capital charges for market risk in the trading

book of a bank are based on the use of VaR at the 99% level and a ten-day horizon. The Solvency II framework uses a value of α equal to 0.995 and a one-year horizon. On the other hand, the backtesting of models that produce VaR figures often needs to be carried out at lower confidence levels using shorter horizons in order to have sufficient statistical power to detect poor model performance.

Model risk and market liquidity. In practice, VaR numbers are sometimes given a very literal interpretation; the statement that the daily VaR at confidence level $\alpha = 99\%$ for a particular portfolio is equal to l is understood to mean that "there is a probability of exactly 1% that the loss on this position will be larger than l". This interpretation is misleading because it neglects estimation error, model risk and market liquidity risk.

We recall that model risk is the risk that our model for the loss distribution is misspecified. For instance, we might work with a normal distribution to model losses, whereas the true distribution is heavy tailed, or we might fail to recognize the presence of volatility clustering or tail dependence (see Chapter 3) in modelling the distribution of the risk-factor changes underlying the losses. Of course, these problems are most pronounced if we are trying to estimate VaR at very high confidence levels. Liquidity risk refers to the fact that any attempt to liquidate a large loss-making position is likely to move the price against us, thus exacerbating the loss.

2.3.3 VaR in Risk Capital Calculations

Quantile-based risk measures are used in many risk capital calculations in practice. In this section we give two examples.

VaR in regulatory capital calculations for the trading book. The VaR risk measure is applied to calculate a number of regulatory capital charges for the trading book of a bank. Under the internal-model approach a bank calculates a daily VaR measure for the distribution of possible ten-day trading book losses based on recent data on risk-factor changes under the assumption that the trading book portfolio is held fixed over this time period. We describe the statistical methodology that is typically used for this calculation in Section 9.2.

While exact details may vary from one national regulator to another, the basic capital charge on day t is usually calculated according to a formula of the form

$$\mathrm{RC}^t = \max\left\{ \mathrm{VaR}_{0.99}^{t,10}, \frac{k}{60}\sum_{i=1}^{60} \mathrm{VaR}_{0.99}^{t-i+1,10} \right\}, \tag{2.20}$$

where $\mathrm{VaR}_{0.99}^{j,10}$ stands for the ten-day VaR at the 99% confidence level, calculated on day j, and where k is a multiplier in the range 3–4 that is determined by the regulator as a function of the overall quality of the bank's internal model. The averaging of the last sixty daily VaR numbers obviously tends to lead to smooth changes in the capital charge over time unless the most recent number $\mathrm{VaR}_{0.99}^{t,10}$ is particularly large.

A number of additional capital charges are added to RC^t. These include a stressed VaR charge and an incremental risk charge, as well as a number of charges that are

designed to take into account so-called specific risks due to idiosyncratic price movements in certain instruments that are not explained by general market-risk factors. The stressed VaR charge is calculated using similar VaR methodology to the standard charge but with data taken from a historical window in which markets were particularly volatile. The incremental risk charge is an estimate of the 99.9% quantile of the distribution of annual losses due to defaults and downgrades for credit-risky instruments in the trading book (excluding securitizations).

The solvency capital requirement in Solvency II. An informal definition of the solvency capital requirement is "the level of capital that enables the insurer to meet its obligations over a one-year time horizon with a high confidence level (99.5%)" (this is taken from a 2007 factsheet produced by De Nederlandsche Bank). We will give an argument that leads to the use of a VaR-based risk measure.

Consider the balance sheet of the insurer in Table 2.2 and assume that the current equity capital is given by $V_t = A_t - B_t$, i.e. the difference between the value of assets and the value of liabilities, or the net asset value; this is also referred to under Solvency II as *own funds*. The liabilities B_t are considered to include all technical provisions computed in a market-consistent way, including risk margins for non-hedgeable risks where necessary.

The insurer wants to ensure that it is solvent in one year's time with high probability α. It considers the possibility that it may need to raise extra capital and makes the following thought calculation. Given its current balance sheet and business model it attempts to determine the minimum amount of extra capital x_0 that it would have to raise now at time t and place in a risk-free investment in order to be solvent in one year's time with probability α. In mathematical notation it needs to determine

$$x_0 = \inf\{x\colon P(V_{t+1} + x(1 + r_{t,1}) \geqslant 0) = \alpha\},$$

where $r_{t,1}$ is the simple risk-free rate for a one-year investment and V_{t+1} is the net asset value in one year's time. If x_0 is negative, then the company is well capitalized at time t and money could be taken out of the company.

An easy calculation gives

$$\begin{aligned} x_0 &= \inf\{x\colon P(-V_{t+1} \leqslant x(1 + r_{t,1})) = \alpha\} \\ &= \inf\{x\colon P(V_t - V_{t+1}/(1 + r_{t,1}) \leqslant x + V_t) = \alpha\}, \end{aligned}$$

which shows that

$$V_t + x_0 = q_\alpha(V_t - V_{t+1}/(1 + r_{t,1})).$$

The sum $V_t + x_0$ gives the solvency capital requirement: namely, the available capital corrected by the amount x_0. Hence, we see that the solvency capital requirement is a quantile of the distribution of $V_t - V_{t+1}/(1 + r_{t,1})$, a loss distribution that takes into account the time value of money through discounting, as discussed in Section 2.2.1. For a well-capitalized company with $x_0 < 0$, the amount $-x_0 = V_t - q_\alpha(V_t - V_{t+1}/(1 + r_{t,1}))$ (own funds minus the solvency capital requirement) is called the excess capital.

2.3.4 Other Risk Measures Based on Loss Distributions

In this section we provide short notes on a number of other statistical summaries of the loss distribution that are frequently used as risk measures in financial and insurance risk management.

Variance. Historically, the variance of the P&L distribution (or, equivalently, the standard deviation) has been the dominating risk measure in finance. To a large extent this is due to the huge impact that the portfolio theory of Markowitz, which uses variance as a measure of risk, has had on theory and practice in finance (see, for example, Markowitz 1952). Variance is a well-understood concept that is easy to use analytically. However, as a risk measure it has two drawbacks. On the technical side, if we want to work with variance, we have to assume that the second moment of the loss distribution exists. While unproblematic for most return distributions in finance, this can cause problems in certain areas of non-life insurance or for the analysis of operational losses (see Section 13.1.4). On the conceptual side, since it makes no distinction between positive and negative deviations from the mean, variance is a good measure of risk only for distributions that are (approximately) symmetric around the mean, such as the normal distribution or a (finite-variance) Student t distribution. However, in many areas of risk management, such as in credit and operational risk management, we deal with loss distributions that are highly skewed.

Lower and upper partial moments. Partial moments are measures of risk based on the lower or upper part of a distribution. In most of the literature on risk management the main concern is with the risk inherent in the lower tail of a P&L distribution, and lower partial moments are used to measure this risk. Under our sign convention we are concerned with the risk inherent in the upper tail of a loss distribution, so we focus on upper partial moments. Given an exponent $k \geqslant 0$ and a reference point q, the upper partial moment $\mathrm{UPM}(k, q)$ is defined as

$$\mathrm{UPM}(k, q) = \int_q^\infty (l - q)^k \, \mathrm{d}F_L(l) \in [0, \infty]. \tag{2.21}$$

Some combinations of k and q have a special interpretation: for $k = 0$ we obtain $P(L \geqslant q)$; for $k = 1$ we obtain $E((L - q)I_{\{L \geqslant q\}})$; for $k = 2$ and $q = E(L)$ we obtain the *upper semivariance* of L. Of course, the higher the value we choose for k, the more conservative our risk measure becomes, since we give more and more weight to large deviations from the reference point q.

Expected shortfall. ES is closely related to VaR and there is an ongoing debate in the risk-management community on the strengths and weaknesses of both risk measures.

Definition 2.12 (expected shortfall). For a loss L with $E(|L|) < \infty$ and df F_L, the ES at confidence level $\alpha \in (0, 1)$ is defined as

$$\mathrm{ES}_\alpha = \frac{1}{1 - \alpha} \int_\alpha^1 q_u(F_L) \, \mathrm{d}u, \tag{2.22}$$

where $q_u(F_L) = F_L^{\leftarrow}(u)$ is the quantile function of F_L.

The condition $E(|L|) < \infty$ ensures that the integral in (2.22) is well defined. By definition, ES is related to VaR by

$$\mathrm{ES}_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)\,\mathrm{d}u.$$

Instead of fixing a particular confidence level α, we average VaR over all levels $u \geqslant \alpha$ and thus "look further into the tail" of the loss distribution. Obviously, ES_α depends only on the distribution of L, and $\mathrm{ES}_\alpha \geqslant \mathrm{VaR}_\alpha$. See Figure 2.2 for a simple illustration of an ES value and its relationship to VaR. The 95% ES value of 4.9 is at least double the 95% VaR value of 2.2 in this case.

For continuous loss distributions an even more intuitive expression can be derived that shows that ES can be interpreted as the expected loss that is incurred in the event that VaR is exceeded.

Lemma 2.13. *For an integrable loss L with continuous df F_L and for any $\alpha \in (0,1)$ we have*

$$\mathrm{ES}_\alpha = \frac{E(L; L \geqslant q_\alpha(L))}{1-\alpha} = E(L \mid L \geqslant \mathrm{VaR}_\alpha), \tag{2.23}$$

where we have used the notation $E(X; A) := E(XI_A)$ for a generic integrable rv X and a generic set $A \in \mathcal{F}$.

Proof. Denote by U an rv with uniform distribution on the interval $[0,1]$. It is a well-known fact from elementary probability theory that the rv $F_L^{\leftarrow}(U)$ has df F_L (see Proposition 7.2 for a proof). We have to show that $E(L; L \geqslant q_\alpha(L)) = \int_\alpha^1 F_L^{\leftarrow}(u)\,\mathrm{d}u$. Now,

$$E(L; L \geqslant q_\alpha(L)) = E(F_L^{\leftarrow}(U); F_L^{\leftarrow}(U) \geqslant F_L^{\leftarrow}(\alpha)) = E(F_L^{\leftarrow}(U); U \geqslant \alpha);$$

in the last equality we used the fact that $F_L^{\leftarrow}$ is strictly increasing since F_L is continuous (see Proposition A.3 (iii)). Thus we get $E(F_L^{\leftarrow}(U); U \geqslant \alpha) = \int_\alpha^1 F_L^{\leftarrow}(u)\,\mathrm{d}u$. The second representation follows since, for a continuous loss distribution F_L, we have $P(L \geqslant q_\alpha(L)) = 1-\alpha$. □

For an extension of this result to loss distributions with atoms, we refer to Proposition 8.13. Next we use Lemma 2.13 to calculate the ES for two common continuous distributions.

Example 2.14 (expected shortfall for Gaussian loss distribution). Suppose that the loss distribution F_L is normal with mean μ and variance σ^2. Fix $\alpha \in (0,1)$. Then

$$\mathrm{ES}_\alpha = \mu + \sigma\frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}, \tag{2.24}$$

where ϕ is the density of the standard normal distribution. The proof is elementary. First note that

$$\mathrm{ES}_\alpha = \mu + \sigma E\left(\frac{L-\mu}{\sigma} \;\middle|\; \frac{L-\mu}{\sigma} \geqslant q_\alpha\left(\frac{L-\mu}{\sigma}\right)\right);$$

Table 2.3. VaR_α and ES_α in the normal and t models for different values of α.

α	0.90	0.95	0.975	0.99	0.995
VaR_α (normal model)	162.1	208.1	247.9	294.3	325.8
VaR_α (t model)	137.1	190.7	248.3	335.1	411.8
ES_α (normal model)	222.0	260.9	295.7	337.1	365.8
ES_α (t model)	223.5	286.5	357.2	466.9	565.7

hence, it suffices to compute the ES for the standard normal rv $\tilde{L} := (L-\mu)/\sigma$. Here we get

$$\mathrm{ES}_\alpha(\tilde{L}) = \frac{1}{1-\alpha}\int_{\Phi^{-1}(\alpha)}^{\infty} l\phi(l)\,\mathrm{d}l = \frac{1}{1-\alpha}[-\phi(l)]_{\Phi^{-1}(\alpha)}^{\infty} = \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}.$$

Example 2.15 (expected shortfall for the Student t loss distribution). Suppose the loss L is such that $\tilde{L} = (L-\mu)/\sigma$ has a standard t distribution with ν degrees of freedom, as in Example 2.11. Suppose further that $\nu > 1$. By the reasoning of Example 2.14, which applies to any location-scale family, we have $\mathrm{ES}_\alpha = \mu + \sigma\,\mathrm{ES}_\alpha(\tilde{L})$. The ES of the standard t distribution is easily calculated by direct integration to be

$$\mathrm{ES}_\alpha(\tilde{L}) = \frac{g_\nu(t_\nu^{-1}(\alpha))}{1-\alpha}\left(\frac{\nu + (t_\nu^{-1}(\alpha))^2}{\nu-1}\right), \tag{2.25}$$

where t_ν denotes the df and g_ν the density of standard t.

Since ES_α can be thought of as an average over all losses that are greater than or equal to VaR_α, it is sensitive to the severity of losses exceeding VaR_α. This advantage of ES is illustrated in the following example.

Example 2.16 (VaR and ES for stock returns). We consider daily losses on a position in a particular stock; the current value of the position equals $V_t = 10\,000$. Recall from Example 2.1 that the loss for this portfolio is given by $L_{t+1}^{\Delta} = -V_t X_{t+1}$, where X_{t+1} represents daily log-returns of the stock. We assume that X_{t+1} has mean 0 and standard deviation $\sigma = 0.2/\sqrt{250}$, i.e. we assume that the stock has an annualized volatility of 20%. We compare two different models for the distribution: namely, (i) a normal distribution, and (ii) a t distribution with $\nu = 4$ degrees of freedom scaled to have standard deviation σ. The t distribution is a symmetric distribution with heavy tails, so that large absolute values are much more probable than in the normal model; it is also a distribution that has been shown to fit well in many empirical studies (see Example 6.14). In Table 2.3 we present VaR_α and ES_α for both models and various values of α. In case (i) these values have been computed using (2.24); the ES for the t model has been computed using (2.25).

Most risk managers would argue that the t model is riskier than the normal model, since under the t distribution large losses are more likely. However, if we use VaR at the 95% or 97.5% confidence level to measure risk, the normal distribution appears to be at least as risky as the t model; only above a confidence level of 99% does the higher risk in the tails of the t model become apparent. On the other

hand, if we use ES, the risk in the tails of the t model is reflected in our risk measurement for lower values of α. Of course, simply going to a 99% confidence level in quoting VaR numbers does not help to overcome this deficiency of VaR, as there are other examples where the higher risk becomes apparent only for confidence levels beyond 99%.

Remark 2.17. It is possible to derive results on the asymptotics of the *shortfall-to-quantile ratio* $\mathrm{ES}_\alpha/\mathrm{VaR}_\alpha$ for $\alpha \to 1$. For the normal distribution we have $\lim_{\alpha\to 1} \mathrm{ES}_\alpha/\mathrm{VaR}_\alpha = 1$; for the t distribution with $\nu > 1$ degrees of freedom we have $\lim_{\alpha\to 1} \mathrm{ES}_\alpha/\mathrm{VaR}_\alpha = \nu/(\nu-1) > 1$. This shows that for a heavy-tailed distribution, the difference between ES and VaR is more pronounced than for the normal distribution. We will take up this issue in more detail in Section 5.2.3 (see also Section 8.4.4).

2.3.5 Coherent and Convex Risk Measures

The premise of this section is the idea of approaching risk measurement by first writing down a list of properties (axioms) that a good risk measure should have. For applications in risk management, such axioms have been proposed by Artzner et al. (1999) (coherent risk measures) and Föllmer and Schied (2002) (convex risk measures). In this section we discuss these axioms in relation to specific examples of risk measures. A longer and more theoretical treatment of coherent and convex risk measures will be given in Chapter 8. It should be mentioned that the idea of having axiomatic systems for risk measures bears some relationship to similar systems for premium principles in the actuarial literature, which have a long and independent history (see, for example, Goovaerts et al. (2003), as well as further references in the Notes and Comments section below).

Axioms for risk measures For the purposes of this section risk measures are real-valued functions defined on a linear space of random variables $\mathcal{M}$, assumed to include constants. There are two possible interpretations of the elements of $\mathcal{M}$. First, elements of $\mathcal{M}$ could be considered as future net asset values of portfolios or positions; in that case, elements of $\mathcal{M}$ will be denoted by V and the current net asset value will be denoted by V_0. Second, elements of $\mathcal{M}$ could represent losses L, where, of course, these are related to future values by the formula $L = -(V - V_0)$ (ignoring any discounting for simplicity).

Correspondingly, there are two possible notions of risk measures on $\mathcal{M}$. On the one hand, we can view the risk measure as the amount of *additional capital* that needs to be added to a position with future net asset value V to make the position acceptable to a regulator or a prudent manager; in this case we write the risk measure as $\tilde{\varrho}(V)$. On the other hand, we might interpret the risk measure as the *total amount of equity capital* that is necessary to back a position with loss L; in this case we write the risk measure as $\varrho(L)$.

These two notions are related by

$$\varrho(L) = V_0 + \tilde{\varrho}(V),$$

since "total capital" is equal to "available capital" plus "additional capital". However, these two different notions do have a bearing on the way in which the axioms are presented and understood. Since in this book we mostly focus on loss distributions, we present the axioms for losses and consider a risk measure $\varrho\colon L \mapsto \varrho(L)$. Note that the alternative notion is frequently found in the literature.

Axiom 2.18 (monotonicity). For $L_1, L_2 \in \mathcal{M}$ such that $L_1 \leqslant L_2$ almost surely, we have $\varrho(L_1) \leqslant \varrho(L_2)$.

From an economic viewpoint this axiom is obvious: positions that lead to higher losses in every state of the world require more risk capital. Positions with $\varrho(L) \leqslant 0$ do not require any capital.

Axiom 2.19 (translation invariance). For all $L \in \mathcal{M}$ and every $l \in \mathbb{R}$ we have $\varrho(L + l) = \varrho(L) + l$.

Axiom 2.19 states that by adding or subtracting a deterministic quantity l to a position leading to the loss L, we alter our capital requirements by exactly that amount. In terms of the alternative notion of a risk measure defined on future net asset values, this axiom implies that $\tilde{\varrho}(V + k) = \tilde{\varrho}(V) - k$ for $k \in \mathbb{R}$. It follows that $\tilde{\varrho}(V + \tilde{\varrho}(V)) = 0$, so a position with future net asset value $V + \tilde{\varrho}(V)$ is immediately acceptable without further injection of capital. This makes sense and implies that risk is measured in monetary terms.

Axiom 2.20 (subadditivity). For all $L_1, L_2 \in \mathcal{M}$ we have $\varrho(L_1 + L_2) \leqslant \varrho(L_1) + \varrho(L_2)$.

The rationale behind Axiom 2.20 is summarized by Artzner et al. (1999) in the statement that "a merger does not create extra risk" (ignoring, of course, any problematic practical aspects of a merger!). Axiom 2.20 is the most debated of the four axioms characterizing coherent risk measures, probably because it rules out VaR as a risk measure in certain situations. We provide some arguments explaining why subadditivity is indeed a reasonable requirement. First, subadditivity reflects the idea that risk can be reduced by diversification, a time-honoured principle in finance and economics. Second, if a regulator uses a non-subadditive risk measure in determining the regulatory capital for a financial institution, that institution has an incentive to legally break up into various subsidiaries in order to reduce its regulatory capital requirements. Similarly, if the risk measure used by an organized exchange in determining the margin requirements of investors is non-subadditive, an investor could reduce the margin he has to pay by opening a different account for every position in his portfolio. Finally, subadditivity makes decentralization of risk-management systems possible. Consider as an example two trading desks with positions leading to losses L_1 and L_2. Imagine that a risk manager wants to ensure that $\varrho(L)$, the risk of the overall loss $L = L_1 + L_2$, is smaller than some number M. If he uses a subadditive risk measure ϱ, he may simply choose bounds M_1 and M_2 such that $M_1 + M_2 \leqslant M$ and impose on each of the desks the constraint that $\varrho(L_i) \leqslant M_i$; subadditivity of ϱ then automatically ensures that $\varrho(L) \leqslant M_1 + M_2 \leqslant M$.

Axiom 2.21 (positive homogeneity). For all $L \in \mathcal{M}$ and every $\lambda > 0$ we have $\varrho(\lambda L) = \lambda \varrho(L)$.

Axiom 2.21 is easily justified if we assume that Axiom 2.20 holds. Subadditivity implies that, for $n \in \mathbb{N}$,

$$\varrho(nL) = \varrho(L + \cdots + L) \leqslant n\varrho(L). \tag{2.26}$$

Since there is no netting or diversification between the losses in this portfolio, it is natural to require that equality should hold in (2.26), which leads to positive homogeneity. Note that subadditivity and positive homogeneity imply that the risk measure ϱ is *convex* on $\mathcal{M}$.

Definition 2.22 (coherent risk measure). A risk measure ϱ whose domain includes the convex cone $\mathcal{M}$ is called *coherent* (on $\mathcal{M}$) if it satisfies Axioms 2.18–2.21.

Axiom 2.21 (positive homogeneity) has been criticized and, in particular, it has been suggested that for large values of the multiplier λ we should have $\varrho(\lambda L) > \lambda \varrho(L)$ in order to penalize a concentration of risk and to account for liquidity risk in a large position. As shown in (2.26), this is impossible for a subadditive risk measure. This problem has led to the study of the larger class of *convex risk measures*. In this class the conditions of subadditivity and positive homogeneity have been relaxed; instead one requires only the weaker property of convexity.

Axiom 2.23 (convexity). For all $L_1, L_2 \in \mathcal{M}$ and all $\lambda \in [0, 1]$ we have $\varrho(\lambda L_1 + (1 - \lambda)L_2) \leqslant \lambda \varrho(L_1) + (1 - \lambda)\varrho(L_2)$.

The economic justification for convexity is again the idea that diversification reduces risk.

Definition 2.24 (convex risk measure). A risk measure ϱ on $\mathcal{M}$ is called *convex* (on $\mathcal{M}$) if it satisfies Axioms 2.18, 2.19 and 2.23.

While every coherent risk measure is convex, the converse is not true. In particular, within the class of convex risk measures it is possible to find risk measures that penalize concentration of risk in the sense that $\varrho(\lambda L) \geqslant \varrho(L)$ for $\lambda > 1$ (see, for example, Example 8.8). On the other hand, for risk measures that are positive, homogeneous convexity and subadditivity are equivalent.

Examples. In view of its practical relevance we begin with a discussion of VaR. It is immediate from the definition of VaR as a quantile of the loss distribution that VaR is translation invariant, monotone and positive homogeneous. However, the following example shows that VaR is in general not subadditive, and hence, in general, neither is it a convex nor a coherent measure of risk.

Example 2.25 (non-subadditivity of VaR for defaultable bonds). Consider a portfolio of two zero-coupon bonds with a maturity of one year that default independently. The default probability of both bonds is assumed to be identical and equal to $p = 0.9\%$. The current price of the bonds and the face value of the bonds is equal to 100, and the bonds pay an interest rate of 5%. If there is no default, the holder

of the bond therefore receives a payment of size 105 in one year; in the case of a default, he receives nothing, i.e. we assume a recovery rate of zero. Denote by L_i the loss incurred by holding one unit of bond i. We have

$$\begin{aligned} P(L_i = -5) &= 1 - p = 0.991 \quad \text{(no default)}, \\ P(L_i = 100) &= p \qquad = 0.009 \quad \text{(default)}. \end{aligned}$$

Set $\alpha = 0.99$. We have $P(L_i < -5) = 0$ and $P(L_i \leqslant -5) = 0.991 > \alpha$, so $\mathrm{VaR}_\alpha(L_i) = -5$.

Now consider a portfolio of one bond from each firm, with corresponding loss $L = L_1 + L_2$. Since the default events of the two firms are independent, we get

$$\begin{aligned} P(L = -10) &= (1-p)^2 \quad = 0.982 \qquad \text{(no default)}, \\ P(L = 95) &= 2p(1-p) = 0.07838 \quad \text{(one default)}, \\ P(L = 200) &= p^2 \qquad = 0.000081 \quad \text{(two defaults)}. \end{aligned}$$

In particular, $P(L \leqslant -10) = 0.982 < 0.99$ and $P(L \leqslant 95) > 0.99$, so $\mathrm{VaR}_\alpha(L) = 95 > -10 = \mathrm{VaR}_\alpha(L_1) + \mathrm{VaR}_\alpha(L_2)$. Hence, VaR is non-subadditive. In fact, in the example, VaR_α punishes diversification, as

$$\mathrm{VaR}_\alpha(0.5L_1 + 0.5L_2) = 0.5\,\mathrm{VaR}_\alpha(L) = 47.5 > \mathrm{VaR}_\alpha(L_1).$$

In Example 2.25 the non-subadditivity of VaR is caused by the fact that the assets making up the portfolio have very skewed loss distributions; such a situation can clearly occur if we have defaultable bonds or options in our portfolio. Note, however, that the assets in this example have an innocuous dependence structure because they are independent.

In fact, the non-subadditivity of VaR can be seen in many different examples. The following is a list of the situations that we will encounter in this book.

- Independent losses with highly skewed discrete distributions, as in Example 2.25.
- Independent losses with continuous light-tailed distributions but low values of α. This will be demonstrated for exponentially distributed losses in Example 7.30 and discussed further in Section 8.3.3.
- Dependent losses with continuous symmetric distributions when the *dependence structure* takes a special form. This will be demonstrated for normally distributed losses in Example 8.39.
- Independent losses with continuous but very heavy-tailed distributions. This can be seen in Example 8.40 for the extreme case of infinite-mean Pareto risks. While less relevant for modelling market and credit risks, infinite-mean distributions are sometimes used to model certain kinds of insurance losses as well as losses due to operational risk (see Chapter 13 for more discussion).

Note that the domain $\mathcal{M}$ is an integral part of the definition of a convex or coherent risk measure. We will often encounter risk measures that are coherent or convex if

restricted to a sufficiently small domain. For example, VaR is subadditive in the idealized situation where all portfolios can be represented as linear combinations of the same set of underlying multivariate normal or, more generally, elliptically distributed risk factors (see Proposition 8.28).

There is an ongoing debate about the practical relevance of the non-subadditivity of VaR. The non-subadditivity can be particularly problematic if VaR is used to set risk limits for traders, as this can lead to portfolios with a high degree of concentration risk. Consider, for instance, in the set-up of Example 2.25 a trader who wants to maximize the expected return of a portfolio in the two defaultable bonds under the constraint that the VaR of his position is smaller than some given positive number; no short selling is permitted. Clearly, an optimal strategy for this trader is to invest all funds in one of the two bonds, a very concentrated position. For an elaboration of this toy example we refer to Frey and McNeil (2002).

Example 2.26 (coherence of expected shortfall). ES, on the other hand, is a coherent risk measure. Translation invariance, monotonicity and positive homogeneity are immediate from the corresponding properties of the quantile. For instance, it holds that

$$\mathrm{ES}_\alpha(\lambda L) = \frac{1}{1-\alpha}\int_\alpha^1 q_u(\lambda L)\,\mathrm{d}u = \frac{1}{1-\alpha}\int_\alpha^1 \lambda q_u(L)\,\mathrm{d}u = \lambda\,\mathrm{ES}_\alpha(L),$$

and similar arguments apply for translation invariance and monotonicity. A general proof of subadditivity is given in Theorem 8.14. Here, we give a simple argument for the case where L_1, L_2 and $L_1 + L_2$ have a continuous distribution. We recall from Lemma 2.13 that for a random variable L with a continuous distribution, it holds that

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha}E(LI_{\{L\geqslant q_\alpha(L)\}}).$$

To simplify the notation let $I_1 := I_{\{L_1\geqslant q_\alpha(L_1)\}}$, $I_2 := I_{\{L_2\geqslant q_\alpha(L_2)\}}$ and $I_{12} := I_{\{L_1+L_2\geqslant q_\alpha(L_1+L_2)\}}$. We calculate that

$$\begin{aligned}(1-\alpha)(\mathrm{ES}_\alpha(L_1) + \mathrm{ES}_\alpha(L_1) - \mathrm{ES}_\alpha(L_1+L_2))&\\ = E(L_1I_1) + E(L_2I_2) - E((L_1+L_2)I_{12})&\\ = E(L_1(I_1 - I_{12})) + E(L_2(I_2 - I_{12})).&\end{aligned}$$

Consider the first term and suppose that $\{L_1 \geqslant q_\alpha(L_1)\}$. It follows that $I_1 - I_{12} \geqslant 0$ and hence that $L_1(I_1 - I_{12}) \geqslant q_\alpha(L_1)(I_1 - I_{12})$. Suppose, on the other hand, that $\{L_1 < q_\alpha(L_1)\}$. It follows that $I_1 - I_{12} \leqslant 0$ and hence that $L_1(I_1 - I_{12}) \geqslant q_\alpha(L_1)(I_1 - I_{12})$. The same reasoning applies to L_2, so in either case we conclude that

$$\begin{aligned}(1-\alpha)(\mathrm{ES}_\alpha(L_1) + \mathrm{ES}_\alpha(L_1) - \mathrm{ES}_\alpha(L_1+L_2))&\\ \geqslant E(q_\alpha(L_1)(I_1 - I_{12})) + E(q_\alpha(L_2)(I_2 - I_{12}))&\\ \geqslant q_\alpha(L_1)E(I_1 - I_{12}) + q_\alpha(L_2)E(I_2 - I_{12})&\\ \geqslant q_\alpha(L_1)((1-\alpha) - (1-\alpha)) + q_\alpha(L_2)((1-\alpha) - (1-\alpha))&\\ = 0,&\end{aligned}$$

which proves subadditivity.

We have now seen two advantages of ES over VaR: ES reflects tail risk better (see Example 2.16) and it is always subadditive. On the other hand, VaR has practical advantages: it is easier to estimate, in particular for heavy-tailed distributions, and VaR estimates are easier to backtest than estimates of ES. We will come back to this point in our discussion of backtesting in Section 9.3.

Example 2.27 (generalized scenario risk measure). The generalized scenario risk measure in (2.16) is another example of a coherent risk measure. Translation invariance, positive homogeneity and monotonicity are clear, so it only remains to check subadditivity. For $i = 1, 2$ denote by $L_i(\boldsymbol{x})$ the hypothetical loss of position i under the scenario $\boldsymbol{x}$ for the risk-factor changes. We observe that

$$\begin{aligned}&\max\{E^P(L_1(\boldsymbol{X})) + L_2(\boldsymbol{X})) \colon P \in \mathcal{P}_{[\mathcal{X},\boldsymbol{w}]}\}\\ &\qquad \leqslant \max\{E^P(L_1(\boldsymbol{X})) \colon P \in \mathcal{P}_{[\mathcal{X},\boldsymbol{w}]}\} + \max\{E^P(L_2(\boldsymbol{X})) \colon P \in \mathcal{P}_{[\mathcal{X},\boldsymbol{w}]}\}.\end{aligned}$$

Example 2.28 (a coherent premium principle). In Fischer (2003), a class of coherent risk measures closely resembling certain actuarial premium principles is proposed. These risk measures are potentially useful for an insurance company that wants to compute premiums on a coherent basis without deviating too far from standard actuarial practice.

Given constants $p > 1$ and $\alpha \in [0, 1)$, this coherent premium principle $\varrho_{[\alpha,p]}$ is defined as follows. Let $\mathcal{M} := L^p(\Omega, \mathcal{F}, P)$, the space of all L with $\|L\|_p := E(|L|^p)^{1/p} < \infty$, and define, for $L \in \mathcal{M}$,

$$\varrho_{[\alpha,p]}(L) = E(L) + \alpha\|(L - E(L))^+\|_p. \tag{2.27}$$

Under (2.27) the risk associated with a loss L is measured by the sum of $E(L)$, the pure actuarial premium for the loss, and a *risk loading* given by a fraction α of the L^p-norm of the positive part of the centred loss $L - E(L)$. This loading can be written more explicitly as $(\int_{E(L)}^{\infty}(l - E(L))^p \,\mathrm{d}F_L(l))^{1/p}$. The higher the values of α and p, the more conservative the risk measure $\varrho_{[\alpha,p]}$ becomes.

The coherence of $\varrho_{[\alpha,p]}$ is easy to check. Translation invariance and positive homogeneity are immediate. To prove subadditivity observe that for any two rvs X and Y we have $(X + Y)^+ \leqslant X^+ + Y^+$. Hence, from Minkowski's inequality (the triangle inequality for the L^p-norm) we obtain that for any two $L_1, L_2 \in \mathcal{M}$,

$$\begin{aligned}\|(L_1 - E(L_1) + L_2 - E(L_2))^+\|_p &\leqslant \|(L_1 - E(L_1))^+ + (L_2 - E(L_2))^+\|_p\\ &\leqslant \|(L_1 - E(L_1))^+\|_p + \|(L_2 - E(L_2))^+\|_p,\end{aligned}$$

which shows that $\varrho_{[\alpha,p]}$ is subadditive. To verify monotonicity, assume that $L_1 \leqslant L_2$ almost surely and write $L = L_1 - L_2$. Since $L \leqslant 0$ almost surely, it follows that $(L - E(L))^+ \leqslant -E(L)$ almost surely, and hence that $\|(L - E(L))^+\|_p \leqslant -E(L)$ and $\varrho_{[\alpha,p]}(L) \leqslant 0$, since $\alpha < 1$. Using the fact that $L_1 = L_2 + L$ and the subadditivity property we obtain

$$\varrho_{[\alpha,p]}(L_1) \leqslant \varrho_{[\alpha,p]}(L_2) + \varrho_{[\alpha,p]}(L) \leqslant \varrho_{[\alpha,p]}(L_2).$$

Notes and Comments

An extensive discussion of different approaches to risk quantification is given in Crouhy, Galai and Mark (2001). Value-at-risk was introduced by JPMorgan in the first version of its RiskMetrics system and was quickly accepted by risk managers and regulators as an industry standard; see also Brown (2012) for a broader view of the history and use of VaR on Wall Street. A number of different notions of VaR are used in practice (see Alexander 2009, Volume 4) but all are related to the idea of a quantile of the P&L distribution.

Expected shortfall was made popular by Artzner et al. (1997, 1999). There are a number of variants of the ES risk measure with a variety of names, such as tail conditional expectation, worst conditional expectation and conditional VaR; all coincide for continuous loss distributions. Acerbi and Tasche (2002) discuss the relationships between the various notions. Risk measures based on loss distributions also appear in the literature under the (somewhat unfortunate) heading of *law-invariant* risk measures.

Example 2.25 is due to Albanese (1997) and Artzner et al. (1999). There are many different examples of the non-subadditivity of VaR in the literature, including the case of independent, infinite-mean Pareto risks (see Embrechts, McNeil and Straumann 2002, Example 7; Denuit and Charpentier 2004, Example 5.2.7). The implications of the non-subadditivity of VaR for portfolio optimization are discussed in Frey and McNeil (2002); see also papers by Basak and Shapiro (2001), Krokhmal, Palmquist and Uryasev (2001) and Emmer, Klüppelberg and Korn (2001).

A class of risk measures that are widely used throughout the hedge fund industry is based on the peak-to-bottom loss over a given period of time in the performance curve of an investment. These measures are typically referred to as (maximal) *drawdown* risk measures (see, for example, Chekhlov, Uryasev and Zabarankin 2005; Jaeger 2005).

The measurement of financial risk and the computation of actuarial premiums are at least conceptually closely related problems, so that the actuarial literature on premium principles is of relevance in financial risk management. We refer to Chapter 3 of Rolski et al. (1999) for an overview; Goovaerts, De Vylder and Haezendonck (1984) provides a specialist account.

Model risk has become a central issue in modern risk management. The problems faced by the hedge fund LTCM in 1998 provide a prime example of model risk in VaR-based risk-management systems. While LTCM had a seemingly sophisticated VaR system in place, errors in parameter estimation, unexpectedly large market moves (heavy tails) and, in particular, vanishing market liquidity drove the hedge fund into near-bankruptcy, causing major financial turbulence around the globe. Jorion (2000) contains an excellent discussion of the LTCM case, in particular comparing a Gaussian-based VaR model with a t-based approach. At a more general level, Jorion (2002a) discusses the various fallacies surrounding VaR-based market-risk-management systems.

3
Empirical Properties of Financial Data

In Chapter 2 we saw that the risk that a financial portfolio loses value over a given time period can be modelled in terms of changes in fundamental underlying risk factors, such as equity prices, interest rates and foreign exchange rates (see, in particular, the examples of Section 2.2.1). To build realistic models for risk-management purposes we need to consider the empirical properties of fundamental risk factors and develop models that share these properties.

In this chapter we first consider the univariate properties of single time series of risk-factor changes in Section 3.1, before reviewing some of the properties of multivariate series in Section 3.2. The features we describe motivate the statistical methodology of Part II of this book.

3.1 Stylized Facts of Financial Return Series

The *stylized facts* of financial time series are a collection of empirical observations, and inferences drawn from these observations, that apply to many time series of risk-factor changes, such as log-returns on equities, indices, exchange rates and commodity prices; these observations are now so entrenched in econometric experience that they have been accorded the status of facts.

The stylized facts that we describe typically apply to time series of daily log-returns and often continue to hold when we consider longer-interval series, such as weekly or monthly returns, or shorter-interval series, such as intra-daily returns. Most risk-management models are based on data collected at these frequencies.

Very-high-frequency financial time series, such as tick-by-tick data, have their own stylized facts, but this will not be a subject of this chapter. Moreover, the properties of very-low-frequency data (such as annual returns) are more difficult to pin down, due to the sparseness of such data and the difficulty of assuming that they are generated under long-term stationary regimes.

For a single time series of financial returns, a version of the stylized facts is as follows.

(1) Return series are not independent and identically distributed (iid), although they show little serial correlation.

(2) Series of absolute or squared returns show profound serial correlation.

(3) Conditional expected returns are close to zero.

(4) Volatility appears to vary over time.

(5) Extreme returns appear in clusters.

(6) Return series are leptokurtic or heavy tailed.

To discuss these observations further we denote a return series by $X_1, \ldots, X_n$ and assume that the returns have been formed by logarithmic differencing of a price, index or exchange-rate series $(S_t)_{t=0,1,\ldots,n}$, so $X_t = \ln(S_t/S_{t-1})$, $t = 1, \ldots, n$.

3.1.1 Volatility Clustering

Evidence for the first two stylized facts is collected in Figures 3.1 and 3.2. Figure 3.1 (a) shows 2608 daily log-returns for the DAX index spanning a decade from 2 January 1985 to 30 December 1994, a period including both the stock market crash of 1987 and the reunification of Germany. Parts (b) and (c) show series of simulated iid data from a normal model and a Student t model, respectively; in both cases the model parameters have been set by fitting the model to the real return data using the method of maximum likelihood under the iid assumption. In the normal case, this means that we simply simulate iid data with distribution $N(\mu, \sigma^2)$, where $\mu = \bar{X} = n^{-1}\sum_{i=1}^n X_i$ and $\sigma^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$. In the t case, the likelihood has been maximized numerically and the estimated degrees of freedom parameter is $\nu = 3.8$.

The simulated normal data are clearly very different from the DAX return data and do not show the same range of extreme values. While the Student t model can generate comparable extreme values to the real data, more careful observation reveals that the real returns exhibit a phenomenon known as *volatility clustering*, which is not present in the simulated series. Volatility clustering is the tendency for extreme returns to be followed by other extreme returns, although not necessarily with the same sign. We can see periods such as the stock market crash of October 1987 or the political uncertainty in the period between late 1989 and German reunification in 1990 in the DAX data: they are marked by large positive and negative moves.

In Figure 3.2 the *correlograms* of the raw data and the absolute data for all three data sets are shown. The correlogram is a graphical display for estimates of serial correlation, and its construction and interpretation are discussed in Section 4.1.3. While there is very little evidence of serial correlation in the raw data for all data sets, the absolute values of the real financial data appear to show evidence of serial dependence. Clearly, more than 5% of the estimated correlations lie outside the dashed lines, which are the 95% confidence intervals for serial correlations when the underlying process consists of iid finite-variance rvs. This serial dependence in the absolute returns would be equally apparent in squared return values, and it seems to confirm the presence of volatility clustering. We conclude that, although there is no evidence against the iid hypothesis for the genuinely iid data, there is strong evidence against the iid hypothesis for the DAX return data.

Table 3.1 contains more evidence against the iid hypothesis for daily stock-return data. The Ljung–Box test of randomness (described in Section 4.1.3) has been performed for the stocks comprising the Dow Jones 30 index in the period 1993–2000.

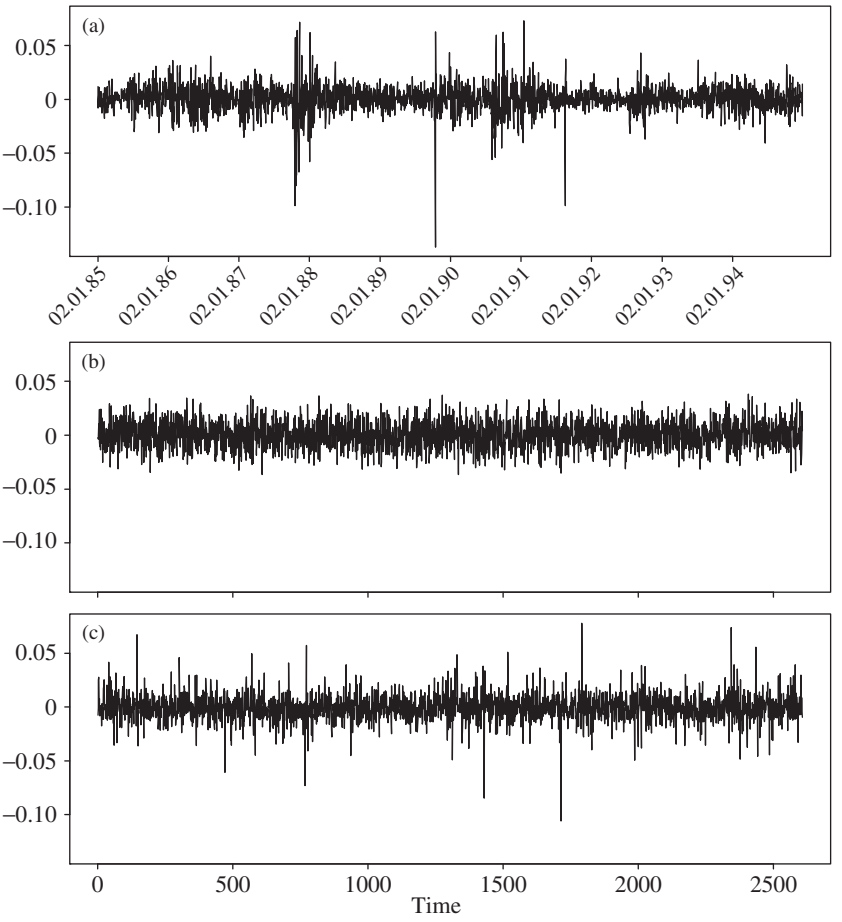

Figure 3.1. (a) Log-returns for the DAX index from 2 January 1985 to 30 December 1994 compared with simulated iid data from (b) a normal and (c) a t distribution, where the parameters have been determined by fitting the models to the DAX data.

In the two columns for daily returns the test is applied, respectively, to the raw return data (LBraw) and their absolute values (LBabs), and p-values are tabulated; these show strong evidence (particularly when applied to absolute values) against the iid hypothesis. If financial log-returns are not iid, then this contradicts the popular *random-walk* hypothesis for the discrete-time development of log-prices (or, in this case, index values). If log-returns are neither iid nor normal, then this contradicts the *geometric Brownian motion* hypothesis for the continuous-time development of prices on which the Black–Scholes–Merton pricing theory is based.

Moreover, if there is serial dependence in financial return data, then the question arises: to what extent can this dependence be used to make *predictions* about the future? This is the subject of the third and fourth stylized facts. It is very difficult to predict the return in the next time period based on historical data alone. This difficulty in predicting future returns is part of the evidence for the well-known

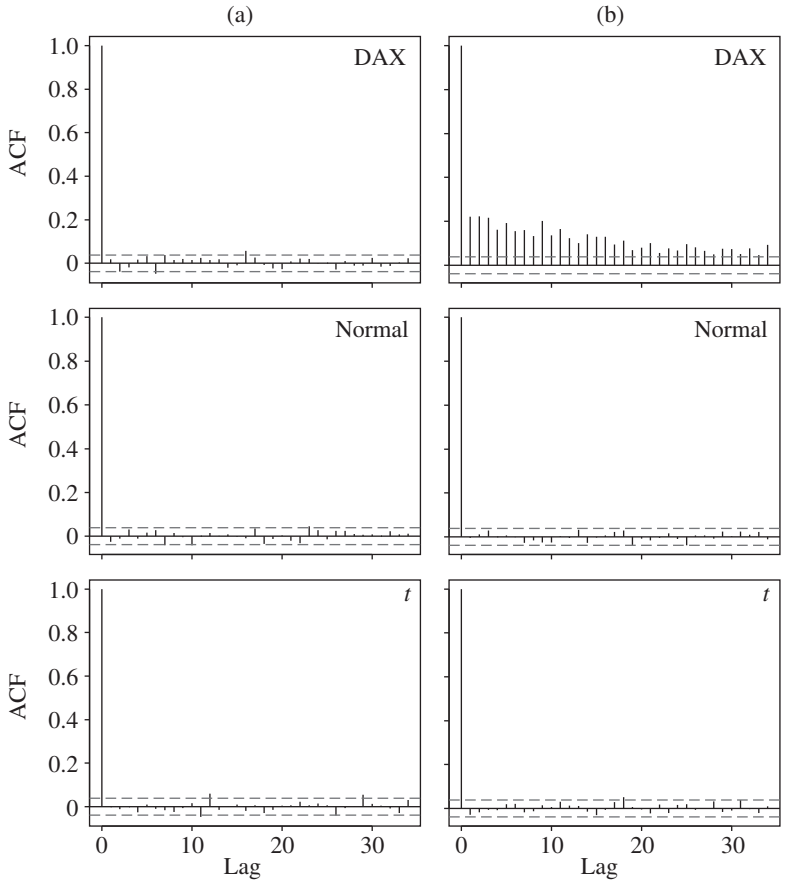

Figure 3.2. Correlograms for (a) the three data sets in Figure 3.1 and (b) the absolute values of these data. Dotted lines mark the standard 95% confidence intervals for the autocorrelations of a process of iid finite-variance rvs.

efficient markets hypothesis in finance, which says that prices react quickly to reflect all the available information about the asset in question.

In empirical terms, the lack of predictability of returns is shown by a lack of serial correlation in the raw return series data. For some series we do sometimes see evidence of correlations at the first lag (or first few lags). A small positive correlation at the first lag would suggest that there is some discernible tendency for a return with a particular sign (positive or negative) to be followed in the next period by a return with the same sign. However, this is not apparent in the DAX data, which suggests that our best estimate for tomorrow's return based on our observations up to today is zero. This idea is expressed in the assertion of the third stylized fact: that conditional expected returns are close to zero.

Volatility is often formally modelled as the *conditional standard deviation* of financial returns given historical information, and, although the conditional expected returns are consistently close to zero, the presence of volatility clustering suggests

Table 3.1. Tests of randomness for returns of Dow Jones 30 stocks in the eight-year period 1993–2000. The columns LBraw and LBabs show p-values for Ljung–Box tests applied to the raw and absolute values, respectively.

		Daily		Monthly	
Name	Symbol	LBraw	LBabs	LBraw	LBabs
Alcoa	AA	0.00	0.00	0.23	0.02
American Express	AXP	0.02	0.00	0.55	0.07
AT&T	T	0.11	0.00	0.70	0.02
Boeing	BA	0.03	0.00	0.90	0.17
Caterpillar	CAT	0.28	0.00	0.73	0.07
Citigroup	C	0.09	0.00	0.91	0.48
Coca-Cola	KO	0.00	0.00	0.50	0.03
DuPont	DD	0.03	0.00	0.75	0.00
Eastman Kodak	EK	0.15	0.00	0.61	0.54
Exxon Mobil	XOM	0.00	0.00	0.32	0.22
General Electric	GE	0.00	0.00	0.25	0.09
General Motors	GM	0.65	0.00	0.81	0.27
Hewlett-Packard	HWP	0.09	0.00	0.21	0.02
Home Depot	HD	0.00	0.00	0.00	0.41
Honeywell	HON	0.44	0.00	0.07	0.30
Intel	INTC	0.23	0.00	0.79	0.62
IBM	IBM	0.18	0.00	0.67	0.28
International Paper	IP	0.15	0.00	0.01	0.09
JPMorgan	JPM	0.52	0.00	0.43	0.12
Johnson & Johnson	JNJ	0.00	0.00	0.11	0.91
McDonald's	MCD	0.28	0.00	0.72	0.68
Merck	MRK	0.05	0.00	0.53	0.65
Microsoft	MSFT	0.28	0.00	0.19	0.13
3M	MMM	0.00	0.00	0.57	0.33
Philip Morris	MO	0.01	0.00	0.68	0.82
Procter & Gamble	PG	0.02	0.00	0.99	0.74
SBC	SBC	0.05	0.00	0.13	0.00
United Technologies	UTX	0.00	0.00	0.12	0.01
Wal-Mart	WMT	0.00	0.00	0.41	0.64
Disney	DIS	0.44	0.00	0.01	0.51

that conditional standard deviations are continually changing in a partly predictable manner. If we know that returns have been large in the last few days, due to market excitement, then there is reason to believe that the distribution from which tomorrow's return is "drawn" should have a large variance. It is this idea that lies behind the time-series models for changing volatility that we will examine in Chapter 4.

Further evidence for volatility clustering is given in Figure 3.3, where the time series of the 100 largest daily losses for the DAX returns and the 100 largest values for the simulated t data are plotted. In Section 5.3.1 we summarize the theory that suggests that the very largest values in iid data will occur like events in a

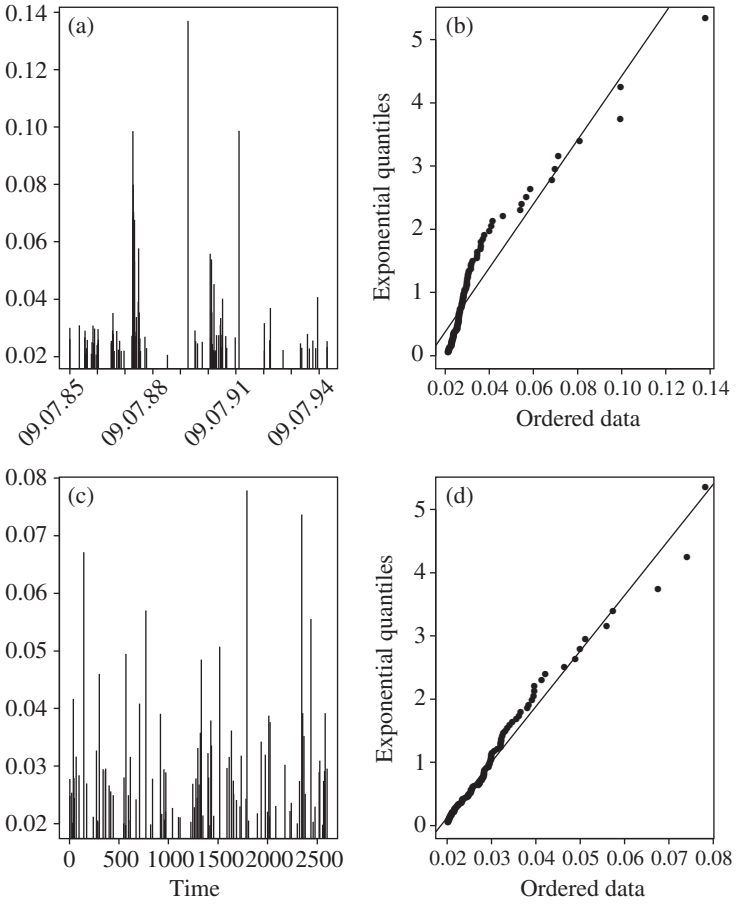

Figure 3.3. Time-series plots of the 100 largest negative values for (a) the DAX returns and (c) the simulated t data as well as (b), (d) Q–Q plots of the waiting times between these extreme values against an exponential reference distribution.

Poisson process, separated by waiting times that are iid with an exponential distribution. Parts (b) and (d) of the figure show Q–Q plots of these waiting times against an exponential reference distribution. While the hypothesis of the Poisson occurrence of extreme values for the iid data is supported, there are too many short waiting times and long waiting times caused by the clustering of extreme values in the DAX data to support the exponential hypothesis. The fifth stylized fact therefore constitutes further strong evidence against the iid hypothesis for return data.

In Chapter 4 we will introduce time-series models that have the volatility clustering behaviour that we observe in real return data. In particular, we will describe ARCH and GARCH models, which can replicate all of the stylized facts we have discussed so far, as well as the typical non-normality of returns addressed by the sixth stylized fact, to which we now turn.

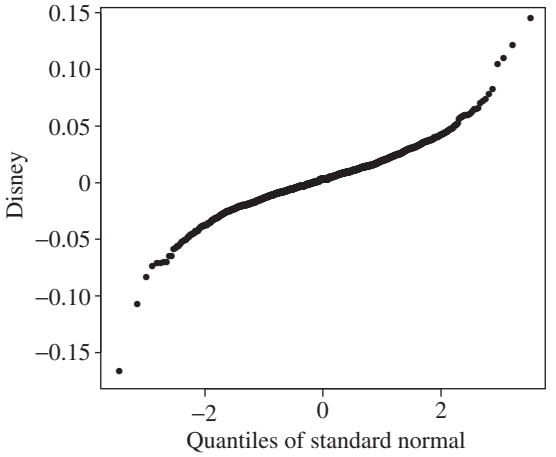

Figure 3.4. A Q–Q plot of daily returns of the Disney share price from 1993 to 2000 against a normal reference distribution.

3.1.2 Non-normality and Heavy Tails

The normal distribution is frequently observed to be a poor model for daily, weekly and even monthly returns. This can be confirmed using various well-known tests of normality, including the Q–Q plot against a standard normal reference distribution, as well as a number of formal numerical tests.

A Q–Q plot (quantile–quantile plot) is a standard visual tool for showing the relationship between empirical quantiles of the data and theoretical quantiles of a reference distribution. A lack of linearity in the Q–Q plot is interpreted as evidence against the hypothesized reference distribution. In Figure 3.4 we show a Q–Q plot of daily returns of the Disney share price from 1993 to 2000 against a normal reference distribution; the inverted S-shaped curve of the points suggests that the more extreme empirical quantiles of the data tend to be larger than the corresponding quantiles of a normal distribution, indicating that the normal distribution is a poor model for these returns.

Common numerical tests include those of Jarque and Bera, Anderson and Darling, Shapiro and Wilk, and D'Agostino. The Jarque–Bera test belongs to the class of omnibus moment tests, i.e. tests that assess simultaneously whether the *skewness* and *kurtosis* of the data are consistent with a Gaussian model. The sample skewness and kurtosis coefficients are defined by

$$b = \frac{(1/n)\sum_{i=1}^{n}(X_i - \bar{X})^3}{((1/n)\sum_{i=1}^{n}(X_i - \bar{X})^2)^{3/2}}, \qquad k = \frac{(1/n)\sum_{i=1}^{n}(X_i - \bar{X})^4}{((1/n)\sum_{i=1}^{n}(X_i - \bar{X})^2)^2}. \tag{3.1}$$

These are designed to estimate the theoretical skewness and kurtosis, which are defined, respectively, by $\beta = E(X-\mu)^3/\sigma^3$ and $\kappa = E(X-\mu)^4/\sigma^4$, where $\mu = E(X)$ and $\sigma^2 = \mathrm{var}(X)$ denote mean and variance; β and κ take the values zero and three for a normal variate X. The Jarque–Bera test statistic is

$$T = \tfrac{1}{6}n(b^2 + \tfrac{1}{4}(k-3)^2), \tag{3.2}$$

Table 3.2. Sample skewness (b) and kurtosis (k) coefficients as well as p-values for Jarque–Bera tests of normality for an arbitrary set of ten of the Dow Jones 30 stocks (see Table 3.1 for names of stocks).

	Daily returns, $n = 2020$			Weekly returns, $n = 416$		
Stock	b	k	p-value	b	k	p-value
AXP	0.05	5.09	0.00	−0.01	3.91	0.00
EK	−1.93	31.20	0.00	−1.13	14.40	0.00
BA	−0.34	10.89	0.00	−0.26	7.54	0.00
C	0.21	5.93	0.00	0.44	5.42	0.00
KO	−0.02	6.36	0.00	−0.21	4.37	0.00
MSFT	−0.22	8.04	0.00	−0.14	5.25	0.00
HWP	−0.23	6.69	0.00	−0.26	4.66	0.00
INTC	−0.56	8.29	0.00	−0.65	5.20	0.00
JPM	0.14	5.25	0.00	−0.20	4.93	0.00
DIS	−0.01	9.39	0.00	0.08	4.48	0.00
	Monthly returns, $n = 96$			Quarterly returns, $n = 32$		
Stock	b	k	p-value	b	k	p-value
AXP	−1.22	5.99	0.00	−1.04	4.88	0.01
EK	−1.52	10.37	0.00	−0.63	4.49	0.08
BA	−0.50	4.15	0.01	−0.15	6.23	0.00
C	−1.10	7.38	0.00	−1.61	7.13	0.00
KO	−0.49	3.68	0.06	−1.45	5.21	0.00
MSFT	−0.40	3.90	0.06	−0.56	2.90	0.43
HWP	−0.33	3.47	0.27	−0.38	3.64	0.52
INTC	−1.04	6.50	0.00	−0.42	3.10	0.62
JPM	−0.51	5.40	0.00	−0.78	7.26	0.00
DIS	0.04	3.26	0.87	−0.49	4.32	0.16

and it has an asymptotic chi-squared distribution with two degrees of freedom under the null hypothesis of normality; sample kurtosis values differing widely from three and skewness values differing widely from zero may lead to rejection of normality.

In Table 3.2 tests of normality are applied to an arbitrary subgroup of ten of the stocks comprising the Dow Jones index. We take eight years of data spanning the period 1993–2000 and form daily, weekly, monthly and quarterly logarithmic returns. For each stock we calculate sample skewness and kurtosis and apply the Jarque–Bera test to the univariate time series. The daily and weekly return data fail all tests; in particular, it is notable that there are some large values for the sample kurtosis. For the monthly data, the null hypothesis of normality is not formally rejected (that is, the p-value is greater than 0.05) for four of the stocks; for quarterly data, it is not rejected for five of the stocks, although here the sample size is small.

The Jarque–Bera test (3.2) clearly rejects the normal hypothesis. In particular, daily financial return data appear to have a much higher kurtosis than is consistent with the normal distribution; their distribution is said to be *leptokurtic*, meaning

that it is more narrow than the normal distribution in the centre but has longer and heavier tails.

Further empirical analysis often suggests that the distribution of daily or other short-interval financial return data has tails that decay slowly according to a power law, rather than the faster, exponential-type decay of the tails of a normal distribution. This means that we tend to see rather more extreme values than might be expected in such return data; we discuss this phenomenon further in Chapter 5, which is devoted to extreme value theory.

3.1.3 Longer-Interval Return Series

As we progressively increase the interval of the returns by moving from daily to weekly, monthly, quarterly and yearly data, the phenomena we have identified tend to become less pronounced. Volatility clustering decreases and returns begin to look both more iid and less heavy tailed.

Beginning with a sample of n returns measured at some time interval (say daily or weekly), we can aggregate these to form longer-interval log-returns. The h-period log-return at time t is given by

$$X_t^{(h)} = \ln\left(\frac{S_t}{S_{t-h}}\right) = \ln\left(\frac{S_t}{S_{t-1}} \cdots \frac{S_{t-h+1}}{S_{t-h}}\right) = \sum_{j=0}^{h-1} X_{t-j}, \tag{3.3}$$

and from our original sample we can form a sample of *non-overlapping* h-period returns $\{X_t^{(h)}: t = h, 2h, \ldots, \lfloor n/h \rfloor h\}$, where $\lfloor \cdot \rfloor$ denotes the integer part or *floor* function; $\lfloor x \rfloor = \max\{k \in \mathbb{Z}: k \leqslant x\}$ is the largest integer not greater than x.

Due to the sum structure of the h-period returns, it is to be expected that some central limit effect takes place, whereby their distribution becomes less leptokurtic and more normal as h is increased. Note that, although we have cast doubt on the iid model for daily data, a central limit theorem also applies to many stationary time-series processes, including the GARCH models that are a focus of Chapter 4.

In Table 3.1 the Ljung–Box tests of randomness have also been applied to non-overlapping monthly return data. For twenty out of thirty stocks the null hypothesis of iid data is not rejected at the 5% level in Ljung–Box tests applied to both the raw and absolute returns. It is therefore harder to find evidence of serial dependence in such monthly returns.

Aggregating data to form non-overlapping h-period returns reduces the sample size from n to $\lfloor n/h \rfloor$, and for longer-period returns (such as quarterly or yearly returns) this may be a very serious reduction in the amount of data. An alternative in this case is to form overlapping returns. For $1 \leqslant k < h$ we can form overlapping returns by taking

$$\{X_t^{(h)}: t = h, h + k, h + 2k, \ldots, h + \lfloor (n - h)/k \rfloor k\}, \tag{3.4}$$

which yields $1 + \lfloor (n - h)/k \rfloor$ values that overlap by an amount $h - k$. In forming overlapping returns we can preserve a large number of data points, but we do build additional serial dependence into the data. Even if the original data were iid, overlapping data would be profoundly dependent, which can greatly complicate their analysis.

Notes and Comments

A number of texts contain extensive empirical analyses of financial return series and discussions of their properties. We mention in particular Taylor (2008), Alexander (2001), Tsay (2002) and Zivot and Wang (2003). For more discussion of the random-walk hypothesis for stock returns, and its shortcomings, see Lo and MacKinlay (1999).

There are countless possible tests of univariate normality and a good starting point is the entry on "departures from normality, tests for" in Volume 2 of the *Encyclopedia of Statistics* (Kotz, Johnson and Read 1985). For an introduction to Q–Q plots see Rice (1995, pp. 353–357); for the widely applied Jarque–Bera test based on the sample skewness and kurtosis, see Jarque and Bera (1987).

3.2 Multivariate Stylized Facts

In risk-management applications we are usually interested in multiple series of financial risk-factor changes. To the stylized facts identified in Section 3.1 we may add a number of stylized facts of a multivariate nature.

We now consider multivariate return data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. Each *component series* $X_{1,j}, \ldots, X_{n,j}$ for $j = 1, \ldots, d$ is a series formed by logarithmic differencing of a daily price, index or exchange-rate series as before. Commonly observed multivariate stylized facts include the following.

(M1) Multivariate return series show little evidence of cross-correlation, except for contemporaneous returns.

(M2) Multivariate series of absolute returns show profound evidence of cross-correlation.

(M3) Correlations between series (i.e. between contemporaneous returns) vary over time.

(M4) Extreme returns in one series often coincide with extreme returns in several other series.

3.2.1 Correlation between Series

The first two observations are fairly obvious extensions of univariate stylized facts (1) and (2) from Section 3.1. Just as the stock returns for, say, Microsoft on days t and $t+h$ (for $h > 0$) show very little serial correlation, so we generally detect very little correlation between the Microsoft return on day t and, say, the Coca-Cola return on day $t + h$. Of course, stock returns on the same day may show non-negligible correlation, due to factors that affect the whole market on that day. When we look at absolute returns we should bear in mind that periods of high or low volatility are generally common to more than one stock. Returns of large magnitude in one stock may therefore tend to be followed on subsequent days by further returns of large magnitude for both that stock and other stocks, which can explain (M2). The issue of cross-correlation and its estimation is a topic in multivariate time-series analysis and is addressed with an example in Section 14.1.

Stylized fact (M3) is a multivariate counterpart to univariate observation (4): that volatility appears to vary with time. It can be interpreted in a couple of ways with reference to different underlying models. On the one hand, we could refer to a model in which there are so-called stationary regimes; during these regimes correlations are fixed but the regimes change from time to time. On the other hand, we could refer to more dynamic models in which a *conditional correlation* is changing all the time. Just as volatility is often formally modelled as the conditional standard deviation of returns given historical information, we can also devise models that feature a changing conditional correlation given historical information. Examples of such models include certain multivariate GARCH models, as discussed in Section 14.2. In the context of such models it is possible to demonstrate (M3) for many pairs of risk-factor return series.

However, we should be careful about drawing conclusions about changing correlations based on more ad hoc analyses. To explain this further we consider two data sets. The first, shown in Figure 3.5, comprises the BMW and Siemens daily log-return series for the period from 23 January 1985 to 22 September 1994; there are precisely 2000 values.

The second data set, shown in Figure 3.6, consists of an equal quantity of randomly generated data from a bivariate t distribution. The parameters have been chosen by fitting the distribution to the BMW-Siemens data by the method of maximum likelihood. The fitted model is estimated to have 2.8 degrees of freedom and estimated correlation 0.72 (see Section 6.2.1 for more details of the multivariate t distribution).

The two data sets show some superficial resemblance. The distribution of values is similar in both cases. However, the simulated data are independent and there is no serial dependence or volatility clustering.

We estimate rolling correlation coefficients for both series using a moving window of twenty-five days, which is approximately the number of trading days in a typical calendar month. These kinds of rolling empirical estimates are quite commonly used in practice to gather evidence of how key model parameters may change. In Figure 3.7 the resulting estimates are shown for the BMW–Siemens log-return data and the iid Student t-distributed data.

Remarkably, there are no obvious differences between the results for the two data sets; if anything, the range of estimated correlation values for the iid data is greater, despite the fact that they are generated from a single stationary model with a fixed correlation of 0.72.

This illustrates that simple attempts to demonstrate (M3) using empirical correlation estimates should be interpreted with care. There is considerable error involved in estimating correlations from small samples, particularly when the underlying distribution is a heavier-tailed bivariate distribution, such as a t distribution, rather than a Gaussian distribution (see also Example 6.30 in this context). The most reliable way to substantiate (M3) and to decide in exactly what way correlation changes is to fit different models for changing correlation and then to make formal statistical comparisons of the models.

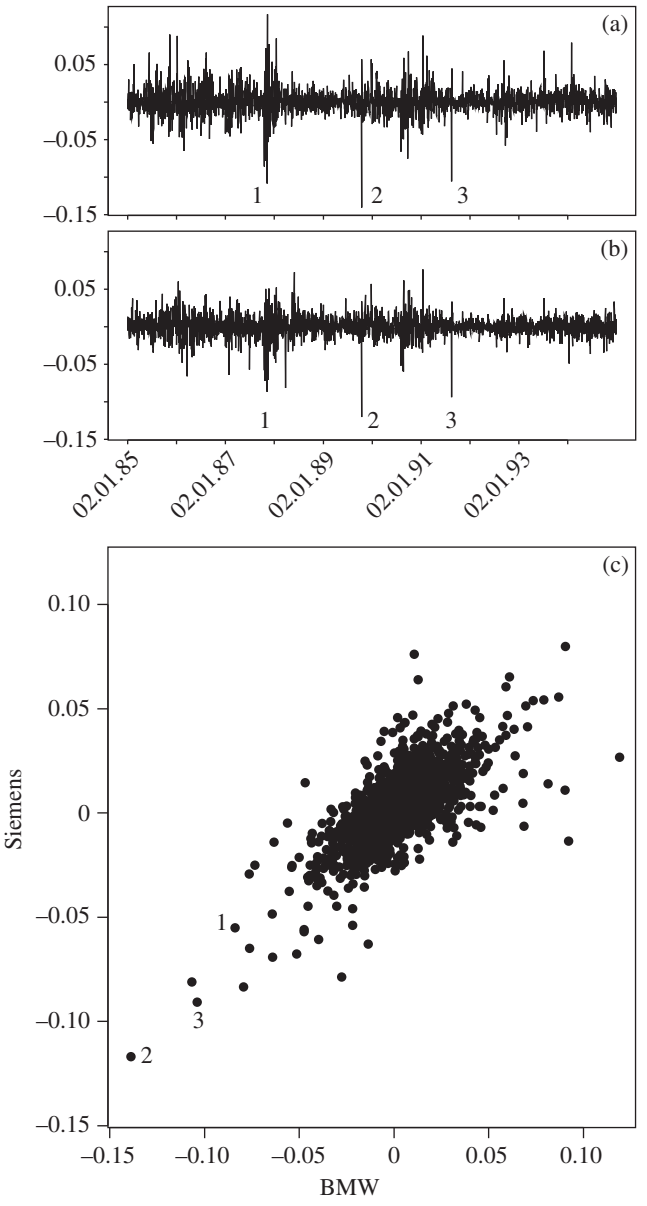

Figure 3.5. (a) BMW and (b) Siemens log-return data for the period from 23 January 1985 to 22 September 1994 together with (c) pairwise scatterplot. Three extreme days on which large negative returns occurred are marked. The dates are 19 October 1987, 16 October 1989 and 19 August 1991 (see Section 3.2.2 for historical commentary).

3.2.2 Tail Dependence

Stylized fact (M4) is often apparent when time series are compared. Consider again the BMW and Siemens log-returns in Figure 3.5. In both the time-series plots and the scatterplot, three days have been indicated with a number. These are days on

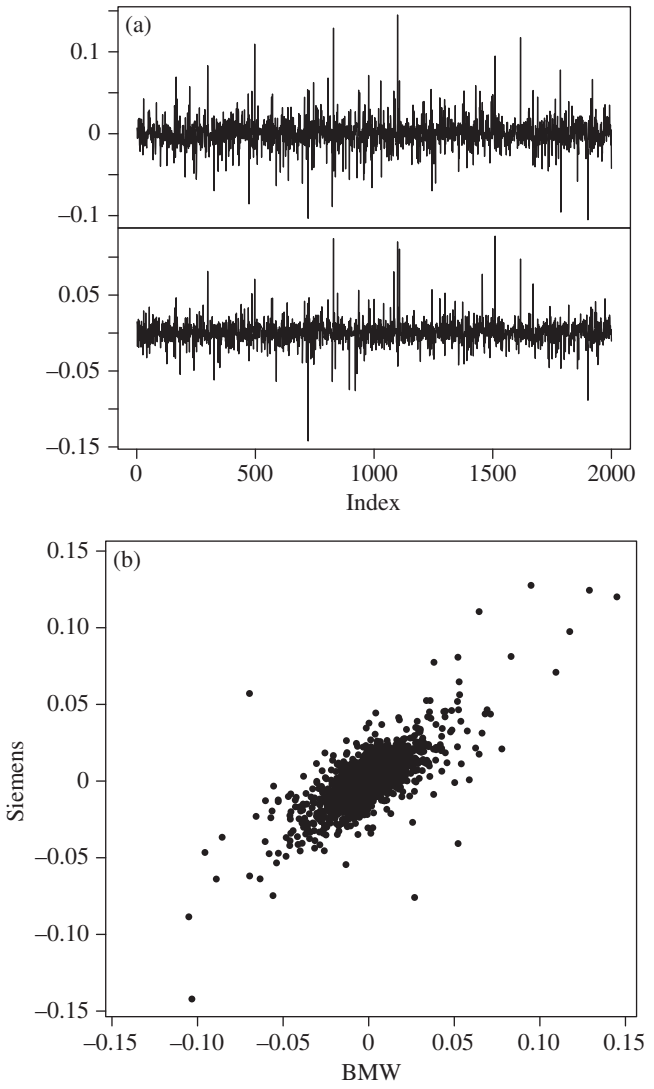

Figure 3.6. Two-thousand iid data points generated from a bivariate t distribution: (a) time series of components and (b) pairwise scatterplot. The parameters have been set by fitting the t distribution to the BMW–Siemens data in Figure 3.5.

which large negative returns were observed for both stocks, and all three occurred during periods of volatility on the German market. They are, respectively, 19 October 1987, Black Monday on Wall Street; 16 October 1989, when over 100 000 Germans protested against the East German regime in Leipzig during the chain of events that led to the fall of the Berlin Wall and German reunification; and 19 August 1991, the day of the coup by communist hardliners during the reforming era of Gorbachev in the USSR. Clearly, these are days on which momentous events led to joint extreme values.

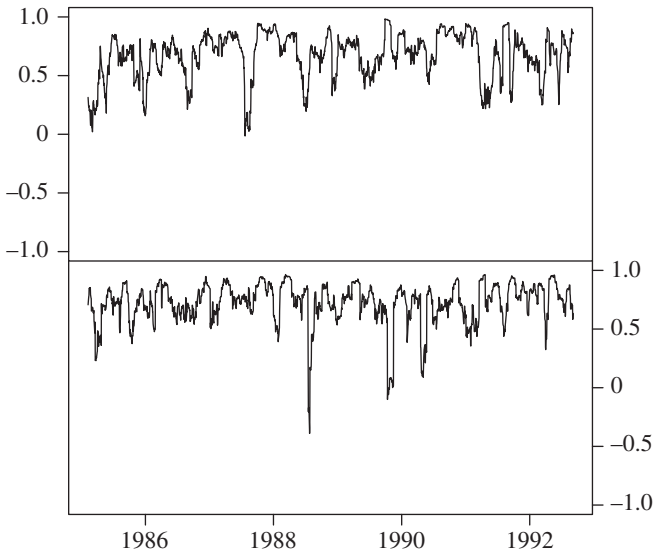

Figure 3.7. Twenty-five-day rolling correlation estimates for the empirical data of Figure 3.5 (top panel) and for the simulated iid data of Figure 3.6 (bottom panel).

Related to stylized fact (M4) is the idea that "correlations go to one in times of market stress". The three extreme days in Figure 3.5 correspond to points that are close to the diagonal of the scatterplot in the lower-left-hand corner, and it is easy to see why one might describe these as occasions on which correlations tend to one. It is quite difficult to formally test the hypothesis that model correlations are higher when volatilities are higher, and this should be done in the context of a multivariate time-series model incorporating either dynamic conditional correlations or regime changes. Once again, we should be cautious about interpreting simple analyses based on empirical correlation estimates, as we now show.

In Figure 3.8 we perform an analysis in which we split the 2000 bivariate return observations into eighty non-overlapping twenty-five-day blocks. In each block we estimate the empirical correlation between the return series and the volatility of the two series. We plot the *Fisher transform* of the estimated correlation against the estimated volatility of the BMW series and then regress the former on the latter. There is a strongly significant regression relationship (shown by the line) between the correlation and volatility estimates. It is tempting to say that in stress periods where volatility is high, correlation is also high. The Fisher transform is a well-known variance-stabilizing transform that is appropriate when correlation is the dependent variable in a regression analysis.

However, when exactly the same exercise is carried out for the data generated from a t distribution (Figure 3.6 (b)), the result is similar. In this case the observation that estimated correlation is higher in periods of higher estimated volatility is a pure artefact of estimation error for both quantities, since the true underlying correlation is fixed.

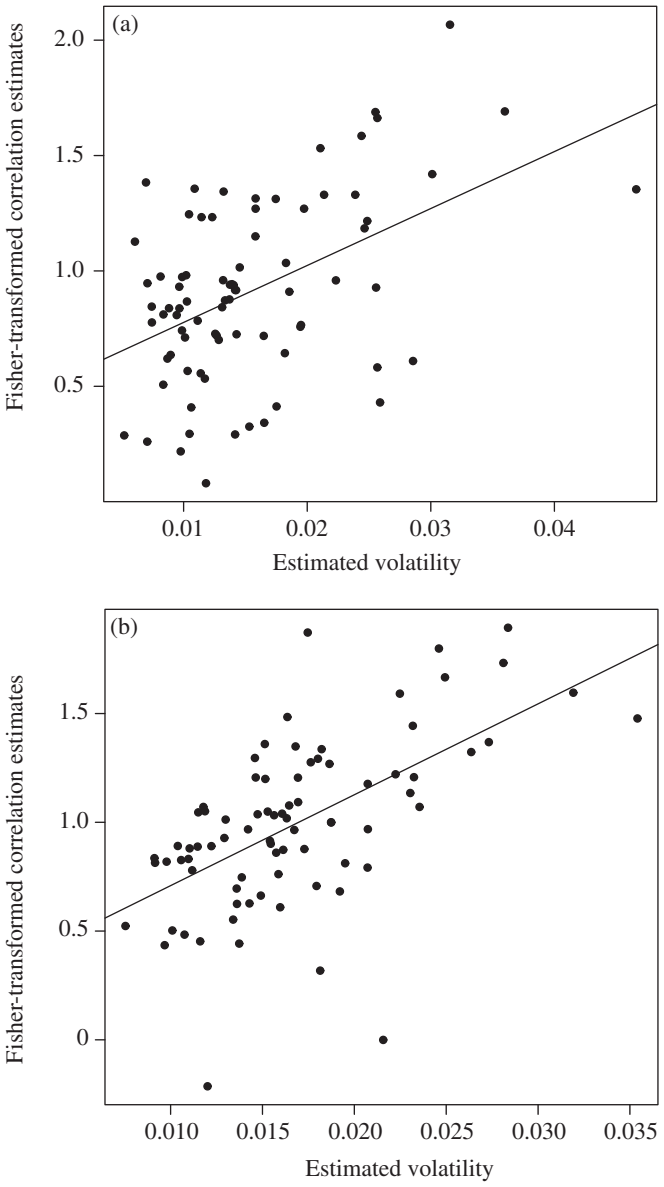

Figure 3.8. Fisher transforms of estimated correlations plotted against estimated volatilities for eighty non-overlapping blocks of twenty-five observations: (a) the BMW–Siemens log-return data in Figure 3.5; and (b) the simulated bivariate t data in Figure 3.6. In both cases there is a significant regression relationship between estimated correlations and estimated volatilities.

This example is not designed to argue against the view that correlations are higher when volatilities are higher; it is simply meant to show that it is difficult to demonstrate this using an ad hoc approach based on estimated correlations. Formal comparison of different multivariate volatility and correlation models with differing

specifications for correlation is required; some models of this kind are described in Section 14.2. Moreover, while it may be partly true that useful multivariate time-series models for returns should have the property that conditional correlations tend to become large when volatilities are large, the phenomenon of simultaneous extreme values can also be addressed in other ways.

For example, we can choose distributions in multivariate models that have so-called *tail dependence* or *extremal dependence*. Loosely speaking, this means models in which the conditional probability of observing an extreme value for one risk factor given that we have observed an extreme value for another is non-negligible and, indeed, is in some cases quite large. A mathematical definition of this notion and a discussion of its importance may be found in Section 7.2.4.

Notes and Comments

Pitfalls in tests for changing correlations are addressed in an interesting paper by Boyer, Gibson and Loretan (1999), which argues against simplistic empirical analyses based on segmenting the data into normal and stressed regimes. See also Loretan and English (2000) for a discussion of correlation breakdowns during periods of market instability. An interesting book on the importance of correlation in risk management is Engle (2009).

Tail dependence has various definitions: see Joe (1997) and Coles, Heffernan and Tawn (1999). The importance of tail dependence in risk management was highlighted in Embrechts, McNeil and Straumann (1999), Embrechts, McNeil and Straumann (2002) and Mashal, Naldi and Zeevi (2003). It is now a recognized issue in the regulatory literature: see, for example, the discussion of tail correlation in the CEIOPS consultation paper on the use of correlation in the standard formula for the solvency capital requirement (CEIOPS 2009).

Part II

Methodology

4
Financial Time Series

Motivated by the discussion of the empirical properties of financial risk-factor change data in Chapter 3, in this chapter we present univariate time-series models that mimic the properties of real return data.

In Section 4.1 we review essential concepts in the analysis of time series, such as stationarity, autocorrelations and their estimation, white noise processes, and ARMA (autoregressive moving-average) processes. We then devote Section 4.2 to univariate ARCH and GARCH (generalized autoregressive conditionally heteroscedastic) processes for capturing the important phenomenon of volatility.

GARCH models are certainly not the only models for describing the volatility of financial returns. Other important classes of model include discrete-time stochastic volatility models, long-memory GARCH models, continuous-time models fitted to discrete data, and models based on realized volatility calculated from high-frequency data; these alternative approaches are not handled in this book.

Our emphasis on GARCH has two main motivations, the first being a practical one. We recall that in risk management we are typically dealing with very large numbers of risk factors, and our philosophy, expounded in Section 1.5, is that broad-brush techniques that capture the main risk features of many time series are more important than very detailed analyses of single series. The GARCH model lends itself to this approach and proves relatively easy to fit. There are also some multivariate extensions (see Chapter 14) that build in simple ways on the univariate models and that may be calibrated to a multivariate series in stages. This ease of use contrasts with other models where the fitting of a single series often presents a computational challenge (e.g. estimation of a stochastic volatility model via filtering or Gibbs sampling), and multivariate extensions have not been widely considered. Moreover, an average financial enterprise will typically collect daily data on its complete set of risk factors for the purposes of risk management, and this rules out some more sophisticated models that require higher-frequency data.

Our second reason for concentrating on ARCH and GARCH models is didactic. These models for volatile return series have a status akin to ARMA models in classical time series; they belong, in our opinion, to the body of standard methodology to which a student of the subject should be exposed. A quantitative risk manager who understands GARCH has a good basis for understanding more complex models and a good framework for talking about historical volatility in a rational way. He/she may also appreciate more clearly the role of more ad hoc volatility

estimation methods such as the exponentially weighted moving-average (EWMA) procedure.

4.1 Fundamentals of Time Series Analysis

This section provides a short summary of the essentials of classical univariate time-series analysis with a focus on that which is relevant for modelling risk-factor return series. We have based the presentation on Brockwell and Davis (1991, 2002), so these texts may be used as supplementary reading.

4.1.1 Basic Definitions

A time-series model for a single risk factor is a discrete-time stochastic process $(X_t)_{t\in\mathbb{Z}}$, i.e. a family of rvs, indexed by the integers and defined on some probability space $(\Omega, \mathcal{F}, P)$.

Moments of a time series. Assuming they exist, we define the *mean function* $\mu(t)$ and the *autocovariance function* $\gamma(t, s)$ of $(X_t)_{t\in\mathbb{Z}}$ by

$$\begin{aligned} \mu(t) &= E(X_t), && t \in \mathbb{Z}, \\ \gamma(t, s) &= E((X_t - \mu(t))(X_s - \mu(s))), && t, s \in \mathbb{Z}. \end{aligned}$$

It follows that the autocovariance function satisfies $\gamma(t, s) = \gamma(s, t)$ for all t, s, and $\gamma(t, t) = \mathrm{var}(X_t)$.

Stationarity. Generally, the processes we consider will be stationary in one or both of the following two senses.

Definition 4.1 (strict stationarity). The time series $(X_t)_{t\in\mathbb{Z}}$ is *strictly* stationary if

$$(X_{t_1}, \ldots, X_{t_n}) \stackrel{\mathrm{d}}{=} (X_{t_1+k}, \ldots, X_{t_n+k})$$

for all $t_1, \ldots, t_n, k \in \mathbb{Z}$ and for all $n \in \mathbb{N}$.

Definition 4.2 (covariance stationarity). The time series $(X_t)_{t\in\mathbb{Z}}$ is *covariance* stationary (or *weakly* or *second-order* stationary) if the first two moments exist and satisfy

$$\begin{aligned} \mu(t) &= \mu, && t \in \mathbb{Z}, \\ \gamma(t, s) &= \gamma(t + k, s + k), && t, s, k \in \mathbb{Z}. \end{aligned}$$

Both these definitions attempt to formalize the notion that the behaviour of a time series is similar in any epoch in which we might observe it. Systematic changes in mean, variance or the covariances between equally spaced observations are inconsistent with stationarity.

It may be easily verified that a strictly stationary time series with finite variance is covariance stationary, but it is important to note that we may define infinite-variance processes (including certain ARCH and GARCH processes) that are strictly stationary but not covariance stationary.

Autocorrelation in stationary time series. The definition of covariance stationarity implies that for all s, t we have $\gamma(t-s,0) = \gamma(t,s) = \gamma(s,t) = \gamma(s-t,0)$, so that the covariance between X_t and X_s only depends on their temporal separation $|s-t|$, which is known as the *lag*. Thus, for a covariance-stationary process we write the autocovariance function as a function of one variable:

$$\gamma(h) := \gamma(h,0), \quad \forall h \in \mathbb{Z}.$$

Noting that $\gamma(0) = \text{var}(X_t)$, $\forall t$, we can now define the autocorrelation function of a covariance-stationary process.

Definition 4.3 (autocorrelation function). The *autocorrelation function* (ACF) $\rho(h)$ of a covariance-stationary process $(X_t)_{t\in\mathbb{Z}}$ is

$$\rho(h) = \rho(X_h, X_0) = \gamma(h)/\gamma(0), \quad \forall h \in \mathbb{Z}.$$

We speak of the autocorrelation or *serial correlation* $\rho(h)$ at lag h. In classical time-series analysis the set of serial correlations and their empirical analogues estimated from data are the objects of principal interest. The study of autocorrelations is known as *analysis in the time domain*.

White noise processes. The basic building blocks for creating useful time-series models are stationary processes without serial correlation, known as *white noise* processes and defined as follows.

Definition 4.4 (white noise). $(X_t)_{t\in\mathbb{Z}}$ is a white noise process if it is covariance stationary with autocorrelation function

$$\rho(h) = \begin{cases} 1, & h = 0, \\ 0, & h \neq 0. \end{cases}$$

A white noise process centred to have mean 0 with variance $\sigma^2 = \text{var}(X_t)$ will be denoted $\text{WN}(0, \sigma^2)$. A simple example of a white noise process is a series of iid rvs with finite variance, and this is known as a *strict white noise* process.

Definition 4.5 (strict white noise). $(X_t)_{t\in\mathbb{Z}}$ is a strict white noise process if it is a series of iid, finite-variance rvs.

A strict white noise (SWN) process centred to have mean 0 and variance σ^2 will be denoted $\text{SWN}(0, \sigma^2)$. Although SWN is the easiest kind of noise process to understand, it is not the only noise that we will use. We will later see that covariance-stationary ARCH and GARCH processes are in fact white noise processes.

Martingale difference. One further noise concept that we use, particularly when we come to discuss volatility and GARCH processes, is that of a martingale-difference sequence. We recall that a *martingale* is a sequence of integrable rvs (M_t) such that the expected value of M_t given the previous history of the sequence is M_{t-1}. This implies that if we define (X_t) by taking first differences of the sequence (M_t), then the expected value of X_t given information about previous values is 0. We have observed in Section 3.1 that this property may be appropriate for financial return

data. A martingale difference is often said to model our winnings in consecutive rounds of a *fair game*.

To discuss this concept more precisely, we assume that the time series $(X_t)_{t\in\mathbb{Z}}$ is adapted to some *filtration* $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ that represents the *accrual of information* over time. The sigma algebra $\mathcal{F}_t$ represents the available information at time t, and typically this will be the information contained in past and present values of the time series itself $(X_s)_{s\leqslant t}$, which we refer to as the *history* up to time t and denote by $\mathcal{F}_t = \sigma(\{X_s : s \leqslant t\})$; the corresponding filtration is known as the *natural filtration*.

Definition 4.6 (martingale difference). The time series $(X_t)_{t\in\mathbb{Z}}$ is known as a martingale-difference sequence with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ if $E|X_t| < \infty$, X_t is $\mathcal{F}_t$-measurable (*adapted*) and

$$E(X_t \mid \mathcal{F}_{t-1}) = 0, \quad \forall t \in \mathbb{Z}.$$

Obviously the unconditional mean of such a process is also zero:

$$E(X_t) = E(E(X_t \mid \mathcal{F}_{t-1})) = 0, \quad \forall t \in \mathbb{Z}.$$

Moreover, if $E(X_t^2) < \infty$ for all t, then autocovariances satisfy

$$\begin{aligned}\gamma(t,s) &= E(X_t X_s) \\ &= \begin{cases} E(E(X_t X_s \mid \mathcal{F}_{s-1})) = E(X_t E(X_s \mid \mathcal{F}_{s-1})) = 0, & t < s, \\ E(E(X_t X_s \mid \mathcal{F}_{t-1})) = E(X_s E(X_t \mid \mathcal{F}_{t-1})) = 0, & t > s. \end{cases}\end{aligned}$$

Thus a finite-variance martingale-difference sequence has zero mean and zero covariance. If the variance is constant for all t, it is a white noise process.

4.1.2 ARMA Processes

The family of classical ARMA processes are widely used in many traditional applications of time-series analysis. They are covariance-stationary processes that are constructed using white noise as a basic building block. As a general notational convention in this section and the remainder of the chapter we will denote white noise by $(\varepsilon_t)_{t\in\mathbb{Z}}$ and *strict* white noise by $(Z_t)_{t\in\mathbb{Z}}$.

Definition 4.7 (ARMA process). Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be $\mathrm{WN}(0, \sigma_\varepsilon^2)$. The process $(X_t)_{t\in\mathbb{Z}}$ is a zero-mean ARMA(p, q) process if it is a covariance-stationary process satisfying difference equations of the form

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \quad \forall t \in \mathbb{Z}. \quad (4.1)$$

(X_t) is an ARMA process with mean μ if the centred series $(X_t - \mu)_{t\in\mathbb{Z}}$ is a zero-mean ARMA(p, q) process.

Note that, according to our definition, there is no such thing as a non-covariance-stationary ARMA process. Whether the process is strictly stationary or not will depend on the exact nature of the driving white noise, also known as the process of *innovations*. If the innovations are iid, or themselves form a strictly stationary process, then the ARMA process will also be strictly stationary.

For all practical purposes we can restrict our study of ARMA processes to *causal* ARMA processes. By this we mean processes satisfying the equations (4.1), which have a representation of the form

$$X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \tag{4.2}$$

where the ψ_i are coefficients that must satisfy

$$\sum_{i=0}^{\infty} |\psi_i| < \infty. \tag{4.3}$$

Remark 4.8. The so-called absolute summability condition (4.3) is a technical condition that ensures that $E|X_t| < \infty$. This guarantees that the infinite sum in (4.2) converges absolutely, almost surely, meaning that both $\sum_{i=0}^{\infty} |\psi_i \| \varepsilon_{t-i}|$ and $\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$ are finite with probability 1 (see Brockwell and Davis 1991, Proposition 3.1.1).

We now verify by direct calculation that causal ARMA processes are indeed covariance stationary and calculate the form of their autocorrelation function, before going on to look at some simple standard examples.

Proposition 4.9. *Any process satisfying (4.2) and (4.3) is covariance stationary, with an autocorrelation function given by*

$$\rho(h) = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+|h|}}{\sum_{i=0}^{\infty} \psi_i^2}, \quad h \in \mathbb{Z}. \tag{4.4}$$

Proof. Obviously, for all t we have $E(X_t) = 0$ and $\operatorname{var}(X_t) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i^2 < \infty$, due to (4.3). Moreover, the autocovariances are given by

$$\operatorname{cov}(X_t, X_{t+h}) = E(X_t X_{t+h}) = E\bigg(\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \sum_{j=0}^{\infty} \psi_j \varepsilon_{t+h-j} \bigg).$$

Since (ε_t) is white noise, it follows that $E(\varepsilon_{t-i}\varepsilon_{t+h-j}) \neq 0 \iff j = i + h$, and hence that

$$\gamma(h) = \operatorname{cov}(X_t, X_{t+h}) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+|h|}, \quad h \in \mathbb{Z}, \tag{4.5}$$

which depends only on the lag h and not on t. The autocorrelation function follows easily. □

Example 4.10 (MA(q) process). It is clear that a pure moving-average process

$$X_t = \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \tag{4.6}$$

forms a simple example of a causal process of the form (4.2). It is easily inferred from (4.4) that the autocorrelation function is given by

$$\rho(h) = \frac{\sum_{i=0}^{q-|h|} \theta_i \theta_{i+|h|}}{\sum_{i=0}^{q} \theta_i^2}, \quad |h| \in \{0, 1, \ldots, q\},$$

where $\theta_0 = 1$. For $|h| > q$ we have $\rho(h) = 0$, and the autocorrelation function is said to *cut off* at lag q. If this feature is observed in the estimated autocorrelations of empirical data, it is often taken as an indicator of moving-average behaviour. A realization of an MA(4) process together with the theoretical form of its ACF is shown in Figure 4.1.

Example 4.11 (AR(1) process). The first-order AR process satisfies the set of difference equations

$$X_t = \phi_1 X_{t-1} + \varepsilon_t, \quad \forall t. \tag{4.7}$$

This process is causal if and only if $|\phi_1| < 1$, and this may be understood intuitively by iterating the equation (4.7) to get

$$\begin{aligned} X_t &= \phi_1(\phi_1 X_{t-2} + \varepsilon_{t-1}) + \varepsilon_{t-2} \\ &= \phi_1^{k+1} X_{t-k-1} + \sum_{i=0}^{k} \phi_1^i \varepsilon_{t-i}. \end{aligned}$$

Using more careful probabilistic arguments it may be shown that the condition $|\phi_1| < 1$ ensures that the first term disappears as $k \to \infty$ and the second term converges. The process

$$X_t = \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i} \tag{4.8}$$

turns out to be the unique solution of the defining equations (4.7). It may be easily verified that this is a process of the form (4.2) and that $\sum_{i=0}^{\infty} |\phi_1|^i = (1 - |\phi_1|)^{-1}$ so that (4.3) is satisfied. Looking at the form of the solution (4.8), we see that the AR(1) process can be represented as an MA(∞) process: an infinite-order moving-average process.

The autocovariance and autocorrelation functions of the process may be calculated from (4.5) and (4.4) to be

$$\gamma(h) = \frac{\phi_1^{|h|} \sigma_\varepsilon^2}{1 - \phi_1^2}, \quad \rho(h) = \phi_1^{|h|}, \quad h \in \mathbb{Z}.$$

Thus the ACF is exponentially decaying with possibly alternating sign. A realization of an AR(1) process together with the theoretical form of its ACF is shown in Figure 4.1.

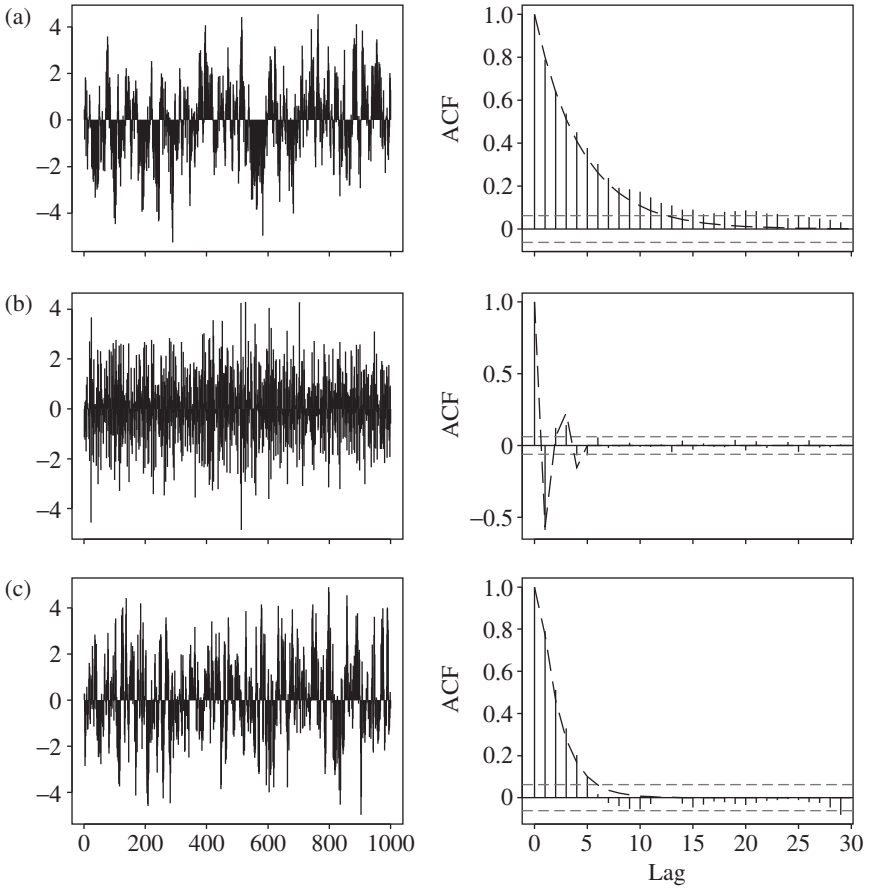

Figure 4.1. A number of simulated ARMA processes with their autocorrelation functions (dashed) and correlograms. Innovations are Gaussian. (a) AR(1), $\phi_1 = 0.8$. (b) MA(4), $\theta_1 = -0.8, 0.4, 0.2, -0.3$. (c) ARMA(1, 1), $\phi_1 = 0.6, \theta_1 = 0.5$.

Remarks on general ARMA theory. In the case of the general ARMA process of Definition 4.7, the issue of whether this process has a causal representation of the form (4.2) is resolved by the study of two polynomials in the complex plane, which are given in terms of the ARMA model parameters by

$$\tilde{\phi}(z) = 1 - \phi_1 z - \cdots - \phi_p z^p,$$
$$\tilde{\theta}(z) = 1 + \theta_1 z + \cdots + \theta_q z^q.$$

Provided that $\tilde{\phi}(z)$ and $\tilde{\theta}(z)$ have no common roots, then the ARMA process is a causal process satisfying (4.2) and (4.3) if and only if $\tilde{\phi}(z)$ has no roots in the unit circle $|z| \leqslant 1$. The coefficients ψ_i in the representation (4.2) are determined by the equation

$$\sum_{i=0}^{\infty} \psi_i z^i = \frac{\tilde{\theta}(z)}{\tilde{\phi}(z)}, \quad |z| \leqslant 1.$$

Example 4.12 (ARMA(1, 1) process). For the process given by

$$X_t - \phi_1 X_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1}, \quad \forall t \in \mathbb{Z},$$

the complex polynomials are $\tilde{\phi}(z) = 1 - \phi_1 z$ and $\tilde{\theta}(z) = 1 + \theta_1 z$, and these have no common roots provided $\phi_1 + \theta_1 \neq 0$. The solution of $\tilde{\phi}(z) = 0$ is $z = 1/\phi_1$ and this is outside the unit circle provided $|\phi_1| < 1$, so that this is the condition for causality (as in the AR(1) model of Example 4.11).

The representation (4.2) can be obtained by considering

$$\sum_{i=0}^{\infty} \psi_i z^i = \frac{1 + \theta_1 z}{1 - \phi_1 z} = (1 + \theta_1 z)(1 + \phi_1 z + \phi_1^2 z^2 + \cdots), \quad |z| \leqslant 1,$$

and is easily calculated to be

$$X_t = \varepsilon_t + (\phi_1 + \theta_1) \sum_{i=1}^{\infty} \phi_1^{i-1} \varepsilon_{t-i}. \tag{4.9}$$

Using (4.4) we may calculate that for $h \neq 0$ the ACF is

$$\rho(h) = \frac{\phi_1^{|h|-1}(\phi_1 + \theta_1)(1 + \phi_1\theta_1)}{1 + \theta_1^2 + 2\phi_1\theta_1}.$$

A realization of an ARMA(1, 1) process together with the theoretical form of its ACF is shown in Figure 4.1.

Invertibility. Equation (4.9) shows how the ARMA(1, 1) process may be thought of as an MA(∞) process. In fact, if we impose the condition $|\theta_1| < 1$, we can also express (X_t) as the AR(∞) process given by

$$X_t = \varepsilon_t + (\phi_1 + \theta_1) \sum_{i=1}^{\infty} (-\theta_1)^{i-1} X_{t-i}. \tag{4.10}$$

If we rearrange this to be an equation for ε_t, then we see that we can, in a sense, "reconstruct" the latest innovation ε_t from the entire history of the process $(X_s)_{s \leqslant t}$. The condition $|\theta_1| < 1$ is known as an *invertibility* condition, and for the general ARMA(p, q) process the invertibility condition is that $\tilde{\theta}(z)$ should have no roots in the unit circle $|z| \leqslant 1$. In practice, the models we fit to real data will be both invertible and causal solutions of the ARMA-defining equations.

Models for the conditional mean. Consider a general invertible ARMA model with non-zero mean. For what comes later it will be useful to observe that we can write such models as

$$X_t = \mu_t + \varepsilon_t, \quad \mu_t = \mu + \sum_{i=1}^{p} \phi_i (X_{t-i} - \mu) + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}. \tag{4.11}$$

Since we have assumed invertibility, the terms ε_{t-j}, and hence μ_t, can be written in terms of the infinite past of the process up to time $t - 1$; μ_t is said to be *measurable* with respect to $\mathcal{F}_{t-1} = \sigma(\{X_s : s \leqslant t - 1\})$.

If we make the assumption that the white noise $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a martingale-difference sequence (see Definition 4.6) with respect to $(\mathcal{F}_t)_{t\in\mathbb{Z}}$, then $E(X_t \mid \mathcal{F}_{t-1}) = \mu_t$. In other words, such an ARMA process can be thought of as putting a particular structure on the conditional mean μ_t of the process. ARCH and GARCH processes will later be seen to put structure on the conditional variance $\mathrm{var}(X_t \mid \mathcal{F}_{t-1})$.

ARIMA models. In traditional time-series analysis we often consider an even larger class of model known as ARIMA models, or autoregressive *integrated* moving-average models. Let ∇ denote the *difference operator*, so that for a time-series process $(Y_t)_{t\in\mathbb{Z}}$ we have $\nabla Y_t = Y_t - Y_{t-1}$. Denote repeated differencing by ∇^d, where

$$\nabla^d Y_t = \begin{cases} \nabla Y_t, & d = 1, \\ \nabla^{d-1}(\nabla Y_t) = \nabla^{d-1}(Y_t - Y_{t-1}), & d > 1. \end{cases} \tag{4.12}$$

The time series (Y_t) is said to be an ARIMA(p, d, q) process if the differenced series (X_t) given by $X_t = \nabla^d Y_t$ is an ARMA(p, q) process. For $d > 1$, ARIMA processes are non-stationary processes. They are popular in practice because the operation of differencing (once or more than once) can turn a data set that is obviously "non-stationary" into a data set that might plausibly be modelled by a stationary ARMA process. For example, if we use an ARMA(p, q) process to model daily log-returns of some price series (S_t), then we are really saying that the original logarithmic price series $(\ln S_t)$ follows an ARIMA$(p, 1, q)$ model.

When the word *integrated* is used in the context of time series it generally implies that we are looking at a non-stationary process that might be made stationary by differencing; see also the discussion of IGARCH models in Section 4.2.2.

4.1.3 Analysis in the Time Domain

We now assume that we have a sample $X_1, \ldots, X_n$ from a covariance-stationary time-series model $(X_t)_{t\in\mathbb{Z}}$. Analysis in the time domain involves calculating empirical estimates of autocovariances and autocorrelations from this random sample and using these estimates to make inferences about the serial dependence structure of the underlying process.

Correlogram. The sample autocovariances are calculated according to

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_{t+h} - \bar{X})(X_t - \bar{X}), \quad 0 \leqslant h < n,$$

where $\bar{X} = \sum_{t=1}^{n} X_t/n$ is the sample mean, which estimates μ, the mean of the time series. From these we calculate the sample ACF:

$$\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0), \quad 0 \leqslant h < n.$$

The *correlogram* is the plot $\{(h, \hat{\rho}(h)) : h = 0, 1, 2, \ldots\}$, which is designed to facilitate the interpretation of the sample ACF. Correlograms for various simulated ARMA processes are shown in Figure 4.1; note that the estimated correlations correspond reasonably closely to the theoretical ACF for these particular realizations.

To interpret such estimators of serial correlation, we need to know something about their behaviour for particular time series. The following general result is for *causal linear processes*, which are processes of the form (4.2) driven by *strict white noise*.

Theorem 4.13. *Let $(X_t)_{t\in\mathbb{Z}}$ be the linear process*

$$X_t - \mu = \sum_{i=0}^{\infty} \psi_i Z_{t-i}, \quad \text{where } \sum_{i=0}^{\infty} |\psi_i| < \infty,\ (Z_t)_{t\in\mathbb{Z}} \sim \mathrm{SWN}(0, \sigma_Z^2).$$

Suppose that either $E(Z_t^4) < \infty$ or $\sum_{i=0}^{\infty} i\psi_i^2 < \infty$. Then, for $h \in \{1, 2, \dots\}$, we have

$$\sqrt{n}(\hat{\boldsymbol{\rho}}(h) - \boldsymbol{\rho}(h)) \xrightarrow{\mathrm{d}} N_h(\mathbf{0}, W),$$

where

$$\hat{\boldsymbol{\rho}}(h) = (\hat{\rho}(1), \dots, \hat{\rho}(h))',$$
$$\boldsymbol{\rho}(h) = (\rho(1), \dots, \rho(h))',$$

N_h denotes an h-dimensional normal distribution (see Section 6.1.3), $\mathbf{0}$ is the h-dimensional vector of zeros, and W is a covariance matrix with elements

$$W_{ij} = \sum_{k=1}^{\infty} (\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k))(\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)).$$

Proof. This follows as a special case of a result in Brockwell and Davis (1991, pp. 221–223). □

The condition $\sum_{i=0}^{\infty} i\psi_i^2 < \infty$ holds for ARMA processes, so ARMA processes driven by SWN fall under the scope of this theorem (regardless of whether fourth moments exist for the innovations).

Trivially, the theorem also applies to SWN itself. For SWN we have

$$\sqrt{n}\hat{\boldsymbol{\rho}}(h) \xrightarrow{\mathrm{d}} N_h(\mathbf{0}, I_h),$$

where I_h denotes the $h \times h$ identity matrix, so for sufficiently large n the sample autocorrelations of data from an SWN process will behave like iid normal observations with mean 0 and variance $1/n$. Ninety-five per cent of the estimated correlations should lie in the interval $(-1.96/\sqrt{n}, 1.96/\sqrt{n})$, and it is for this reason that correlograms are drawn with confidence bands at these values. If more than 5% of estimated correlations lie outside these bounds, then this is considered as evidence against the null hypothesis that the data are strict white noise.

Remark 4.14. In light of the discussion of the asymptotic behaviour of sample autocorrelations for SWN, it might be asked how these estimators behave for white noise. However, this is an extremely general question because white noise encompasses a variety of possible underlying processes (including the standard ARCH and GARCH processes we later address) that only share second-order properties (finiteness of variance and lack of serial correlation). In some cases the standard

Gaussian confidence bands apply; in some cases they do not. For a GARCH process the critical issue turns out to be the heaviness of the tail of the stationary distribution (see Mikosch and Stărică (2000) for more details).

Portmanteau tests. It is often useful to combine the visual analysis of the correlogram with a formal numerical test of the strict white noise hypothesis, and a popular test is that of Ljung and Box, as applied in Section 3.1.1. Under the null hypothesis of SWN, the statistic

$$Q_{\mathrm{LB}} = n(n+2)\sum_{j=1}^{h}\frac{\hat{\rho}(j)^2}{n-j}$$

has an asymptotic chi-squared distribution with h degrees of freedom. This statistic is generally preferred to the simpler Box–Pierce statistic $Q_{\mathrm{BP}} = n\sum_{j=1}^{h}\hat{\rho}(j)^2$, which also has an asymptotic χ^2_h distribution under the null hypothesis, although the chi-squared approximation may not be so good in smaller samples. These tests are the most commonly applied portmanteau tests.

If a series of rvs forms an SWN process, then the series of absolute or squared variables must also be iid. It is a good idea to also apply the correlogram and Ljung–Box tests to absolute values as a further test of the SWN hypothesis. We prefer to perform tests of the SWN hypothesis on the absolute values rather than the squared values because the squared series is only an SWN (according to the definition we use) when the underlying series has a finite fourth moment. Daily log-return data often point to models with an infinite fourth moment.

4.1.4 Statistical Analysis of Time Series

In practice, the statistical analysis of time-series data $X_1, \ldots, X_n$ follows a programme consisting of the following stages.

Preliminary analysis. The data are plotted and the plausibility of a single stationary model is considered. There are also a number of formal numerical tests of stationarity that may be carried out at this point (see Notes and Comments for details).

Since we concentrate here on differenced logarithmic value series, we will assume that at most minor preliminary manipulation of our data is required. Classical time-series analysis has many techniques for removing *trends and seasonalities* from "non-stationary" data; these techniques are discussed in all standard texts, including Brockwell and Davis (2002) and Chatfield (2003). While certain kinds of financial time series, such as earnings time series, certainly do show seasonal patterns, we will assume that such effects are relatively minor in the kinds of daily or weekly return series that are the basis of risk-management methods. If we were to base our risk management on high-frequency data, preliminary cleaning would be more of an issue, since these show clear diurnal cycles and other deterministic features (see Dacorogna et al. 2001).

Obviously, the assumption of stationarity becomes more questionable if we take long data windows, or if we choose windows in which well-known economic policy shifts have taken place. Although the markets change constantly there will always be

a tension between our desire to use the most up-to-date data and our need to include enough data to have precision in statistical estimation. Whether half a year of data, one year, five years or ten years are appropriate will depend on the situation. It is certainly a good idea to perform a number of analyses with different data windows and to investigate the sensitivity of statistical inference to the amount of data.

Analysis in the time domain. Having settled on the data, the techniques of Section 4.1.3 come into play. By applying correlograms and portmanteau tests such as Ljung–Box to both the raw data and their absolute values, the SWN hypothesis is evaluated. If it cannot be rejected for the data in question, then the formal time-series analysis is over and simple distributional fitting could be used instead of dynamic modelling.

For daily risk-factor return series we expect to quickly reject the SWN hypothesis. Despite the fact that correlograms of the raw data may show little evidence of serial correlation, correlograms of the absolute data are likely to show evidence of strong serial dependence. In other words, the data may support a white noise model but not a strict white noise model. In this case, ARMA modelling is not required, but the volatility models of Section 4.2 may be useful.

If the correlogram does provide evidence of the kind of serial correlation patterns produced by ARMA processes, then we can attempt to fit ARMA processes to data.

Model fitting. A traditional approach to model fitting first attempts to *identify the order* of a suitable ARMA process using the correlogram and a further tool known as the partial correlogram (not described in this book but found in all standard texts). For example, the presence of a *cut-off* at lag q in the correlogram (see Example 4.10) is taken as a diagnostic for pure moving-average behaviour of order q (and similar behaviour in a partial correlogram indicates pure AR behaviour). With modern computing power it is now quite easy to simply fit a variety of MA, AR and ARMA models and to use a model-selection criterion like that of Akaike (described in Section A.3.6) to choose the "best" model. There are also automated model choice procedures such as the method of Tsay and Tiao (1984).

Sometimes there are a priori reasons for expecting certain kinds of model to be most appropriate. For example, suppose we analyse longer-period returns that overlap, as in (3.4). Consider the case where the raw data are daily returns and we build weekly returns. In (3.4) we set $h = 5$ (to get weekly returns) and $k = 1$ (to get as much data as possible). Assuming that the underlying data are genuinely from a white noise process $(X_t)_{t\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$, the weekly aggregated returns at times t and $t + l$ satisfy

$$\mathrm{cov}(X_t^{(5)}, X_{t+l}^{(5)}) = \mathrm{cov}\left(\sum_{i=0}^{4} X_{t-i}, \sum_{j=0}^{4} X_{t+l-j}\right) = \begin{cases} (5-l)\sigma^2, & l = 0, \ldots, 4, \\ 0, & l \geqslant 5, \end{cases}$$

so that the overlapping returns have the correlation structure of an MA(4) process, and this would be a natural choice of time-series model for them.

Having chosen the model to fit, there are a number of possible fitting methods, including specialized methods for AR processes, such as Yule–Walker, that make

minimal assumptions concerning the distribution of the white noise innovations; we refer to the standard time-series literature for more details. In Section 4.2.4 we discuss the method of (conditional) maximum likelihood, which may be used to fit ARMA models with (or without) GARCH errors to data.

Residual analysis and model comparison. Recall the representation of a causal and invertible ARMA process in (4.11), and suppose we have fitted such a process and estimated the parameters ϕ_i and θ_j. The residuals are inferred values $\hat{\varepsilon}_t$ for the unobserved innovations ε_t and they are calculated recursively from the data and the fitted model using the equations

$$\hat{\varepsilon}_t = X_t - \hat{\mu}_t, \quad \hat{\mu}_t = \hat{\mu} + \sum_{i=1}^{p} \hat{\phi}_i (X_{t-i} - \hat{\mu}) + \sum_{j=1}^{q} \hat{\theta}_j \hat{\varepsilon}_{t-j}, \tag{4.13}$$

where the values $\hat{\mu}_t$ are sometimes known as the *fitted values*. Obviously, we have a problem calculating the first few values of $\hat{\varepsilon}_t$ due to the finiteness of our data sample and the infinite nature of the recursions (4.13). One of many possible solutions is to set $\hat{\varepsilon}_{-q+1} = \hat{\varepsilon}_{-q+2} = \cdots = \hat{\varepsilon}_0 = 0$ and $X_{-p+1} = X_{-p+2} = \cdots = X_0 = \bar{X}$ and then to use (4.13) for $t = 1, \ldots, n$. Since the first few values will be influenced by these starting values, they might be ignored in later analyses.

The residuals $(\hat{\varepsilon}_t)$ should behave like a realization of a white noise process, since this is our model assumption for the innovations, and this can be assessed by constructing their correlogram. If there is still evidence of serial correlation in the correlogram, then this suggests that a good ARMA model has not yet been found. Moreover, we can use portmanteau tests to test formally that the residuals behave like a realization of a strict white noise process. If the residuals behave like SWN, then no further time-series modelling is required; if they behave like WN but not SWN, then the volatility models of Section 4.2 may be required.

It is usually possible to find more than one reasonable ARMA model for the data, and formal model-comparison techniques may be required to decide on an overall best model or models. The Akaike information criterion described in Section A.3.6 might be used, or one of a number of variants on this criterion that are often preferred for time series (see Brockwell and Davis 2002, Section 5.5.2).

4.1.5 Prediction

There are many approaches to the forecasting or prediction of time series, and we summarize two that extend easily to the case of GARCH models. The first strategy makes use of fitted ARMA (or ARIMA) models and is sometimes called the *Box–Jenkins approach* (Box and Jenkins 1970). The second strategy is a model-free approach to forecasting known as *exponential smoothing*, which is related to the exponentially weighted moving-average technique for predicting volatility.

Prediction using ARMA models. Consider the invertible ARMA model and its representation in (4.11). Let $\mathcal{F}_t$ denote the history of the process up to and including time t, as before, and assume that the innovations $(\varepsilon_t)_{t\in\mathbb{Z}}$ have the martingale-difference property with respect to $(\mathcal{F}_t)_{t\in\mathbb{Z}}$.

For the prediction problem it will be convenient to denote our sample of n data by $X_{t-n+1}, \ldots, X_t$. We assume that these are realizations of rvs following a particular ARMA model. Our aim is to predict X_{t+1} or, more generally, X_{t+h}, and we denote our prediction by $P_t X_{t+h}$. The method we describe assumes that we have access to the infinite history of the process up to time t and derives a formula that is then approximated for our finite sample.

As a predictor of X_{t+h} we use the conditional expectation $E(X_{t+h} \mid \mathcal{F}_t)$. Among all predictions $P_t X_{t+h}$ based on the infinite history of the process up to time t, this predictor minimizes the mean squared prediction error $E((X_{t+h} - P_t X_{t+h})^2)$.

The basic idea is that, for $h \geqslant 1$, the prediction $E(X_{t+h} \mid \mathcal{F}_t)$ is recursively evaluated in terms of $E(X_{t+h-1} \mid \mathcal{F}_t)$. We use the fact that $E(\varepsilon_{t+h} \mid \mathcal{F}_t) = 0$ (the martingale-difference property of innovations) and that the rvs $(X_s)_{s \leqslant t}$ and $(\varepsilon_s)_{s \leqslant t}$ are "known" at time t. The assumption of invertibility (4.10) ensures that the innovation ε_t can be written as a function of the infinite history of the process $(X_s)_{s \leqslant t}$. To illustrate the approach it will suffice to consider an ARMA(1, 1) model, the generalization to ARMA(p, q) models following easily.

Example 4.15 (prediction for the ARMA(1, 1) model). Suppose an ARMA(1, 1) model of the form (4.11) has been fitted to the data, and its parameters μ, ϕ_1 and θ_1 have been determined. Our one-step prediction for X_{t+1} is

$$E(X_{t+1} \mid \mathcal{F}_t) = \mu_{t+1} = \mu + \phi_1(X_t - \mu) + \theta_1 \varepsilon_t,$$

since $E(\varepsilon_{t+1} \mid \mathcal{F}_t) = 0$. For a two-step prediction we get

$$\begin{aligned} E(X_{t+2} \mid \mathcal{F}_t) = E(\mu_{t+2} \mid \mathcal{F}_t) &= \mu + \phi_1(E(X_{t+1} \mid \mathcal{F}_t) - \mu) \\ &= \mu + \phi_1^2(X_t - \mu) + \phi_1 \theta_1 \varepsilon_t, \end{aligned}$$

and in general we have

$$E(X_{t+h} \mid \mathcal{F}_t) = \mu + \phi_1^h(X_t - \mu) + \phi_1^{h-1} \theta_1 \varepsilon_t.$$

Without knowing all historical values of $(X_s)_{s \leqslant t}$ this predictor cannot be evaluated exactly, because we do not know ε_t exactly, but it can be accurately approximated if n is reasonably large. The easiest way of doing this is to substitute the model residual $\hat{\varepsilon}_t$ calculated from (4.13) for ε_t. Note that $\lim_{h \to \infty} E(X_{t+h} \mid \mathcal{F}_t) = \mu$, almost surely, so that the prediction converges to the estimate of the unconditional mean of the process for longer time horizons.

Exponential smoothing. This is a popular technique that is used for both prediction of time-series and trend estimation. Here we do not necessarily assume that the data come from a stationary model, although we do assume that there is no deterministic seasonal component in the model. In general, the method is less well suited to return series with frequently changing signs and is better suited to undifferenced price or value series. It forms the basis of a very common method of volatility prediction (see Section 4.2.5).

Suppose our data represent realizations of rvs $Y_{t-n+1}, \ldots, Y_t$, considered without reference to any concrete parametric model. As a forecast for Y_{t+1} we use a prediction of the form

$$P_t Y_{t+1} = \sum_{i=0}^{n-1} \lambda(1-\lambda)^i Y_{t-i}, \quad \text{where } 0 < \lambda < 1.$$

Thus we weight the data from most recent to most distant with a sequence of exponentially decreasing weights that sum to almost one. It is easily calculated that

$$\begin{aligned} P_t Y_{t+1} = \sum_{i=0}^{n-1} \lambda(1-\lambda)^i Y_{t-i} &= \lambda Y_t + (1-\lambda) \sum_{j=0}^{n-2} \lambda(1-\lambda)^j Y_{t-1-j} \\ &= \lambda Y_t + (1-\lambda) P_{t-1} Y_t, \end{aligned} \tag{4.14}$$

so that the prediction at time t is obtained from the prediction at time $t-1$ by a simple recursive scheme. The choice of λ is subjective; the larger the value, the more weight is put on the most recent observation. Empirical validation studies with different data sets can be used to determine a value of λ that gives good results; Chatfield (2003) reports that values between 0.1 and 0.3 are commonly used in practice.

Note that, although the method is commonly seen as a model-free forecasting technique, it can be shown to be the natural prediction method based on conditional expectation for a non-stationary ARIMA(0, 1, 1) model.

Notes and Comments

There are many texts covering the subject of classical time-series analysis, including Box and Jenkins (1970), Priestley (1981), Abraham and Ledolter (1983), Brockwell and Davis (1991, 2002), Hamilton (1994) and Chatfield (2003). Our account of basic concepts, ARMA models and analysis in the time domain closely follows Brockwell and Davis (1991), which should be consulted for the rigorous background to ideas we can only summarize. We have not discussed analysis of time series in the frequency domain, which is less common for financial time series; for this subject see, again, Brockwell and Davis (1991) or Priestley (1981).

For more on tests of the strict white noise hypothesis (that is, tests of randomness), see Brockwell and Davis (2002). Original references for the Box–Pierce and Ljung–Box tests are Box and Pierce (1970) and Ljung and Box (1978).

There is a large econometrics literature on tests of stationarity and unit-root tests, where the latter are effectively tests of the null hypothesis of non-stationary random-walk behaviour. Particular examples are the Dickey–Fuller and Phillips–Perron unit-root tests (Dickey and Fuller 1979; Phillips and Perron 1988) and the KPSS test of stationarity (Kwiatkowski et al. 1992).

There is a vast literature on forecasting and prediction in linear models. A good non-mathematical introduction is found in Chatfield (2003). The approach we describe based on the infinite history of the time series is discussed in greater detail in Hamilton (1994). Brockwell and Davis (2002) concentrate on exact linear

prediction methods for finite samples. A general review of exponential smoothing is found in Gardner (1985).

4.2 GARCH Models for Changing Volatility

The most important models for daily risk-factor return series are addressed in this section. We give definitions of ARCH (autoregressive conditionally heteroscedastic) and GARCH (generalized ARCH) models and discuss some of their mathematical properties before going on to talk about their use in practice.

4.2.1 ARCH Processes

Definition 4.16. Let $(Z_t)_{t\in\mathbb{Z}}$ be SWN(0, 1). The process $(X_t)_{t\in\mathbb{Z}}$ is an ARCH(p) process if it is strictly stationary and if it satisfies, for all $t \in \mathbb{Z}$ and some strictly positive-valued process $(\sigma_t)_{t\in\mathbb{Z}}$, the equations

$$X_t = \sigma_t Z_t, \tag{4.15}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2, \tag{4.16}$$

where $\alpha_0 > 0$ and $\alpha_i \geqslant 0$, $i = 1, \ldots, p$.

Let $\mathcal{F}_t = \sigma(\{X_s : s \leqslant t\})$ again denote the sigma algebra representing the history of the process up to time t, so that $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ is the natural filtration. The construction (4.16) ensures that σ_t is *measurable* with respect to $\mathcal{F}_{t-1}$, and the process $(\sigma_t)_{t\in\mathbb{Z}}$ is said to be previsible. This allows us to calculate that, provided $E(|X_t|) < \infty$,

$$E(X_t \mid \mathcal{F}_{t-1}) = E(\sigma_t Z_t \mid \mathcal{F}_{t-1}) = \sigma_t E(Z_t \mid \mathcal{F}_{t-1}) = \sigma_t E(Z_t) = 0, \tag{4.17}$$

so that the ARCH process has the martingale-difference property with respect to $(\mathcal{F}_t)_{t\in\mathbb{Z}}$. If the process is covariance stationary, it is simply a white noise, as discussed in Section 4.1.1.

Remark 4.17. Note that the independence of Z_t and $\mathcal{F}_{t-1}$ that we have assumed above follows from the fact that an ARCH process must be causal, i.e. the equations (4.15) and (4.16) must have a solution of the form $X_t = f(Z_t, Z_{t-1}, \ldots)$ for some f, so that Z_t is independent of previous values of the process. This contrasts with ARMA models, where the equations can have non-causal solutions (see Brockwell and Davis 1991, Example 3.1.2).

If we simply assume that the process is a covariance-stationary white noise (for which we will give a condition in Proposition 4.18), then $E(X_t^2) < \infty$ and

$$\operatorname{var}(X_t \mid \mathcal{F}_{t-1}) = E(\sigma_t^2 Z_t^2 \mid \mathcal{F}_{t-1}) = \sigma_t^2 \operatorname{var}(Z_t) = \sigma_t^2.$$

Thus the model has the interesting property that its conditional standard deviation σ_t, or *volatility*, is a continually changing function of the previous squared values of the process. If one or more of $|X_{t-1}|, \ldots, |X_{t-p}|$ are particularly *large*, then X_t is effectively drawn from a distribution with large variance, and may itself be large; in

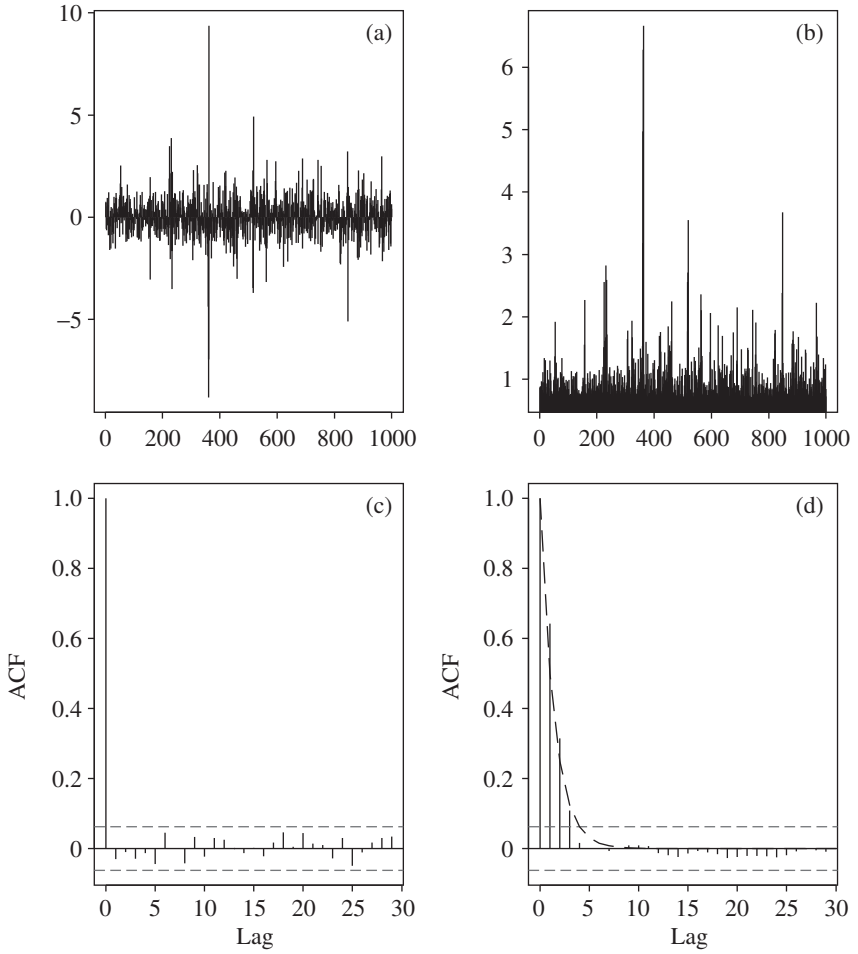

Figure 4.2. A simulated ARCH(1) process with Gaussian innovations and parameters $\alpha_0 = \alpha_1 = 0.5$: (a) the realization of the process; (b) the realization of the volatility; and correlograms of (c) the raw and (d) the squared values. The process is covariance stationary with unit variance and a finite fourth moment (since $\alpha_1 < 1/\sqrt{3}$) and the squared values follow an AR(1) process. The true form of the ACF of the squared values is represented by the dashed line in the correlogram.

this way the model generates volatility clusters. The name ARCH refers to this structure: the model is autoregressive, since X_t clearly depends on previous X_{t-i}, and conditionally heteroscedastic, since the conditional variance changes continually.

The distribution of the innovations $(Z_t)_{t\in\mathbb{Z}}$ can in principle be any zero-mean, unit-variance distribution. For statistical fitting purposes we may or may not choose to actually specify the distribution, depending on whether we implement a maximum likelihood (ML), quasi-maximum likelihood (QML) or non-parametric fitting method (see Section 4.2.4). For ML the most common choices are standard normal innovations or scaled t innovations. By the latter we mean that $Z_t \sim t_1(\nu, 0, (\nu-2)/\nu)$, in the notation of Example 6.7, so that the variance

of the distribution is one. We keep these choices in mind when discussing further theoretical properties of ARCH and GARCH models.

The ARCH(1) *model.* In the rest of this section we analyse some of the properties of the ARCH(1) model. These properties extend to the whole class of ARCH and GARCH models but are most easily introduced in the simplest case. A simulated realization of an ARCH(1) process with Gaussian innovations and the corresponding realization of the volatility process are shown in Figure 4.2.

Using $X_t^2 = \sigma_t^2 Z_t^2$ and (4.16) in the case $p = 1$, we deduce that the squared ARCH(1) process satisfies

$$X_t^2 = \alpha_0 Z_t^2 + \alpha_1 Z_t^2 X_{t-1}^2. \tag{4.18}$$

A detailed mathematical analysis of the ARCH(1) model involves the study of equation (4.18), which is a *stochastic recurrence equation* (SRE). Much as for the AR(1) model in Example 4.11, we would like to know when this equation has stationary solutions expressed in terms of the infinite history of the innovations, i.e. solutions of the form $X_t^2 = f(Z_t, Z_{t-1}, \dots)$.

For ARCH models we have to distinguish carefully between solutions that are covariance stationary and solutions that are only strictly stationary. It is possible to have ARCH(1) models with infinite variance, which obviously cannot be covariance stationary.

Stochastic recurrence relations. The detailed theory required to analyse stochastic recurrence relations of the form (4.18) is outside the scope of this book, and we give only brief notes to indicate the ideas involved. Our treatment is based on Brandt (1986), Mikosch (2003) and Mikosch (2013); see Notes and Comments at the end of this section for further references.

Equation (4.18) is a particular example of a class of recurrence equations of the form

$$Y_t = A_t Y_{t-1} + B_t, \tag{4.19}$$

where $(A_t)_{t\in\mathbb{Z}}$ and $(B_t)_{t\in\mathbb{Z}}$ are sequences of iid rvs. Sufficient conditions for a solution are that

$$E(\ln^+ |B_t|) < \infty \quad \text{and} \quad E(\ln |A_t|) < 0, \tag{4.20}$$

where $\ln^+ x = \max(0, \ln x)$. The unique solution is given by

$$Y_t = B_t + \sum_{i=1}^{\infty} B_{t-i} \prod_{j=0}^{i-1} A_{t-j}, \tag{4.21}$$

where the sum converges absolutely, almost surely.

We can develop some intuition for the conditions (4.20) and the form of the solution (4.21) by iterating equation (4.19) k times to obtain

$$\begin{aligned} Y_t &= A_t(A_{t-1}Y_{t-2} + B_{t-1}) + B_t \\ &= B_t + \sum_{i=1}^{k} B_{t-i} \prod_{j=0}^{i-1} A_{t-j} + Y_{t-k-1} \prod_{i=0}^{k} A_{t-i}. \end{aligned}$$

The conditions (4.20) ensure that the middle term on the right-hand side converges absolutely and the final term disappears. In particular, note that

$$\frac{1}{k+1}\sum_{i=0}^{k}\ln|A_{t-i}| \xrightarrow{\text{a.s.}} E(\ln|A_t|) < 0$$

by the strong law of large numbers. So

$$\prod_{i=0}^{k}|A_{t-i}| = \exp\left(\sum_{i=0}^{k}\ln|A_{t-i}|\right) \xrightarrow{\text{a.s.}} 0,$$

which shows the importance of the $E(\ln|A_t|) < 0$ condition. The solution (4.21) to the SRE is a strictly stationary process (being a function of iid variables $(A_s, B_s)_{s\leqslant t}$), and the $E(\ln|A_t|) < 0$ condition turns out to be the key to the strict stationarity of ARCH and GARCH models.

Stationarity of ARCH(1). The squared ARCH(1) model (4.18) is an SRE of the form (4.19) with $A_t = \alpha_1 Z_t^2$ and $B_t = \alpha_0 Z_t^2$. Thus the conditions in (4.20) translate into the requirements that $E(\ln^+|\alpha_0 Z_t^2|) < \infty$, which is automatically true for the ARCH(1) process as we have defined it, and $E(\ln(\alpha_1 Z_t^2)) < 0$. This is the condition for a strictly stationary solution of the ARCH(1) equations, and it can be shown that it is in fact a necessary and sufficient condition for strict stationarity (see Bougerol and Picard 1992). From (4.21), the solution of equation (4.18) takes the form

$$X_t^2 = \alpha_0\sum_{i=0}^{\infty}\alpha_1^i\prod_{j=0}^{i}Z_{t-j}^2. \tag{4.22}$$

If the (Z_t) are standard normal innovations, then the condition for a strictly stationary solution is approximately $\alpha_1 < 3.562$; perhaps somewhat surprisingly, if the (Z_t) are scaled t innovations with four degrees of freedom and variance 1, the condition is $\alpha_1 < 5.437$. Strict stationarity depends on the distribution of the innovations but covariance stationarity does not; the necessary and sufficient condition for covariance stationarity is always $\alpha_1 < 1$, as we now prove.

Proposition 4.18. *The* ARCH(1) *process is a covariance-stationary white noise process if and only if* $\alpha_1 < 1$. *The variance of the covariance-stationary process is given by* $\alpha_0/(1-\alpha_1)$.

Proof. Assuming covariance stationarity, it follows from (4.18) and $E(Z_t^2) = 1$ that

$$\sigma_x^2 = E(X_t^2) = \alpha_0 + \alpha_1 E(X_{t-1}^2) = \alpha_0 + \alpha_1\sigma_x^2.$$

Clearly, $\sigma_x^2 = \alpha_0/(1-\alpha_1)$ and we must have $\alpha_1 < 1$.

Conversely, if $\alpha_1 < 1$, then, by Jensen's inequality,

$$E(\ln(\alpha_1 Z_t^2)) \leqslant \ln(E(\alpha_1 Z_t^2)) = \ln(\alpha_1) < 0,$$

and we can use (4.22) to calculate that

$$E(X_t^2) = \alpha_0\sum_{i=0}^{\infty}\alpha_1^i = \frac{\alpha_0}{1-\alpha_1}.$$

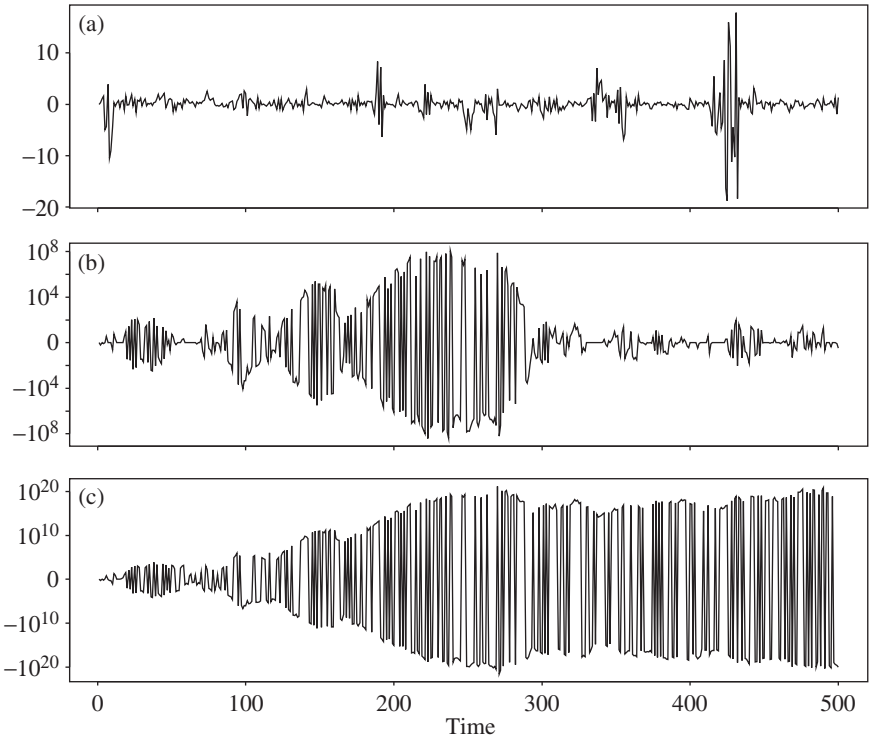

Figure 4.3. (a), (b) Strictly stationary ARCH(1) models with Gaussian innovations that are not covariance stationary ($\alpha_1 = 1.2$ and $\alpha_1 = 3$, respectively). (c) A non-stationary (explosive) process generated by the ARCH(1) equations with $\alpha_1 = 4$. Note that (b) and (c) use a special double-logarithmic y-axis where all values less than one in modulus are plotted at zero.

The process $(X_t)_{t\in\mathbb{Z}}$ is a martingale difference with a finite, non-time-dependent second moment. Hence it is a white noise process. □

See Figure 4.3 for examples of non-covariance-stationary ARCH(1) models as well as an example of a non-stationary (explosive) process generated by the ARCH(1) equations. The process in Figure 4.2 is covariance stationary.

On the stationary distribution of X_t. It is clear from (4.22) that the distribution of the (X_t) in an ARCH(1) model bears a complicated relationship to the distribution of the innovations (Z_t). Even if the innovations are Gaussian, the stationary distribution of the time series is not Gaussian, but rather a leptokurtic distribution with more slowly decaying tails.

Moreover, from (4.15) we see that the distribution of X_t is a normal mixture distribution of the kind discussed in Section 6.2. Its distribution depends on the distribution of σ_t, which has no simple form.

Proposition 4.19. *For $m \geqslant 1$, the strictly stationary* ARCH(1) *process has finite moments of order $2m$ if and only if $E(Z_t^{2m}) < \infty$ and $\alpha_1 < (E(Z_t^{2m}))^{-1/m}$.*

Proof. We rewrite (4.22) in the form $X_t^2 = Z_t^2 \sum_{i=0}^{\infty} Y_{t,i}$ for positive rvs $Y_{t,i} = \alpha_0 \alpha_1^i \prod_{j=1}^{i} Z_{t-j}^2$, $i \geqslant 1$, and $Y_{t,0} = \alpha_0$. For $m \geqslant 1$ the following inequalities hold (the latter being Minkowski's inequality):

$$E(Y_{t,1}^m) + E(Y_{t,2}^m) \leqslant E((Y_{t,1} + Y_{t,2})^m) \leqslant ((E(Y_{t,1}^m))^{1/m} + (E(Y_{t,2}^m))^{1/m})^m.$$

Since

$$E(X_t^{2m}) = E(Z_t^{2m}) E\left(\left(\sum_{i=0}^{\infty} Y_{t,i}\right)^m\right),$$

it follows that

$$E(Z_t^{2m}) \sum_{i=0}^{\infty} E(Y_{t,i}^m) \leqslant E(X_t^{2m}) \leqslant E(Z_t^{2m}) \left(\sum_{i=0}^{\infty} (E(Y_{t,i}^m))^{1/m}\right)^m.$$

Since $E(Y_{t,i}^m) = \alpha_0^m \alpha_1^{im} (E(Z_t^{2m}))^i$, it may be deduced that all three quantities are finite if and only if $E(Z_t^{2m}) < \infty$ and $\alpha_1^m E(Z_t^{2m}) < 1$. □

For example, for a finite fourth moment ($m = 2$) we require $\alpha_1 < 1/\sqrt{3}$ in the case of Gaussian innovations and $\alpha_1 < 1/\sqrt{6}$ in the case of t innovations with six degrees of freedom; for t innovations with four degrees of freedom, the fourth moment is undefined.

Assuming the existence of a finite fourth moment, it is easy to calculate its value, and also that of the kurtosis of the process. We square both sides of (4.18), take expectations of both sides and then solve for $E(X_t^4)$ to obtain

$$E(X_t^4) = \frac{\alpha_0^2 E(Z_t^4)(1 - \alpha_1^2)}{(1 - \alpha_1)^2 (1 - \alpha_1^2 E(Z_t^4))}.$$

The kurtosis of the stationary distribution κ_X can then be calculated to be

$$\kappa_X = \frac{E(X_t^4)}{E(X_t^2)^2} = \frac{\kappa_Z (1 - \alpha_1^2)}{(1 - \alpha_1^2 \kappa_Z)},$$

where $\kappa_Z = E(Z_t^4)$ denotes the kurtosis of the innovations. Clearly, when $\kappa_Z > 1$, the kurtosis of the stationary distribution is inflated in comparison with that of the innovation distribution; for Gaussian or t innovations, $\kappa_X > 3$, so the stationary distribution is leptokurtic. The kurtosis of the process in Figure 4.2 is 9.

Parallels with the AR(1) *process.* We now turn our attention to the serial dependence structure of the squared series in the case of covariance stationarity ($\alpha_1 < 1$). We write the squared process as

$$X_t^2 = \sigma_t^2 Z_t^2 = \sigma_t^2 + \sigma_t^2 (Z_t^2 - 1). \tag{4.23}$$

Setting $V_t = \sigma_t^2 (Z_t^2 - 1)$, we note that $(V_t)_{t \in \mathbb{Z}}$ forms a martingale-difference series, since $E|V_t| < \infty$ and $E(V_t \mid \mathcal{F}_{t-1}) = \sigma_t^2 E(Z_t^2 - 1) = 0$. Now we rewrite (4.23) as $X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + V_t$, and observe that this closely resembles an AR(1) process for X_t^2, except that V_t is not necessarily a white noise process. If we restrict our attention to processes where $E(X_t^4)$ is finite, then V_t has a finite and constant second

moment and is a white noise process. Under this assumption, X_t^2 is an AR(1) process, according to Definition 4.7, of the form

$$\left(X_t^2 - \frac{\alpha_0}{1-\alpha_1}\right) = \alpha_1\left(X_{t-1}^2 - \frac{\alpha_0}{1-\alpha_1}\right) + V_t.$$

It has mean $\alpha_0/(1-\alpha_1)$ and we can use Example 4.11 to conclude that the autocorrelation function is $\rho(h) = \alpha_1^{|h|}$, $h \in \mathbb{Z}$. Figure 4.2 shows an example of an ARCH(1) process with finite fourth moment whose squared values follow an AR(1) process.

4.2.2 GARCH Processes

Definition 4.20. Let $(Z_t)_{t\in\mathbb{Z}}$ be SWN(0, 1). The process $(X_t)_{t\in\mathbb{Z}}$ is a GARCH(p, q) process if it is strictly stationary and if it satisfies, for all $t \in \mathbb{Z}$ and some strictly positive-valued process $(\sigma_t)_{t\in\mathbb{Z}}$, the equations

$$X_t = \sigma_t Z_t, \qquad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{4.24}$$

where $\alpha_0 > 0$, $\alpha_i \geqslant 0$, $i = 1, \ldots, p$, and $\beta_j \geqslant 0$, $j = 1, \ldots, q$.

The GARCH processes are *generalized* ARCH processes in the sense that the squared volatility σ_t^2 is allowed to depend on previous squared volatilities, as well as previous squared values of the process.

The GARCH(1, 1) *model.* In practice, low-order GARCH models are most widely used and we will concentrate on the GARCH(1, 1) model. In this model periods of high volatility tend to be *persistent*, since $|X_t|$ has a chance of being large if *either* $|X_{t-1}|$ is large *or* σ_{t-1} is large; the same effect can be achieved in ARCH models of high order, but lower-order GARCH models achieve this effect more parsimoniously. A simulated realization of a GARCH(1, 1) process with Gaussian innovations and its volatility are shown in Figure 4.4; in comparison with the ARCH(1) model of Figure 4.2, it is clear that the volatility persists longer at higher levels before decaying to lower levels.

Stationarity. It follows from (4.24) that for a GARCH(1, 1) model we have

$$\sigma_t^2 = \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta_1)\sigma_{t-1}^2, \tag{4.25}$$

which is again an SRE of the form $Y_t = A_t Y_{t-1} + B_t$, as in (4.19). This time it is an SRE for $Y_t = \sigma_t^2$ rather than X_t^2, but its analysis follows easily from the ARCH(1) case.

The condition $E(\ln|A_t|) < 0$ for a strictly stationary solution of (4.19) translates to the condition $E(\ln(\alpha_1 Z_t^2 + \beta_1)) < 0$ for (4.25), and the general solution (4.21) becomes

$$\sigma_t^2 = \alpha_0 + \alpha_0 \sum_{i=1}^{\infty} \prod_{j=1}^{i} (\alpha_1 Z_{t-j}^2 + \beta_1). \tag{4.26}$$

If $(\sigma_t^2)_{t\in\mathbb{Z}}$ is a strictly stationary process, then so is $(X_t)_{t\in\mathbb{Z}}$, since $X_t = \sigma_t Z_t$ and $(Z_t)_{t\in\mathbb{Z}}$ is simply strict white noise. The solution of the GARCH(1, 1) defining equations is then

$$X_t = Z_t \sqrt{\alpha_0 \Big(1 + \sum_{i=1}^{\infty} \prod_{j=1}^{i} (\alpha_1 Z_{t-j}^2 + \beta_1) \Big)}, \tag{4.27}$$

and we can use this to derive the condition for covariance stationarity.

Proposition 4.21. *The* GARCH(1, 1) *process is a covariance-stationary white noise process if and only if* $\alpha_1 + \beta_1 < 1$. *The variance of the covariance-stationary process is given by* $\alpha_0/(1 - \alpha_1 - \beta_1)$.

Proof. We use a similar argument to Proposition 4.18 and make use of (4.27). □

Fourth moments and kurtosis. Using a similar approach to Proposition 4.19 we can use (4.27) to derive conditions for the existence of higher moments of a covariance-stationary GARCH(1, 1) process. For the existence of a fourth moment, a necessary and sufficient condition is that $E((\alpha_1 Z_t^2 + \beta_1)^2) < 1$, or alternatively that

$$(\alpha_1 + \beta_1)^2 < 1 - (\kappa_Z - 1)\alpha_1^2.$$

Assuming this to be true, we calculate the fourth moment and kurtosis of X_t. We square both sides of (4.25) and take expectations to obtain

$$E(\sigma_t^4) = \alpha_0^2 + (\alpha_1^2 \kappa_Z + \beta_1^2 + 2\alpha_1\beta_1) E(\sigma_t^4) + 2\alpha_0(\alpha_1 + \beta_1) E(\sigma_t^2).$$

Solving for $E(\sigma_t^4)$, recalling that $E(\sigma_t^2) = E(X_t^2) = \alpha_0/(1 - \alpha_1 - \beta_1)$, and setting $E(X_t^4) = \kappa_Z E(\sigma_t^4)$, we obtain

$$E(X_t^4) = \frac{\alpha_0^2 \kappa_Z (1 - (\alpha_1 + \beta_1)^2)}{(1 - \alpha_1 - \beta_1)^2 (1 - \alpha_1^2 \kappa_Z - \beta_1^2 - 2\alpha_1\beta_1)},$$

from which it follows that

$$\kappa_X = \frac{\kappa_Z (1 - (\alpha_1 + \beta_1)^2)}{(1 - (\alpha_1 + \beta_1)^2 - (\kappa_Z - 1)\alpha_1^2)}.$$

Again it is clear that the kurtosis of X_t is greater than that of Z_t whenever $\kappa_Z > 1$, such as for Gaussian and scaled t innovations. The kurtosis of the GARCH(1, 1) model in Figure 4.4 is 3.77.

Parallels with the ARMA(1, 1) *process.* Using the same representation as in equation (4.23), the covariance-stationary GARCH(1, 1) process may be written as

$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + V_t,$$

where V_t is a martingale difference, given by $V_t = \sigma_t^2(Z_t^2 - 1)$. Since $\sigma_{t-1}^2 = X_{t-1}^2 - V_{t-1}$, we may write

$$X_t^2 = \alpha_0 + (\alpha_1 + \beta_1) X_{t-1}^2 - \beta_1 V_{t-1} + V_t, \tag{4.28}$$

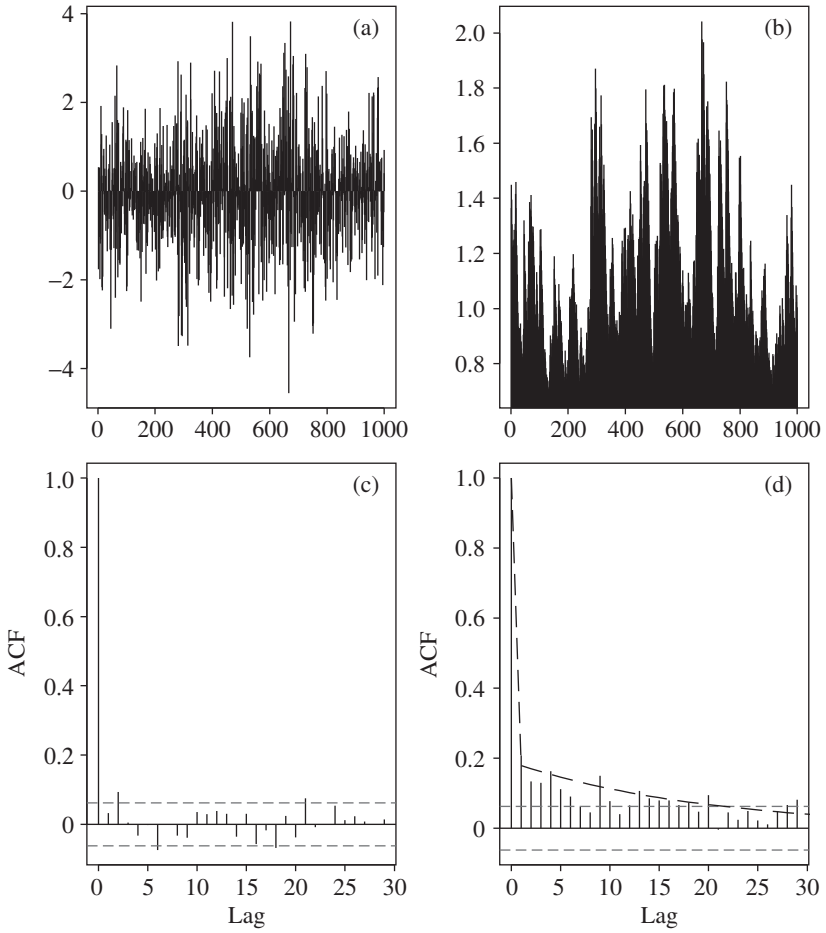

Figure 4.4. A GARCH(1, 1) process with Gaussian innovations and parameters $\alpha_0 = 0.5$, $\alpha_1 = 0.1, \beta_1 = 0.85$: (a) the realization of the process; (b) the realization of the volatility; and correlograms of (c) the raw and (d) the squared values. The process is covariance stationary with unit variance and a finite fourth moment and the squared values follow an ARMA(1, 1) process. The true form of the ACF of the squared values is shown by a dashed line in the correlogram.

which begins to resemble an ARMA(1, 1) process for X_t^2. If we further assume that $E(X_t^4) < \infty$, then, recalling that $\alpha_1 + \beta_1 < 1$, we have formally that

$$\left(X_t^2 - \frac{\alpha_0}{1-\alpha_1-\beta_1}\right) = (\alpha_1+\beta_1)\left(X_{t-1}^2 - \frac{\alpha_0}{1-\alpha_1-\beta_1}\right) - \beta_1 V_{t-1} + V_t$$

is an ARMA(1, 1) process. Figure 4.4 shows an example of a GARCH(1, 1) process with finite fourth moment whose squared values follow an ARMA(1, 1) process.

The GARCH(p, q) *model.* Higher-order ARCH and GARCH models have the same general behaviour as ARCH(1) and GARCH(1, 1), but their mathematical analysis becomes more tedious. The condition for a strictly stationary solution of the

defining SRE has been derived by Bougerol and Picard (1992), but it is complicated. The necessary and sufficient condition that this solution is covariance stationary is $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$.

A squared GARCH(p, q) process has the structure

$$X_t^2 = \alpha_0 + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) X_{t-i}^2 - \sum_{j=1}^{q} \beta_j V_{t-j} + V_t,$$

where $\alpha_i = 0$ for $i = p+1, \ldots, q$ if $q > p$, or $\beta_j = 0$ for $j = q+1, \ldots, p$ if $p > q$. This resembles the ARMA$(\max(p, q), q)$ process and is formally such a process provided $E(X_t)^4 < \infty$.

Integrated GARCH. The study of integrated GARCH (or IGARCH) processes has been motivated by the fact that, in some applications of GARCH modelling to daily or higher-frequency risk-factor return series, the estimated ARCH and GARCH coefficients $(\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$ are observed to sum to a number very close to 1, and sometimes even slightly larger than 1. In a model where $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j \geqslant 1$, the process has *infinite variance* and is thus non-covariance-stationary. The special case where $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1$ is known as IGARCH and has received some attention.

For simplicity, consider the IGARCH(1, 1) model. We use (4.28) to conclude that the squared process must satisfy

$$\nabla X_t^2 = X_t^2 - X_{t-1}^2 = \alpha_0 - (1 - \alpha_1) V_{t-1} + V_t,$$

where V_t is a noise sequence defined by $V_t = \sigma_t^2 (Z_t^2 - 1)$ and $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + (1 - \alpha_1)\sigma_{t-1}^2$. This equation is reminiscent of an ARIMA(0, 1, 1) model (see (4.12)) for X_t^2, although the noise V_t is not white noise, nor is it strictly speaking a martingale difference according to Definition 4.6. $E(V_t \mid \mathcal{F}_{t-1})$ is undefined since $E(\sigma_t^2) = E(X_t^2) = \infty$, and therefore $E|V_t|$ is undefined.

4.2.3 Simple Extensions of the GARCH Model

Many variants and extensions of the basic GARCH model have been proposed. We mention only a few (see Notes and Comments for further reading).

ARMA models with GARCH errors. We have seen that ARMA processes are driven by a white noise $(\varepsilon_t)_{t \in \mathbb{Z}}$ and that a covariance-stationary GARCH process is an example of a white noise. In this section we put the ARMA and GARCH models together by setting the ARMA error ε_t equal to $\sigma_t Z_t$, where σ_t follows a GARCH volatility specification in terms of historical values of ε_t. This gives us a flexible family of ARMA models with GARCH errors that combines the features of both model classes.

Definition 4.22. Let $(Z_t)_{t \in \mathbb{Z}}$ be SWN(0, 1). The process $(X_t)_{t \in \mathbb{Z}}$ is said to be an ARMA(p_1, q_1) process with GARCH(p_2, q_2) errors if it is covariance stationary

and satisfies difference equations of the form

$$
\begin{aligned}
X_t &= \mu_t + \sigma_t Z_t, \\
\mu_t &= \mu + \sum_{i=1}^{p_1} \phi_i (X_{t-i} - \mu) + \sum_{j=1}^{q_1} \theta_j (X_{t-j} - \mu_{t-j}), \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p_2} \alpha_i (X_{t-i} - \mu_{t-i})^2 + \sum_{j=1}^{q_2} \beta_j \sigma_{t-j}^2,
\end{aligned}
$$

where $\alpha_0 > 0$, $\alpha_i \geqslant 0$, $i = 1, \ldots, p_2$, $\beta_j \geqslant 0$, $j = 1, \ldots, q_2$, and $\sum_{i=1}^{p_2} \alpha_i + \sum_{j=1}^{q_2} \beta_j < 1$.

To be consistent with the previous definition of an ARMA process we build the covariance-stationarity condition for the GARCH errors into the definition. For the ARMA process to be a causal and invertible linear process, as before, the polynomials $\tilde{\phi}(z) = 1 - \phi_1 z - \cdots - \phi_{p_1} z^{p_1}$ and $\tilde{\theta}(z) = 1 + \theta_1 z + \cdots + \theta_{q_1} z^{q_1}$ should have no common roots and no roots inside the unit circle.

Let $(\mathcal{F}_t)_{t \in \mathbb{Z}}$ denote the natural filtration of $(X_t)_{t \in \mathbb{Z}}$, and assume that the ARMA model is invertible. The invertibility of the ARMA process ensures that μ_t is $\mathcal{F}_{t-1}$-measurable as in (4.11). Moreover, since σ_t depends on the infinite history $(X_s - \mu_s)_{s \leqslant t-1}$, the ARMA invertibility also ensures that σ_t is $\mathcal{F}_{t-1}$-measurable. Simple calculations show that $\mu_t = E(X_t \mid \mathcal{F}_{t-1})$ and $\sigma_t^2 = \operatorname{var}(X_t \mid \mathcal{F}_{t-1})$, so that μ_t and σ_t^2 are the conditional mean and variance of the new process.

GARCH with leverage. One of the main criticisms of the standard ARCH and GARCH models is the rigidly symmetric way in which the volatility reacts to recent returns, regardless of their sign. Economic theory suggests that market information should have an asymmetric effect on volatility, whereby bad news leading to a fall in the equity value of a company tends to increase the volatility. This phenomenon has been called a *leverage effect*, because a fall in equity value causes an increase in the debt-to-equity ratio, or so-called leverage, of a company and should consequently make the stock more volatile. At a less theoretical level it seems reasonable that falling stock values might lead to a higher level of investor nervousness than rises in value of the same magnitude.

One method of adding a leverage effect to a GARCH(1, 1) model is by introducing an additional parameter into the volatility equation (4.24) to get

$$
\sigma_t^2 = \alpha_0 + \alpha_1 (X_{t-1} + \delta |X_{t-1}|)^2 + \beta_1 \sigma_{t-1}^2. \tag{4.29}
$$

We assume that $\delta \in [-1, 1]$ and $\alpha_1 \geqslant 0$, as in the GARCH(1, 1) model. Observe that (4.29) may be written as

$$
\sigma_t^2 = \begin{cases} \alpha_0 + \alpha_1 (1 + \delta)^2 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, & X_{t-1} \geqslant 0, \\ \alpha_0 + \alpha_1 (1 - \delta)^2 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, & X_{t-1} < 0, \end{cases}
$$

and hence that

$$
\frac{\partial \sigma_t^2}{\partial X_{t-1}^2} = \begin{cases} \alpha_1 (1 + \delta)^2 \sigma_{t-1}^2, & X_{t-1} \geqslant 0, \\ \alpha_1 (1 - \delta)^2 \sigma_{t-1}^2, & X_{t-1} < 0. \end{cases}
$$

The response of volatility to the magnitude of the most recent return depends on the sign of that return, and we generally expect $\delta < 0$, so bad news has the greater effect.

Threshold GARCH. Observe that (4.29) may easily be rewritten in the form

$$\sigma_t^2 = \alpha_0 + \tilde{\alpha}_1 X_{t-1}^2 + \tilde{\delta}_1 I_{\{X_{t-1}<0\}} X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \tag{4.30}$$

where $\tilde{\alpha}_1 = \alpha_1(1+\delta)^2$ and $\tilde{\delta} = -4\delta\alpha_1$. Equation (4.30) gives the most common version of a threshold GARCH (or TGARCH) model. In effect, a threshold has been set at level zero, and at time t the dynamics depend on whether the previous value of the process X_{t-1} (or innovation Z_{t-1}) was below or above this threshold. However, it is also possible to set non-zero thresholds in TGARCH models, so this represents a more general class of model than GARCH with leverage.

In a less common version of threshold GARCH, the coefficients of the GARCH effects depend on the signs of previous values of the process; this gives a first-order process of the form

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \delta I_{\{X_{t-1}<0\}} \sigma_{t-1}^2. \tag{4.31}$$

Remark 4.23. Note, also, that a further way to introduce asymmetry into a GARCH model is to explicitly use an asymmetric innovation distribution (albeit normalized to have mean 0 and variance 1). Candidate distributions could come from the generalized hyperbolic family of Section 6.2.3.

4.2.4 Fitting GARCH Models to Data

Building the likelihood. In practice, the most widely used approach to fitting GARCH models to data is maximum likelihood. We consider in turn the fitting of the ARCH(1) and GARCH(1, 1) models, from which the fitting of general ARCH(p) and GARCH(p, q) models easily follows.

For the ARCH(1) and GARCH(1, 1) models, suppose we have a total of $n+1$ data values $X_0, X_1, \ldots, X_n$. It is useful to recall that we can write the joint density of the corresponding rvs as

$$f_{X_0,\ldots,X_n}(x_0, \ldots, x_n) = f_{X_0}(x_0) \prod_{t=1}^{n} f_{X_t|X_{t-1},\ldots,X_0}(x_t \mid x_{t-1}, \ldots, x_0). \tag{4.32}$$

For the pure ARCH(1) process, which is first-order Markovian, the conditional densities $f_{X_t|X_{t-1},\ldots,X_0}$ in (4.32) depend on the past only through the value of σ_t or, equivalently, X_{t-1}. The conditional density is easily calculated to be

$$f_{X_t|X_{t-1},\ldots,X_0}(x_t \mid x_{t-1}, \ldots, x_0) = f_{X_t|X_{t-1}}(x_t \mid x_{t-1}) = \frac{1}{\sigma_t} f_Z\left(\frac{x_t}{\sigma_t}\right), \tag{4.33}$$

where $\sigma_t = (\alpha_0 + \alpha_1 x_{t-1}^2)^{1/2}$ and $f_Z(z)$ denotes the density of the innovations $(Z_t)_{t\in\mathbb{Z}}$. We recall that this must have mean 0 and variance 1, and typical choices would be the standard normal density or the density of a t distribution scaled to have unit variance.

However, the marginal density f_{X_0} in (4.32) is not known in a tractable closed form for ARCH and GARCH models, and this poses a problem for basing a likelihood on (4.32). The solution employed in practice is to construct the *conditional likelihood* given X_0, which is calculated from

$$f_{X_1,\ldots,X_n|X_0}(x_1,\ldots,x_n \mid x_0) = \prod_{t=1}^{n} f_{X_t|X_{t-1},\ldots,X_0}(x_t \mid x_{t-1},\ldots,x_0). \tag{4.34}$$

For the ARCH(1) model this follows from (4.33) and is

$$L(\alpha_0,\alpha_1;\boldsymbol{X}) = f_{X_1,\ldots,X_n|X_0}(X_1,\ldots,X_n \mid X_0) = \prod_{t=1}^{n} \frac{1}{\sigma_t} f_Z\left(\frac{X_t}{\sigma_t}\right),$$

with $\sigma_t = (\alpha_0 + \alpha_1 X_{t-1}^2)^{1/2}$. For an ARCH($p$) model we would use analogous arguments to write down a likelihood conditional on the first p values.

In the GARCH(1, 1) model, σ_t is recursively defined in terms of σ_{t-1}, and here, instead of using (4.34), we construct the joint density of $X_1,\ldots,X_n$ conditional on realized values of both X_0 and σ_0, which is

$$f_{X_1,\ldots,X_n|X_0,\sigma_0}(x_1,\ldots,x_n \mid x_0,\sigma_0) = \prod_{t=1}^{n} f_{X_t|X_{t-1},\ldots,X_0,\sigma_0}(x_t \mid x_{t-1},\ldots,x_0,\sigma_0).$$

The conditional densities $f_{X_t|X_{t-1},\ldots,X_0,\sigma_0}$ depend on the past only through the value of σ_t, which is given recursively from $\sigma_0, X_0, \ldots, X_{t-1}$ using $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1\sigma_{t-1}^2$. This gives us the conditional likelihood

$$L(\alpha_0,\alpha_1,\beta_1;\boldsymbol{X}) = \prod_{t=1}^{n} \frac{1}{\sigma_t} f_Z\left(\frac{X_t}{\sigma_t}\right), \quad \sigma_t = \sqrt{\alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1\sigma_{t-1}^2}.$$

The problem remains that the value of σ_0^2 is not actually observed, and this is usually solved by choosing a starting value, such as the sample variance of $X_1,\ldots,X_n$, or even simply zero.

For a GARCH(p, q) model, we would assume that we had $n + p$ data values labelled $X_{-p+1},\ldots,X_0,X_1,\ldots,X_n$. We would evaluate the likelihood conditional on the (observed) values of $X_{-p+1},\ldots,X_0$ as well as the (unobserved) values of $\sigma_{-q+1},\ldots,\sigma_0$, for which starting values would be used as above. For example, if $p = 1$ and $q = 3$, we require starting values for σ_0, σ_{-1} and σ_{-2}.

A similar approach can be used to develop a likelihood for an ARMA model with GARCH errors. In this case we would end up with a conditional likelihood of the form

$$L(\boldsymbol{\theta};\boldsymbol{X}) = \prod_{t=1}^{n} \frac{1}{\sigma_t} f_Z\left(\frac{X_t - \mu_t}{\sigma_t}\right),$$

where σ_t follows a GARCH specification and μ_t follows an ARMA specification as in Definition 4.22, and all unknown parameters (possibly including unknown parameters of the innovation distribution) have been collected in the vector $\boldsymbol{\theta}$. We could of course also consider models with leverage or threshold effects.

Deriving parameter estimates. Consider, then, a log-likelihood of the form

$$\ln L(\boldsymbol{\theta}; \boldsymbol{X}) = \sum_{t=1}^{n} l_t(\boldsymbol{\theta}), \tag{4.35}$$

where l_t denotes the log-likelihood contribution arising from the tth observation. The maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ maximizes the (conditional) log-likelihood in (4.35) and, being in general a local maximum, solves the likelihood equations

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\boldsymbol{\theta}; \boldsymbol{X}) = \sum_{t=1}^{n} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \tag{4.36}$$

where the left-hand side is also known as the *score vector* of the conditional likelihood. The equations (4.36) are usually solved numerically using so-called modified Newton–Raphson procedures. A particular method that is widely used for GARCH models is the BHHH method of Berndt et al. (1974).

In describing the behaviour of parameter estimates in the following paragraphs, we distinguish two situations. In the first situation we assume that the model that has been fitted has been *correctly specified*, so that the data are truly generated by a time-series model with both the assumed dynamic form and innovation distribution. We describe the asymptotic behaviour of the maximum likelihood estimates (MLEs) under this idealization.

In the second situation we assume that the correct dynamic form is fitted but that the innovations are erroneously assumed to be Gaussian. Under this misspecification, the model fitting procedure is known as *quasi-maximum likelihood* (QML) and the estimates obtained are QMLEs. Essentially, the Gaussian likelihood is treated as an objective function to be maximized rather than a proper likelihood; our intuition suggests that this may still give reasonable parameter estimates, and this turns out to be the case under appropriate assumptions about the true innovation distribution.

Properties of MLEs. It helps to recall at this point the asymptotic distribution theory for MLEs in the classical iid case, which is summarized in Section A.3. The asymptotic results we give for GARCH models have a similar form to the results in the iid case, but it is important to realize that this is not simply an application of these results. The asymptotics have been separately and laboriously derived in a series of papers for which starting references are given in Notes and Comments. We will give results for pure GARCH models without ARMA components or additional leverage structure, which have been studied rigorously, but the form of the results will apply more generally.

For a pure GARCH(p, q) model with Gaussian innovations it can be shown that (assuming the model has been correctly specified)

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathrm{d}} N_{p+q+1}(\mathbf{0}, I(\boldsymbol{\theta})^{-1}),$$

where

$$I(\boldsymbol{\theta}) = E\left(\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right) = -E\left(\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right) \tag{4.37}$$

is the Fisher information matrix arising from any single observation. Thus we have consistent and asymptotically normal estimates of the GARCH parameters. In practice, the *expected* information matrix $I(\boldsymbol{\theta})$ is approximated by an *observed* information matrix, and here we could take the observed information matrix coming from either of the equivalent forms for the expected information matrix in (4.37). That is, we could use

$$\bar{I}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right) \quad \text{or} \quad \bar{J}(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}, \tag{4.38}$$

where the first matrix is said to have *outer-product* form and the second is said to have *Hessian* form. These matrices are estimated by evaluating them at the MLEs to get $\bar{I}(\hat{\boldsymbol{\theta}})$ or $\bar{J}(\hat{\boldsymbol{\theta}})$. In practice, the derivatives of the log-likelihood at the MLE are often approximated using first- and second-order differences.

If the model is correctly specified, the estimates $\bar{I}(\hat{\boldsymbol{\theta}})$ and $\bar{J}(\hat{\boldsymbol{\theta}})$ should be broadly similar, being estimators based on two different expressions for the same Fisher information matrix. In practice, we could also estimate $I(\boldsymbol{\theta})$ by $\bar{J}(\hat{\boldsymbol{\theta}})\bar{I}(\hat{\boldsymbol{\theta}})^{-1}\bar{J}(\hat{\boldsymbol{\theta}})$, and this anticipates the so-called *sandwich estimator* that is used in the QML procedure.

Properties of QMLEs. In this approach we assume that the true data-generating mechanism is a GARCH(p, q) model with non-Gaussian innovations, but we attempt to estimate the parameters of the process by maximizing the likelihood for a GARCH(p, q) model with Gaussian innovations. We still obtain consistent estimators of the model parameters and, if the true innovation distribution has a finite fourth moment, we again get asymptotic normality; however, the form of the asymptotic covariance matrix changes.

We now distinguish between matrices $I(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$, given by

$$I(\boldsymbol{\theta}) = E\left(\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right), \qquad J(\boldsymbol{\theta}) = -E\left(\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right),$$

where the expectation is now taken with respect to the true model (not the misspecified Gaussian model). The matrices $I(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$ differ in general (unless the Gaussian model is correct). It may be shown that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathrm{d}} N_{p+q+1}(\mathbf{0}, J(\boldsymbol{\theta})^{-1} I(\boldsymbol{\theta}) J(\boldsymbol{\theta})^{-1}), \tag{4.39}$$

and the asymptotic covariance matrix is said to be of sandwich form; it can be estimated by $\bar{J}(\hat{\boldsymbol{\theta}})^{-1}\bar{I}(\hat{\boldsymbol{\theta}})\bar{J}(\hat{\boldsymbol{\theta}})^{-1}$, where $\bar{I}(\boldsymbol{\theta})$ and $\bar{J}(\boldsymbol{\theta})$ are defined in (4.38). If the model-checking procedures described below suggest that the dynamics have been adequately described by the GARCH model, but the Gaussian assumption seems doubtful, then standard errors for parameter estimates should be based on this covariance matrix estimate.

Model checking. As with ARMA models, it is usual to check fitted GARCH models using residuals. We consider a general ARMA–GARCH model of the form $X_t - \mu_t = \varepsilon_t = \sigma_t Z_t$, with μ_t and σ_t as in Definition 4.22. In this model we distinguish between *unstandardized* and *standardized* residuals. The former are the residuals

$\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$ from the ARMA part of the model; they are calculated using the approach in (4.13), and under the hypothesized model they should behave like a realization of a pure GARCH process. The latter are reconstructed realizations of the SWN that is assumed to drive the GARCH part of the model, and they are calculated from the former by

$$\hat{Z}_t = \hat{\varepsilon}_t/\hat{\sigma}_t, \qquad \hat{\sigma}_t^2 = \hat{\alpha}_0 + \sum_{i=1}^{p_2} \hat{\alpha}_i \hat{\varepsilon}_{t-i}^2 + \sum_{j=1}^{q_2} \hat{\beta}_j \hat{\sigma}_{t-j}^2. \tag{4.40}$$

To use (4.40) we need some initial values, and one solution is to set required starting values of $\hat{\varepsilon}_t$ equal to zero and required starting values of the volatility $\hat{\sigma}_t$ equal to either the sample variance or zero. Because the first few values will be influenced by these starting values, as well as the starting values required to calculate the unstandardized residuals, they may be ignored in later analyses.

The standardized residuals should behave like an SWN and this can be investigated by constructing correlograms of raw and absolute values and applying portmanteau tests of strict white noise, as described in Section 4.1.3.

Assuming that the SWN hypothesis is not rejected, so that the dynamics have been satisfactorily captured, the validity of the distribution used in the ML fitting can also be investigated using Q–Q plots and goodness-of-fit tests for the normal or scaled t distributions. If the Gaussian likelihood does a reasonable job of estimating dynamics, but the residuals do not behave like iid standard normal observations, then the QML fitting philosophy can be adopted and standard errors can be estimated using the sandwich estimator implied by (4.39) above.

This opens up the possibility of *two-stage analyses*, where first the dynamics are estimated by QML methods and then the innovation distribution is modelled using the residuals from the dynamic model as data. The first stage is sometimes called *pre-whitening* of the data. In the second stage we might consider using heavier-tailed models than the Gaussian that also allow some asymmetry in the innovations.

A disadvantage of the two-stage approach is that the error from the time-series modelling propagates through to the distributional fitting in the second stage and the overall error is hard to quantify, but the procedure does lead to more transparency in model building and allows us to separate the tasks of volatility modelling and modelling the shocks that drive the process. In higher-dimensional risk-factor modelling, it may be a useful pragmatic approach.

Example 4.24 (GARCH model for Microsoft log-returns). We consider the Microsoft daily log-returns for the period 1997–2000 (1009 values), as shown in Figure 4.5. Although the raw returns show no evidence of serial correlation (see Figure 4.6), their absolute values do show serial correlation and they fail a Ljung–Box test (based on the first ten estimated correlations) at the 5% level.

For these data, models with Student t innovations are clearly preferred to models with Gaussian innovations, so we adopt an ML approach to fitting models with t innovations. We compare the standard GARCH(1, 1) model (with a constant mean term) with models that incorporate ARMA structure (AR(1), MA(1) and ARMA(1, 1))

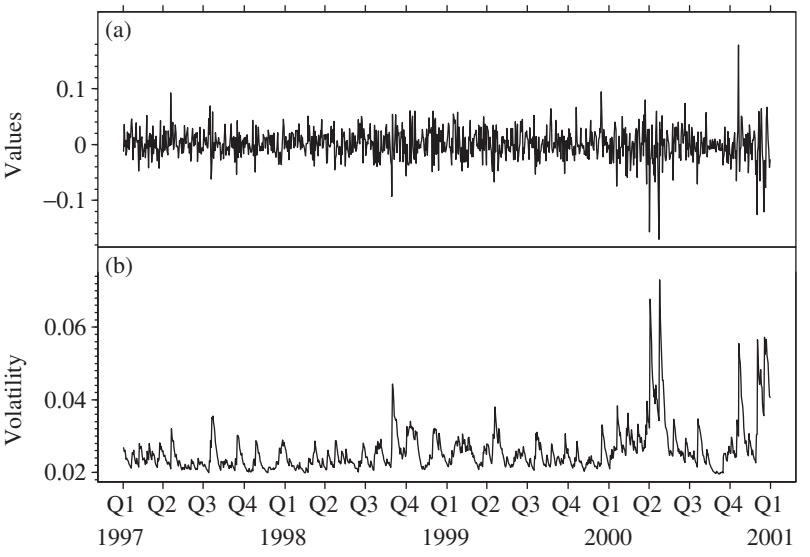

Figure 4.5. Microsoft log-returns 1997–2000; data and estimate of volatility from a GARCH(1, 1) model with a leverage term. (a) Original series. (b) Conditional standard deviation.

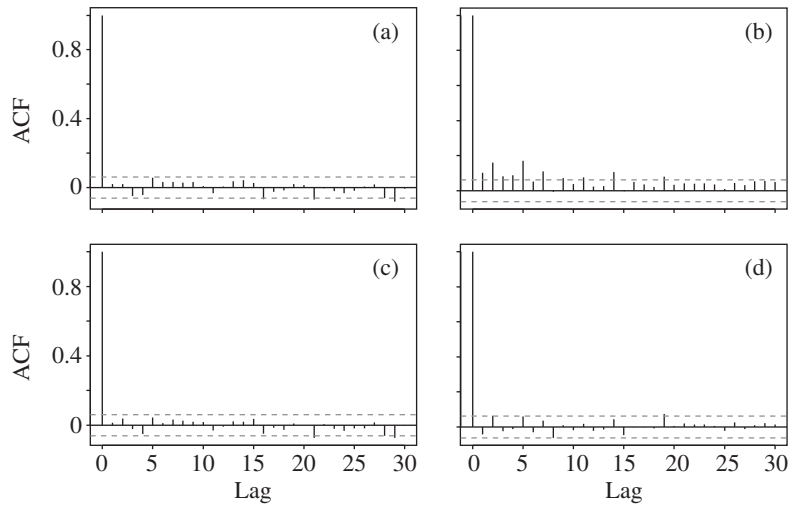

Figure 4.6. Microsoft log-returns 1997–2000; correlograms of data ((a) raw and (b) absolute values) and residuals ((c) raw and (d) absolute values) from a GARCH(1, 1) model.

for the conditional mean; the ARMA structure seems to offer little improvement in the model, and the basic GARCH(1, 1) model is favoured in an Akaike comparison. However, a model with a leverage term as in (4.29) does seem to offer an improvement. Both the raw and absolute standardized residuals obtained from this model show no visual evidence of serial correlation (see again Figure 4.6) and they do not fail Ljung–Box tests. The estimated degrees-of-freedom parameter of the (scaled)

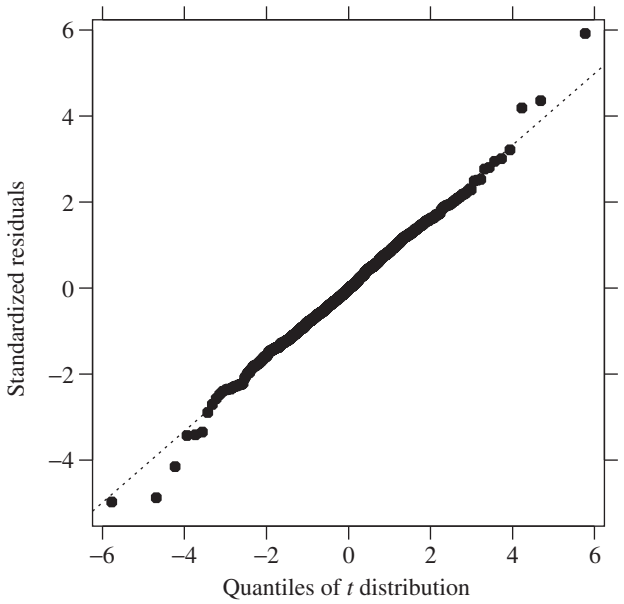

Figure 4.7. Microsoft log-returns 1997–2000; Q–Q plot of residuals from a GARCH(1, 1) model with leverage against a Student t distribution with 6.30 degrees of freedom.

Table 4.1. Analysis of Microsoft log-returns for the period 1997–2000; ML estimates of parameters and standard errors for a GARCH(1, 1) model with a leverage term under the assumption of t innovations.

Parameter	Estimate	Standard error	Ratio
μ	9.35×10^{-4}	7.21×10^{-4}	1.30
α_0	7.79×10^{-5}	3.07×10^{-5}	2.54
α_1	0.108	0.0369	2.91
β_1	0.778	0.0673	11.57
δ	−0.178	0.123	−1.45

t distribution is 6.30 (the standard error is 1.07) and a Q–Q plot of the residuals against this reference distribution reveals a satisfactory correspondence (see Figure 4.7). The estimates of the remaining parameters (with standard errors) in this model are given in Table 4.1.

4.2.5 Volatility Forecasting and Risk Measure Estimation

In this section we assume that our underlying model is a strictly and covariance-stationary time-series process (X_t) adapted to a filtration $(\mathcal{F}_t)$ satisfying equations of the form

$$X_t = \mu_t + \sigma_t Z_t, \tag{4.41}$$

where μ_t and σ_t are $\mathcal{F}_{t-1}$-measurable and Z_t is an innovation variable with mean 0 and variance 1 that is independent of $\mathcal{F}_{t-1}$. Examples fitting into the framework

of (4.41) are any of the ARCH and GARCH models discussed in this chapter as well as causal and invertible ARMA models with GARCH errors.

Our task is to forecast σ_{t+h} for $h \geqslant 1$ based on a sample of n data $X_{t-n+1}, \ldots, X_t$, which are assumed to be generated by the process (4.41). As in Section 4.1.5 we assume that we have observed the infinite history of the process up to time t and derive prediction formulas that we adapt to take account of the finiteness of the sample.

Since

$$E(\sigma_{t+h}^2 \mid \mathcal{F}_t) = E((X_{t+h} - \mu_{t+h})^2 \mid \mathcal{F}_t),$$

our forecasting problem is closely related to the problem of predicting $(X_{t+h} - \mu_{t+h})^2$, and we can use a similar approach to prediction to the one described in Section 4.1.5. We first derive prediction equations under explicit assumptions about the underlying model (i.e. when we specify the structure of σ_t and μ_t in (4.41)) before presenting the more ad hoc technique of exponentially weighted moving-average (EWMA) prediction. Finally, we describe how forecasts of volatility form the basis for estimates of value-at-risk and expected shortfall.

GARCH-based volatility prediction. Assume that a GARCH model has been fitted and its parameters estimated; we will suppress estimator notation for the parameters in the remainder of the section. We make calculations for two simple models, from which the general procedure for more complex models should be clear.

Example 4.25 (prediction in the GARCH(1, 1) model). Suppose that we use a pure GARCH(1, 1) model as in Definition 4.20, which conforms to (4.41) with $\mu_t = 0$. Since $E(X_{t+h} \mid \mathcal{F}_t) = 0$ (the martingale-difference property of the GARCH process), optimal predictions of X_{t+h} are zero. A natural prediction of X_{t+1}^2 based on $\mathcal{F}_t$ is its conditional mean σ_{t+1}^2 given by

$$E(X_{t+1}^2 \mid \mathcal{F}_t) = \sigma_{t+1}^2 = \alpha_0 + \alpha_1 X_t^2 + \beta_1 \sigma_t^2,$$

and, if $E(X_t^4) < \infty$, this is the optimal squared error prediction. Note that the prediction of the random variable X_{t+1}^2 based on the information $\mathcal{F}_t$ is the value of σ_{t+1}^2, which is *known* at time t, being a function of the history of the process.

In practice, we have to make an approximation based on this formula because the infinite series of past values that would allow us to calculate σ_t^2 is not available to us. A natural approach in applications is to approximate σ_t^2 by an estimate of squared volatility $\hat{\sigma}_t^2$ calculated from the residual equations (4.40). Our approximate forecast of X_{t+1}^2 also functions as an estimate of the squared volatility at time $t + 1$ and is given by

$$\hat{\sigma}_{t+1}^2 = \hat{E}(X_{t+1}^2 \mid \mathcal{F}_t) = \alpha_0 + \alpha_1 X_t^2 + \beta_1 \hat{\sigma}_t^2. \tag{4.42}$$

Thus equation (4.42) can be thought of as a recursive scheme for estimating volatility one step ahead.

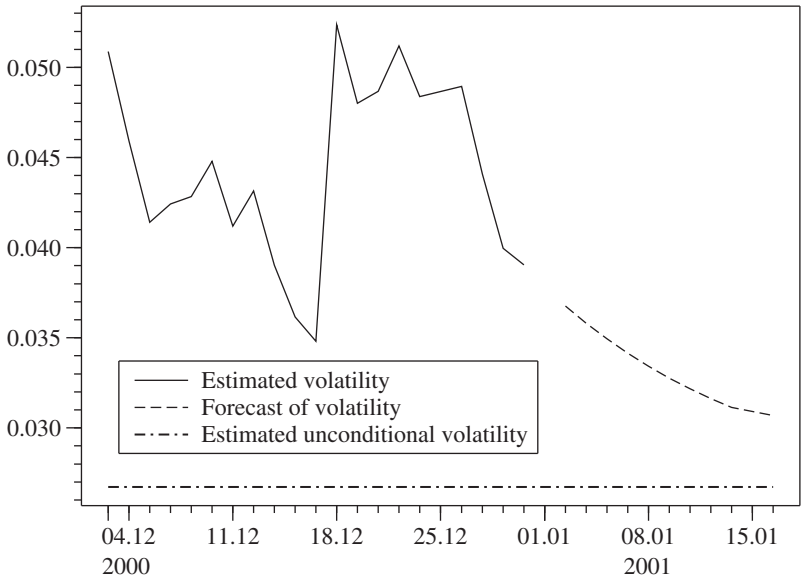

Figure 4.8. Estimate of volatility for the final days of the year 2000 and predictions of volatility for the first ten days of 2001 based on a GARCH(1, 1) model (without leverage) fitted to the Microsoft return data in Example 4.24.

When we look $h > 1$ steps ahead given the information at time t, both X_{t+h}^2 and σ_{t+h}^2 are rvs. Their predictions coincide and are

$$\begin{aligned} E(X_{t+h}^2 \mid \mathcal{F}_t) &= E(\sigma_{t+h}^2 \mid \mathcal{F}_t) \\ &= \alpha_0 + \alpha_1 E(X_{t+h-1}^2 \mid \mathcal{F}_t) + \beta_1 E(\sigma_{t+h-1}^2 \mid \mathcal{F}_t) \\ &= \alpha_0 + (\alpha_1 + \beta_1) E(X_{t+h-1}^2 \mid \mathcal{F}_t), \end{aligned}$$

so that a general formula is

$$E(X_{t+h}^2 \mid \mathcal{F}_t) = \alpha_0 \sum_{i=0}^{h-1} (\alpha_1 + \beta_1)^i + (\alpha_1 + \beta_1)^{h-1} (\alpha_1 X_t^2 + \beta_1 \sigma_t^2),$$

and we obtain a practical formula by substituting an estimate of squared volatility $\hat{\sigma}_t^2$ as before. As $h \to \infty$ we observe that $E(\sigma_{t+h}^2 \mid \mathcal{F}_t) \to \alpha_0/(1 - \alpha_1 - \beta_1)$, almost surely, so that the prediction of squared volatility converges to the unconditional variance of the process. A concrete example of volatility prediction in a GARCH(1, 1) model is given in Figure 4.8 for the Microsoft data analysed in Example 4.24.

We now consider a second example, which combines what we know about prediction in ARMA and GARCH models.

Example 4.26 (prediction in an ARMA(1, 1)–GARCH(1, 1) model). Suppose that we use an ARMA(1, 1) model with GARCH(1, 1) errors as in Definition 4.22. This also conforms to (4.41), and prediction formulas for this model follow easily

from Examples 4.15 and 4.25. We calculate that

$$E(X_{t+h} \mid \mathcal{F}_t) = \mu + \phi_1^h(X_t - \mu) + \phi_1^{h-1}\theta_1\varepsilon_t, \tag{4.43}$$

$$\operatorname{var}(X_{t+h} \mid \mathcal{F}_t) = \alpha_0 \sum_{i=0}^{h-1}(\alpha_1 + \beta_1)^i + (\alpha_1 + \beta_1)^{h-1}(\alpha_1\varepsilon_t^2 + \beta_1\sigma_t^2), \tag{4.44}$$

and these are approximated by substituting inferred values for ε_t and σ_t obtained from the residual equations (4.40). Equation (4.43) yields predictions of μ_{t+h} or X_{t+h}, and equation (4.44) yields predictions of $(X_{t+h} - \mu_{t+h})^2$ or σ_{t+h}^2.

Exponential smoothing for volatility. Now suppose that we do not want to make detailed assumptions about the structure of σ_t and μ_t in (4.42). We consider a simpler scheme for forecasting volatility that builds on the exponential smoothing idea of Section 4.1.5. We recall from (4.14) that a forecast $P_t(X_{t+1})$ of X_{t+1} based on time-t information can be constructed using an updating scheme of the form

$$P_t X_{t+1} = \lambda X_t + (1 - \lambda) P_{t-1} X_t \tag{4.45}$$

for an appropriately chosen value of the parameter λ. If we apply this scheme to the prediction of $(X_{t+1} - \mu_{t+1})^2$, we obtain

$$P_t(X_{t+1} - \mu_{t+1})^2 = \alpha(X_t - \mu_t)^2 + (1 - \alpha) P_{t-1}(X_t - \mu_t)^2 \tag{4.46}$$

for an appropriately chosen value of the parameter α. Of course, in addition to choosing α, we also need to insert an estimate of the unobserved conditional mean μ_t to use (4.46).

Since $\sigma_{t+1}^2 = E((X_{t+1} - \mu_{t+1})^2 \mid \mathcal{F}_t)$, we can also use (4.46) as an exponential smoothing scheme for the unobserved squared volatility. This yields a recursive scheme for the one-step-ahead volatility forecast given by

$$\hat{\sigma}_{t+1}^2 = \alpha(X_t - \hat{\mu}_t)^2 + (1 - \alpha)\hat{\sigma}_t^2, \tag{4.47}$$

which defines the EWMA procedure. For many risk-factor return series, the conditional mean appears to be close to zero (recall the stylized facts of return series in Section 3.1) and we often set $\hat{\mu}_t = 0$. Alternatively, we can apply the exponential smoothing idea to the conditional mean and replace $\hat{\mu}_t$ by an estimate $P_{t-1}X_t$ derived using the recursive scheme (4.45). Typical values for α are generally small; for example, in the RiskMetrics methodology widely used by banks, a value of $\alpha = 0.06$ has been recommended (Mina and Xiao 2001).

If we compare (4.47) with the one-step-ahead volatility estimation scheme defined by a GARCH(1, 1) model in (4.42), it is tempting to say that EWMA corresponds to estimating volatility using a conditional-expectation-based technique in an IGARCH model, where the parameter α_0 equals zero. This analogy should be used with care; GARCH and IGARCH models with $\alpha_0 = 0$ are not well defined, and the solution of the stochastic recurrence relation in (4.27) vanishes. Moreover, IGARCH is not covariance stationary. It is better to regard EWMA as a sensible model-free approach to volatility forecasting based on the classical technique of exponential smoothing.

Estimates of VaR and expected shortfall. Finally, we suppose that the data $X_{t-n+1}, \ldots, X_t$ can be interpreted as financial losses and we consider the application of risk measures based on loss distributions (see Section 2.3.1) to the conditional distribution $F_{X_{t+1}|\mathcal{F}_t}$. For example, the data may represent *negative* log-returns on an asset price rather than returns. In particular, we look at the estimation of value-at-risk and expected shortfall for the distribution $F_{X_{t+1}|\mathcal{F}_t}$.

Writing F_Z for the df of the innovations (Z_t), the $\mathcal{F}_t$-measurability of μ_{t+1} and σ_{t+1} implies that

$$F_{X_{t+1}|\mathcal{F}_t}(x) = P(\mu_{t+1} + \sigma_{t+1}Z_{t+1} \leqslant x \mid \mathcal{F}_t) = F_Z((x - \mu_{t+1})/\sigma_{t+1}).$$

Let VaR_α^t denote the α-quantile of $F_{X_{t+1}|\mathcal{F}_t}$ and let ES_α^t denote the corresponding expected shortfall. Using the approach of Examples 2.11 and 2.14 we obtain

$$\mathrm{VaR}_\alpha^t = \mu_{t+1} + \sigma_{t+1}q_\alpha(Z), \qquad \mathrm{ES}_\alpha^t = \mu_{t+1} + \sigma_{t+1}\,\mathrm{ES}_\alpha(Z), \tag{4.48}$$

where we write Z for a generic rv with df F_Z.

It is clear that if we can estimate μ_{t+1} and σ_{t+1}, then we only need to be able to estimate $q_\alpha(Z)$ and $\mathrm{ES}_\alpha(Z)$ for the innovation distribution to obtain estimates of the risk measures in (4.48). This task can be accomplished in both a parametric and a non-parametric (or semi-parametric) setting. If we estimate a fully specified GARCH-type model using the ML approach of Section 4.2.4, then it is mostly straightforward to calculate $q_\alpha(Z)$ and $\mathrm{ES}_\alpha(Z)$ for the estimated innovation distribution. If, on the other hand, we use a QML method to fit a GARCH-type model or, even more simply, we use exponential smoothing techniques to estimate the volatility and conditional mean, then we can form residuals $\hat{Z}_s = (X_s - \hat{\mu}_s)/\hat{\sigma}_s$ for $s = t - n + 1, \ldots, n$ and apply quantile and expected shortfall estimation techniques to these residuals; statistical methods for estimating risk measures from data are discussed in Section 9.2.6.

Notes and Comments

The ARCH process was originally proposed by Engle (1982), and the GARCH process by Bollerslev (1986), who gave the condition for covariance stationarity. Overview texts on GARCH models include the books by Gouriéroux (1997) and Francq and Zakoïan (2010) and a number of useful review articles including Bollerslev, Chou and Kroner (1992), Bollerslev, Engle and Nelson (1994) and Shephard (1996). There are also substantial sections on GARCH models in the books by Alexander (2001), Tsay (2002) and Zivot and Wang (2003). The IGARCH model was first discussed by Engle and Bollerslev (1986).

The condition for strict stationarity of GARCH models was derived by Nelson (1990) in the case of the GARCH(1, 1) model and by Bougerol and Picard (1992) for GARCH(p, q). The necessary theory involves the study of stochastic recurrence relations and goes back to Kesten (1973); Brandt (1986) is also a useful reference. Readable accounts of this theory may be found in Embrechts, Klüppelberg and Mikosch (1997), Mikosch and Stărică (2000) and Mikosch (2003, 2013).

For more on the derivation of conditional likelihood functions for ARCH and GARCH models, see Hamilton (1994) and Tsay (2002). The BHHH algorithm (Berndt et al. 1974) is the most commonly used approach to numerically maximizing the likelihood. For an informative general discussion of numerical optimization procedures in the context of maximum likelihood, see Hamilton (1994, pp. 133–142). Standard general references on the QML approach are White (1981) and Gouriéroux, Montfort and Trognon (1984).

The essential asymptotic properties of MLEs and QMLEs in GARCH models are described in many publications, but a detailed mathematical proof has often lagged behind the assertions. Early papers appealed to regularity conditions for conditionally specified models such as those of Crowder (1976), which are essentially unverifiable. Lee and Hansen (1994) and Lumsdaine (1996) proved consistency and asymptotic normality of QMLEs in the GARCH(1, 1) model. More recently, Berkes, Horváth and Kokoszka (2003) have extended this to the GARCH(p, q) model under minimal assumptions, and Straumann (2005) and Straumann and Mikosch (2006) have given similar results for a wide variety of first-order models.

From a more practical point-of-view, it is not easy to estimate GARCH model parameters to a high degree of accuracy because of the flatness of the typical likelihoods and the non-negligible influence of starting values in finite samples. Readers who write their own code may wish to compare their estimates with benchmark studies by McCullough and Renfro (1999) and Brooks, Burke and Persand (2001).

Alternative innovation distributions to the Gaussian and scaled t distributions that have been considered include the generalized error distribution (GED) in Nelson (1991) and the normal inverse Gaussian (NIG) in Venter and de Jongh (2002); the latter authors present extensive evidence that the NIG is a good choice of innovation distribution for practical work and that GARCH inference based on the NIG is relatively robust to misspecification of the distribution.

A great many extensions to the GARCH class have been proposed and thorough surveys may be found in Bollerslev, Engle and Nelson (1994) and Shephard (1996). Leverage effects in the GARCH model and the more general PGARCH (power GARCH) model are examined in Ding, Granger and Engle (1993). Various threshold GARCH models have been suggested; the model (4.30) is of the type suggested by Glosten, Jagannathan and Runkle (1993), while (4.31) is the switching-volatility GARCH (SV-GARCH) model of Fornari and Mele (1997). There have been proposals for non-parametric ARCH and GARCH modelling, including the multiplicative ARCH(p)-model of Yang, Härdle and Nielsen (1999) and the non-parametric GARCH procedure of Bühlmann and McNeil (2002). For long-memory processes modelling volatility, see the book by Beran et al. (2013).

The use of the EWMA (exponentially weighted moving-average) volatility estimation method based on exponential smoothing was popularized by the RiskMetrics Group at JPMorgan (JPMorgan 1996; Mina and Xiao 2001). See also Zivot and Wang (2003) for examples of the use of this method.

5
Extreme Value Theory

Extreme value theory (EVT) is a branch of probability concerned with limiting laws for extreme values in large samples. The theory contains many important results describing the behaviour of sample maxima and minima, upper-order statistics (such as the kth largest value in a sample) and sample values exceeding high thresholds. Our interest in this theory centres on the application of the results to developing models for the extremal behaviour of financial risk factors. In particular, we are interested in models for the tail of the distribution of financial risk-factor changes. We have observed at various points in Chapters 3 and 4 that risk-factor changes are frequently heavy tailed when compared with a normal distribution.

Much of this chapter is based on the presentation of EVT in Embrechts, Klüppelberg and Mikosch (1997) (henceforth, EKM), and whenever theoretical detail is missing the reader should consult that text. We concentrate on describing the statistical models suggested by EVT, while briefly summarizing the theoretical ideas on which the statistical methods are based.

We focus on two main kinds of model for extreme values. The most traditional models are the block maxima models described in Section 5.1: these are models for the largest observations collected from large samples of identically distributed observations. A more modern and powerful group of models are those for threshold exceedances, described in Section 5.2. These are models for all large observations that exceed some high level, and they are generally considered to be the most useful for practical applications, due to their more efficient use of the (often limited) data on extreme outcomes.

The models for threshold exceedances can be embedded in an elegant point process framework that simultaneously addresses their occurrence in time as well as the magnitude of excess losses over the threshold. This is the so-called peaks-over-threshold (POT) model, which is presented in Section 5.3. The POT model serves as a starting point for developing more dynamic descriptions of the occurrence and magnitude of extremes using self-exciting (Hawkes) processes. These advanced dynamic models are treated in Chapter 16 along with multivariate EVT.

5.1 Maxima

To begin with we consider a sequence of iid rvs $(X_i)_{i\in\mathbb{N}}$ representing financial losses. These may have a variety of interpretations, such as operational losses, insurance losses and losses on a credit portfolio over fixed time intervals. Later we relax the

assumption of independence and consider that the rvs form a strictly stationary time series of dependent losses; they might be (negative) returns on an investment in a single stock, an index, or a portfolio of investments.

5.1.1 Generalized Extreme Value Distribution

Convergence of sums. The role of the generalized extreme value (GEV) distribution in the theory of extremes is analogous to that of the normal distribution (and, more generally, the stable laws) in the central limit theory for sums of rvs. Assuming that the underlying rvs $X_1, X_2, \ldots$ are iid with a finite variance, and writing $S_n = X_1 + \cdots + X_n$ for the sum of the first n rvs, the standard version of the central limit theorem (CLT) says that appropriately normalized sums $(S_n - a_n)/b_n$ converge in distribution to the standard normal distribution as n goes to infinity. The appropriate normalization uses sequences of normalizing constants (a_n) and (b_n) defined by $a_n = nE(X_1)$ and $b_n = \sqrt{n \operatorname{var}(X_1)}$. In mathematical notation we have

$$\lim_{n\to\infty} P\left(\frac{S_n - a_n}{b_n} \leqslant x\right) = \Phi(x), \quad x \in \mathbb{R}.$$

Convergence of maxima. Classical EVT is concerned with limiting distributions for normalized maxima. We denote the maximum of n iid rvs $X_1, \ldots, X_n$ by $M_n = \max(X_1, \ldots, X_n)$ and refer to this also as an n-block maximum. The only possible non-degenerate limiting distributions for normalized maxima as n goes to infinity are in the GEV family.

Definition 5.1 (generalized extreme value distribution). The df of the (standard) GEV distribution is given by

$$H_\xi(x) = \begin{cases} \exp(-(1+\xi x)^{-1/\xi}), & \xi \neq 0, \\ \exp(-\mathrm{e}^{-x}), & \xi = 0, \end{cases}$$

where $1 + \xi x > 0$. A three-parameter family is obtained by defining $H_{\xi,\mu,\sigma}(x) := H_\xi((x-\mu)/\sigma)$ for a location parameter $\mu \in \mathbb{R}$ and a scale parameter $\sigma > 0$.

The parameter ξ is known as the *shape* parameter of the GEV distribution, and H_ξ defines a *type* of distribution, meaning a family of distributions specified up to location and scaling (see Section A.1.1 for a formal definition). The extreme value distribution in Definition 5.1 is generalized in the sense that the parametric form subsumes three types of distribution that are known by other names according to the value of ξ: when $\xi > 0$ the distribution is a Fréchet distribution; when $\xi = 0$ it is a Gumbel distribution; and when $\xi < 0$ it is a Weibull distribution. We also note that for fixed x we have $\lim_{\xi\to 0} H_\xi(x) = H_0(x)$ (from either side), so that the parametrization in Definition 5.1 is *continuous* in ξ, which facilitates the use of this distribution in statistical modelling.

The df and density of the GEV distribution are shown in Figure 5.1 for the three cases $\xi = 0.5$, $\xi = 0$ and $\xi = -0.5$, corresponding to Fréchet, Gumbel and Weibull types, respectively. Observe that the Weibull distribution is a short-tailed distribution with a so-called finite *right endpoint*. The right endpoint of a distribution will be

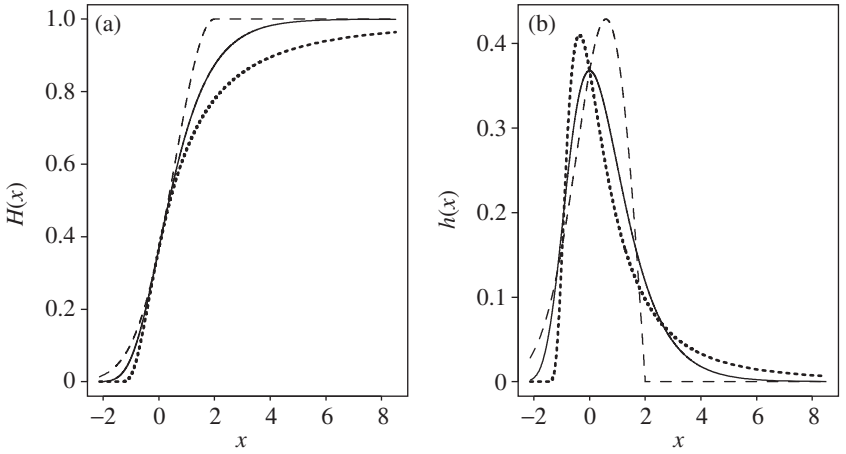

Figure 5.1. (a) The df of a standard GEV distribution in three cases: the solid line corresponds to $\xi = 0$ (Gumbel); the dotted line is $\xi = 0.5$ (Fréchet); and the dashed line is $\xi = -0.5$ (Weibull). (b) Corresponding densities. In all cases, $\mu = 0$ and $\sigma = 1$.

denoted by $x_F = \sup\{x \in \mathbb{R}\colon F(x) < 1\}$. The Gumbel and Fréchet distributions have infinite right endpoints, but the decay of the tail of the Fréchet distribution is much slower than that of the Gumbel distribution.

Suppose that maxima M_n of iid rvs converge in distribution as $n \to \infty$ under an appropriate normalization. Recalling that $P(M_n \leqslant x) = F^n(x)$, we observe that this convergence means that there exist sequences of real constants (d_n) and (c_n), where $c_n > 0$ for all n, such that

$$\lim_{n\to\infty} P((M_n - d_n)/c_n \leqslant x) = \lim_{n\to\infty} F^n(c_n x + d_n) = H(x) \tag{5.1}$$

for some non-degenerate df $H(x)$. The role of the GEV distribution in the study of maxima is formalized by the following definition and theorem.

Definition 5.2 (maximum domain of attraction). If (5.1) holds for some non-degenerate df H, then F is said to be in the maximum domain of attraction of H, written $F \in \mathrm{MDA}(H)$.

Theorem 5.3 (Fisher–Tippett, Gnedenko). *If $F \in \mathrm{MDA}(H)$ for some non-degenerate df H, then H must be a distribution of type H_ξ, i.e. a GEV distribution.*

Remarks 5.4.

(1) If convergence of normalized maxima takes place, the type of the limiting distribution (as specified by ξ) is uniquely determined, although the location and scaling of the limit law (μ and σ) depend on the exact normalizing sequences chosen; this is guaranteed by the so-called convergence to types theorem (EKM, p. 554). It is always possible to choose these sequences such that the limit appears in the standard form H_ξ.

(2) By non-degenerate df we mean a distribution that is not concentrated on a single point.

Examples. We calculate two examples to show how the GEV limit emerges for two well-known underlying distributions and appropriately chosen normalizing sequences. To discover how normalizing sequences may be constructed in general, we refer to Section 3.3 of EKM.

Example 5.5 (exponential distribution). If the underlying distribution is an exponential distribution with df $F(x) = 1 - \mathrm{e}^{-\beta x}$ for $\beta > 0$ and $x \geqslant 0$, then by choosing normalizing sequences $c_n = 1/\beta$ and $d_n = (\ln n)/\beta$ we can directly calculate the limiting distribution of maxima using (5.1). We get

$$F^n(c_n x + d_n) = \left(1 - \frac{1}{n}\mathrm{e}^{-x}\right)^n, \quad x \geqslant -\ln n,$$
$$\lim_{n\to\infty} F^n(c_n x + d_n) = \exp(-\mathrm{e}^{-x}), \qquad x \in \mathbb{R},$$

from which we conclude that $F \in \mathrm{MDA}(H_0)$.

Example 5.6 (Pareto distribution). If the underlying distribution is a Pareto distribution (Pa(α, κ)) with df $F(x) = 1 - (\kappa/(\kappa + x))^\alpha$ for $\alpha > 0, \kappa > 0$ and $x \geqslant 0$, we can take normalizing sequences $c_n = \kappa n^{1/\alpha}/\alpha$ and $d_n = \kappa n^{1/\alpha} - \kappa$. Using (5.1) we get

$$F^n(c_n x + d_n) = \left(1 - \frac{1}{n}\left(1 + \frac{x}{\alpha}\right)^{-\alpha}\right)^n, \quad 1 + \frac{x}{\alpha} \geqslant n^{-1/\alpha},$$
$$\lim_{n\to\infty} F^n(c_n x + d_n) = \exp\left(-\left(1 + \frac{x}{\alpha}\right)^{-\alpha}\right), \qquad 1 + \frac{x}{\alpha} > 0,$$

from which we conclude that $F \in \mathrm{MDA}(H_{1/\alpha})$.

Convergence of minima. The limiting theory for convergence of maxima encompasses the limiting behaviour of minima using the identity

$$\min(X_1, \ldots, X_n) = -\max(-X_1, \ldots, -X_n). \tag{5.2}$$

It is not difficult to see that normalized minima of iid samples with df F will converge in distribution if the df $\tilde{F}(x) = 1 - F(-x)$, which is the df of the rvs $-X_1, \ldots, -X_n$, is in the maximum domain of attraction of an extreme value distribution. Writing $M_n^* = \max(-X_1, \ldots, -X_n)$ and assuming that $\tilde{F} \in \mathrm{MDA}(H_\xi)$, we have

$$\lim_{n\to\infty} P\left(\frac{M_n^* - d_n}{c_n} \leqslant x\right) = H_\xi(x),$$

from which it follows easily, using (5.2), that

$$\lim_{n\to\infty} P\left(\frac{\min(X_1, \ldots, X_n) + d_n}{c_n} \leqslant x\right) = 1 - H_\xi(-x).$$

Thus appropriate limits for minima are distributions of type $1 - H_\xi(-x)$. For a symmetric distribution F we have $\tilde{F}(x) = F(x)$, so that if H_ξ is the limiting type of distribution for maxima for a particular value of ξ, then $1 - H_\xi(-x)$ is the limiting type of distribution for minima.

5.1.2 Maximum Domains of Attraction

For most applications it is sufficient to note that essentially all the common continuous distributions of statistics or actuarial science are in MDA(H_ξ) for some value of ξ. In this section we consider the issue of which underlying distributions lead to which limits for maxima.

The Fréchet case. The distributions that lead to the Fréchet limit $H_\xi(x)$ for $\xi > 0$ have a particularly elegant characterization involving *slowly varying* or *regularly varying* functions.

Definition 5.7 (slowly varying and regularly varying functions).

(i) A positive, Lebesgue-measurable function L on $(0, \infty)$ is slowly varying at ∞ (written $L \in \mathcal{R}_0$) if

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1, \quad t > 0.$$

(ii) A positive, Lebesgue-measurable function h on $(0, \infty)$ is regularly varying at ∞ with index $\rho \in \mathbb{R}$ if

$$\lim_{x\to\infty} \frac{h(tx)}{h(x)} = t^\rho, \quad t > 0.$$

Slowly varying functions are functions that, in comparison with power functions, change relatively slowly for large x, an example being the logarithm $L(x) = \ln(x)$. Regularly varying functions are functions that can be represented by power functions multiplied by slowly varying functions, i.e. $h(x) = x^\rho L(x)$ for some $L \in \mathcal{R}_0$.

Theorem 5.8 (Fréchet MDA, Gnedenko). *For $\xi > 0$,*

$$F \in \text{MDA}(H_\xi) \iff \bar{F}(x) = x^{-1/\xi} L(x) \text{ for some function } L \in \mathcal{R}_0. \tag{5.3}$$

This means that distributions giving rise to the Fréchet case are distributions with tails that are regularly varying functions with a negative index of variation. Their tails decay essentially like a power function, and the rate of decay $\alpha = 1/\xi$ is often referred to as the *tail index* of the distribution. A consequence of Theorem 5.8 is that the right endpoint of any distribution in the Fréchet MDA satisfies $x_F = \infty$.

These distributions are the most studied distributions in EVT and they are of particular interest in financial applications because they are heavy-tailed distributions with infinite higher moments. If X is a non-negative rv whose df F is an element of MDA(H_ξ) for $\xi > 0$, then it may be shown that $E(X^k) = \infty$ for $k > 1/\xi$ (EKM, p. 568). If, for some small $\varepsilon > 0$, the distribution is in MDA($H_{(1/2)+\varepsilon}$), it is an infinite-variance distribution, and if the distribution is in MDA($H_{(1/4)+\varepsilon}$), it is a distribution with infinite fourth moment.

Example 5.9 (Pareto distribution). In Example 5.6 we verified by direct calculation that normalized maxima of iid Pareto variates converge to a Fréchet distribution. Observe that the tail of the Pareto df in (A.19) may be written $\bar{F}(x) = x^{-\alpha} L(x)$, where it may be easily checked that $L(x) = (\kappa^{-1} + x^{-1})^{-\alpha}$ is a slowly varying

function; indeed, as $x \to \infty$, $L(x)$ converges to the constant κ^α. Thus we verify that the Pareto df has the form (5.3).

Further examples of distributions giving rise to the Fréchet limit for maxima include the Fréchet distribution itself and the inverse gamma, Student t, loggamma, F and Burr distributions. We will provide further demonstrations for some of these distributions in Section 16.1.1.

The Gumbel case. The characterization of distributions in this class is more complicated than in the Fréchet class. We have seen in Example 5.5 that the exponential distribution is in the Gumbel class and, more generally, it could be said that the distributions in this class have tails that have an essentially exponential decay. A positive-valued rv with a df in MDA(H_0) has finite moments of any positive order, i.e. $E(X^k) < \infty$ for every $k > 0$ (EKM, p. 148).

However, there is a great deal of variety in the tails of distributions in this class, so, for example, both the normal and lognormal distributions belong to the Gumbel class (EKM, pp. 145–147). The normal distribution, as discussed in Section 6.1.4, is thin tailed, but the lognormal distribution has much heavier tails, and we would need to collect a lot of data from the lognormal distribution before we could distinguish its tail behaviour from that of a distribution in the Fréchet class. Moreover, it should be noted that the right endpoints of distributions in this class satisfy $x_F \leqslant \infty$, so the case $x_F < \infty$ is possible.

In financial modelling it is often erroneously assumed that the only interesting models for financial returns are the power-tailed distributions of the Fréchet class. The Gumbel class is also interesting because it contains many distributions with much heavier tails than the normal, even if these are not regularly varying power tails. Examples are hyperbolic and generalized hyperbolic distributions (with the exception of the special boundary case that is Student t).

Other distributions in MDA(H_0) include the gamma, chi-squared, standard Weibull (to be distinguished from the Weibull special case of the GEV distribution) and Benktander type I and II distributions (which are popular actuarial loss distributions), and the Gumbel itself. We provide demonstrations for some of these examples in Section 16.1.2.

The Weibull case. This is perhaps the least important case for financial modelling, at least in the area of market risk, since the distributions in this class all have finite *right endpoints*. Although all potential financial and insurance losses are, in practice, bounded, we will still tend to favour models that have infinite support for loss modelling. An exception may be in the area of credit risk modelling, where we will see in Chapter 10 that probability distributions on the unit interval [0, 1] are very useful. A characterization of the Weibull class is as follows.

Theorem 5.10 (Weibull MDA, Gnedenko). *For $\xi < 0$,*

$$F \in \mathrm{MDA}(H_\xi) \iff x_F < \infty \text{ and } \bar{F}(x_F - x^{-1}) = x^{1/\xi} L(x)$$

for some function $L \in \mathcal{R}_0$.

It can be shown (EKM, p. 137) that a beta distribution with density $f_{\alpha,\beta}$ as given in (A.10) is in MDA($H_{-1/\beta}$). This includes the special case of the uniform distribution for $\beta = \alpha = 1$.

5.1.3 Maxima of Strictly Stationary Time Series

The standard theory of the previous sections concerns maxima of iid sequences. With financial time series in mind, we now look briefly at the theory for maxima of strictly stationary time series and find that the same types of limiting distribution apply.

In this section let $(X_i)_{i\in\mathbb{Z}}$ denote a strictly stationary time series with stationary distribution F, and let $(\tilde{X}_i)_{i\in\mathbb{Z}}$ denote the *associated iid process*, i.e. a strict white noise process with the same df F. Let $M_n = \max(X_1, \ldots, X_n)$ and $\tilde{M}_n = \max(\tilde{X}_1, \ldots, \tilde{X}_n)$ denote maxima of the original series and the iid series, respectively.

For many processes $(X_i)_{i\in\mathbb{N}}$, it may be shown that there exists a real number θ in (0, 1] such that

$$\lim_{n\to\infty} P\left(\frac{\tilde{M}_n - d_n}{c_n} \leqslant x\right) = H(x) \tag{5.4}$$

for a non-degenerate limit $H(x)$ if and only if

$$\lim_{n\to\infty} P\left(\frac{M_n - d_n}{c_n} \leqslant x\right) = H^\theta(x). \tag{5.5}$$

For such processes this value θ is known as the *extremal index* of the process (not to be confused with the tail index of distributions in the Fréchet class). A formal definition is more technical (see Notes and Comments) but the basic ideas behind (5.4) and (5.5) are easily explained.

For processes with an extremal index, normalized maxima converge in distribution provided that maxima of the associated iid process converge in distribution: that is, provided the underlying distribution F is in MDA(H_ξ) for some ξ. Moreover, since $H_\xi^\theta(x)$ can be easily verified to be a distribution of the same type as $H_\xi(x)$, the limiting distribution of the normalized maxima of the dependent series is a GEV distribution with exactly the same ξ parameter as the limit for the associated iid data; only the location and scaling of the distribution may change.

Writing $u = c_n x + d_n$, we observe that, for large enough n, (5.4) and (5.5) imply that

$$P(M_n \leqslant u) \approx P^\theta(\tilde{M}_n \leqslant u) = F^{n\theta}(u), \tag{5.6}$$

so that for u large, the probability distribution of the maximum of n observations from the time series with extremal index θ can be approximated by the distribution of the maximum of $n\theta < n$ observations from the associated iid series. In a sense, $n\theta$ can be thought of as counting the number of roughly independent *clusters* of observations in n observations, and θ is often interpreted as the reciprocal of the mean cluster size.

Table 5.1. Approximate values of the extremal index as a function of the parameter α_1 for the ARCH(1) process in (4.22).

α_1	0.1	0.3	0.5	0.7	0.9
θ	0.999	0.939	0.835	0.721	0.612

Not every strictly stationary process has an extremal index (see p. 418 of EKM for a counterexample) but, for the kinds of time-series processes that interest us in financial modelling, an extremal index generally exists. Essentially, we only have to distinguish between the cases when $\theta = 1$ and the cases when $\theta < 1$: for the former, there is no tendency to cluster at high levels, and large sample maxima from the time series behave exactly like maxima from similarly sized iid samples; for the latter, we must be aware of a tendency for extreme values to cluster.

- Strict white noise processes (iid rvs) have extremal index $\theta = 1$.
- ARMA processes with Gaussian strict white noise innovations have $\theta = 1$ (EKM, pp. 216–218). However, if the innovation distribution is in MDA(H_ξ) for $\xi > 0$, then $\theta < 1$ (EKM, pp. 415, 416).
- ARCH and GARCH processes have $\theta < 1$ (EKM, pp. 476–480).

The final fact is particularly relevant to our financial applications, since we saw in Chapter 4 that ARCH and GARCH processes provide good models for many financial return series.

Example 5.11 (the extremal index of the ARCH(1) process). In Table 5.1 we reproduce some results from de Haan et al. (1989), who calculate approximate values for the extremal index of the ARCH(1) process (see Definition 4.16) using a Monte Carlo simulation approach. Clearly, the stronger the ARCH effect (that is, the larger the magnitude of the parameter α_1), the greater the tendency of the process to cluster. For a process with parameter 0.9, the extremal index value $\theta = 0.612$ is interpreted as suggesting that the average cluster size is $1/\theta = 1.64$.

5.1.4 The Block Maxima Method

Fitting the GEV distribution. Suppose we have data from an unknown underlying distribution F, which we suppose lies in the domain of attraction of an extreme value distribution H_ξ for some ξ. If the data are realizations of iid variables, or variables from a process with an extremal index such as GARCH, the implication of the theory is that the true distribution of the n-block maximum M_n can be approximated for large enough n by a three-parameter GEV distribution $H_{\xi,\mu,\sigma}$.

We make use of this idea by fitting the GEV distribution $H_{\xi,\mu,\sigma}$ to data on the n-block maximum. Obviously we need repeated observations of an n-block maximum, and we assume that the data can be divided into m blocks of size n. This makes most sense when there are natural ways of blocking the data. The method has its origins in hydrology, where, for example, daily measurements of water levels might be divided into yearly blocks and the yearly maxima collected. Analogously, we will consider

financial applications where daily return data (recorded on trading days) are divided into yearly (or semesterly or quarterly) blocks and the maximum daily falls within these blocks are analysed.

We denote the block maximum of the jth block by M_{nj}, so our data are $M_{n1}, \ldots, M_{nm}$. The GEV distribution can be fitted using various methods, including maximum likelihood. An alternative is the method of probability-weighted moments (see Notes and Comments). In implementing maximum likelihood it will be assumed that the block size n is quite large so that, regardless of whether the underlying data are dependent or not, the block maxima observations can be taken to be independent. In this case, writing $h_{\xi,\mu,\sigma}$ for the density of the GEV distribution, the log-likelihood is easily calculated to be

$$\begin{aligned} &l(\xi, \mu, \sigma; M_{n1}, \ldots, M_{nm}) \\ &\quad = \sum_{i=1}^{m} \ln h_{\xi,\mu,\sigma}(M_{ni}) \\ &\quad = -m \ln \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{m} \ln \left(1 + \xi \frac{M_{ni} - \mu}{\sigma}\right) - \sum_{i=1}^{m} \left(1 + \xi \frac{M_{ni} - \mu}{\sigma}\right)^{-1/\xi}, \end{aligned}$$

which must be maximized subject to the parameter constraints that $\sigma > 0$ and $1 + \xi(M_{ni} - \mu)/\sigma > 0$ for all i. While this represents an irregular likelihood problem, due to the dependence of the parameter space on the values of the data, the consistency and asymptotic efficiency of the resulting MLEs can be established for the case when $\xi > -\frac{1}{2}$ using results in Smith (1985).

In determining the number and size of the blocks (m and n, respectively), a trade-off necessarily takes place: roughly speaking, a large value of n leads to a more accurate approximation of the block maxima distribution by a GEV distribution and a low bias in the parameter estimates; a large value of m gives more block maxima data for the ML estimation and leads to a low variance in the parameter estimates. Note also that, in the case of dependent data, somewhat larger block sizes than are used in the iid case may be advisable; dependence generally has the effect that convergence to the GEV distribution is slower, since the effective sample size is $n\theta$, which is smaller than n.

Example 5.12 (block maxima analysis of S&P return data). Suppose we turn the clock back and imagine it is the early evening of Friday 16 October 1987. An unusually turbulent week in the equity markets has seen the S&P 500 index fall by 9.12%. On that Friday alone the index is down 5.16% on the previous day, the largest one-day fall since 1962.

We fit the GEV distribution to annual maximum daily percentage falls in value for the S&P index. Using data going back to 1960, shown in Figure 5.2, gives us twenty-eight observations of the annual maximum fall (including the latest observation from the incomplete year 1987). The estimated parameter values are $\hat{\xi} = 0.29$, $\hat{\mu} = 2.03$ and $\hat{\sigma} = 0.72$ with standard errors 0.21, 0.16 and 0.14, respectively. Thus the fitted distribution is a heavy-tailed Fréchet distribution with an infinite fourth moment,

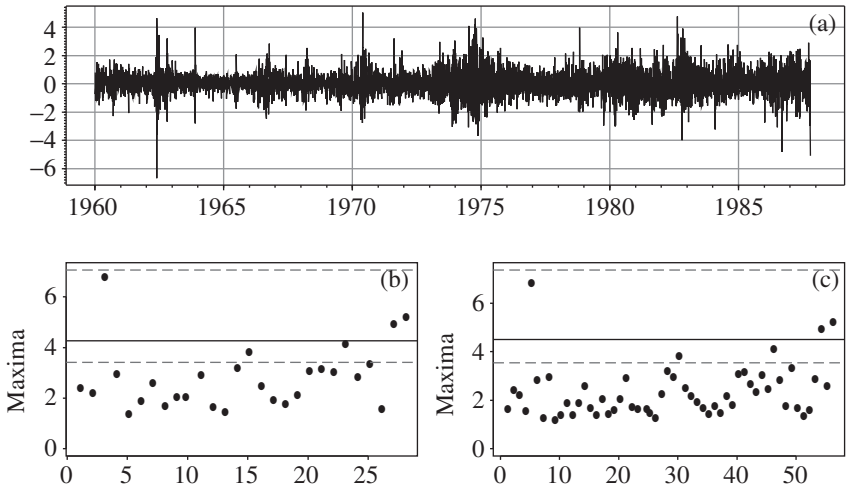

Figure 5.2. (a) S&P percentage returns for the period 1960 to 16 October 1987. (b) Annual maxima of daily falls in the index; superimposed is an estimate of the ten-year return level with associated 95% confidence interval (dotted lines). (c) Semesterly maxima of daily falls in the index; superimposed is an estimate of the 20-semester return level with associated 95% confidence interval. See Examples 5.12 and 5.15 for full details.

suggesting that the underlying distribution is heavy tailed. Note that the standard errors imply considerable uncertainty in our analysis, as might be expected with only twenty-four observations of maxima. In fact, in a likelihood ratio test of the null hypothesis that a Gumbel model fits the data ($H_0 \colon \xi = 0$), the null hypothesis cannot be rejected.

To increase the number of blocks we also fit a GEV model to 56 semesterly maxima and obtain the parameter estimates $\hat{\xi} = 0.33$, $\hat{\mu} = 1.68$ and $\hat{\sigma} = 0.55$ with standard errors 0.14, 0.09 and 0.07. This model has an even heavier tail, and the null hypothesis that a Gumbel model is adequate is now rejected.

Return levels and stress losses. The fitted GEV model can be used to estimate two related quantities that describe the occurrence of *stress events*. On the one hand, we can estimate the size of a stress event that occurs with prescribed frequency (the *return-level* problem). On the other hand, we can estimate the frequency of a stress event that has a prescribed size (the *return-period* problem).

Definition 5.13 (return level). Let H denote the df of the true distribution of the n-block maximum. The k n-block return level is $r_{n,k} = q_{1-1/k}(H)$, i.e. the $(1 - 1/k)$-quantile of H.

The k n-block return level can be roughly interpreted as that level which is exceeded in one out of every k n-blocks on average. For example, the ten-trading-year return level $r_{260,10}$ is that level which is exceeded in one out of every ten years on average. (In the notation we assume that every year has 260 trading days, although this is only an average and there will be slight differences from year to year.) Using

our fitted model we would estimate a return level by

$$\hat{r}_{n,k} = H^{-1}_{\hat{\xi},\hat{\mu},\hat{\sigma}}\left(1 - \frac{1}{k}\right) = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left(\left(-\ln\left(1 - \frac{1}{k}\right)\right)^{-\hat{\xi}} - 1\right). \tag{5.7}$$

Definition 5.14 (return period). Let H denote the df of the true distribution of the n-block maximum. The return period of the event $\{M_n > u\}$ is given by $k_{n,u} = 1/\bar{H}(u)$.

Observe that the return period $k_{n,u}$ is defined in such a way that the $k_{n,u}$ n-block return level is u. In other words, in $k_{n,u}$ n-blocks we would expect to observe a single block in which the level u was exceeded. If there was a strong tendency for the extreme values to cluster, we might expect to see multiple exceedances of the level within that block. Assuming that H is the df of a GEV distribution and using our fitted model, we would estimate the return period by $\hat{k}_{n,u} = 1/\bar{H}_{\hat{\xi},\hat{\mu},\hat{\sigma}}(u)$.

Note that both $\hat{r}_{n,k}$ and $\hat{k}_{n,u}$ are simple functionals of the estimated parameters of the GEV distribution. As well as calculating point estimates for these quantities we should give confidence intervals that reflect the error in the parameter estimates of the GEV distribution. A good method is to base such confidence intervals on the likelihood ratio statistic, as described in Section A.3.5. To do this we reparametrize the GEV distribution in terms of the quantity of interest. For example, in the case of return level, let $\phi = H^{-1}_{\xi,\mu,\sigma}(1 - (1/k))$ and parametrize the GEV distribution by $\boldsymbol{\theta} = (\phi, \xi, \sigma)'$ rather than $\boldsymbol{\theta} = (\xi, \mu, \sigma)'$. The maximum likelihood estimate of ϕ is the estimate (5.7), and a confidence interval can be constructed according to the method in Section A.3.5 (see (A.28) in particular).

Example 5.15 (stress losses for S&P return data). We continue Example 5.12 by estimating the ten-year return level and the 20-semester return level based on data up to 16 October 1987, using (5.7) for the point estimate and the likelihood ratio method as described above to get confidence intervals. The point estimate of the ten-year return level is 4.3% with a 95% confidence interval of (3.4, 7.3); the point estimate of the 20-semester return level is 4.5% with a 95% confidence interval of (3.5, 7.1). Clearly, there is some uncertainty about the size of events of this frequency even with 28 years or 56 semesters of data.

The day after the end of our data set, 19 October 1987, was Black Monday. The index fell by the unprecedented amount of 20.5% in one day. This event is well outside our confidence interval for a ten-year loss. If we were to estimate a 50-year return level (an event beyond our experience if we have 28 years of data), then our point estimate would be 7.2 with a confidence interval of (4.8, 23.4), so the 1987 crash lies close to the upper boundary of our confidence interval for a much rarer event. But the 28 maxima are really too few to get a reliable estimate for an event as rare as the 50-year event.

If we turn the problem around and attempt to estimate the return period of a 20.5% loss, the point estimate is 1631 years (i.e. almost a two-millennium event) but the 95% confidence interval encompasses everything from 42 years to essentially never! The analysis of semesterly maxima gives only moderately more informative

results: the point estimate is 1950 semesters; the confidence interval runs from 121 semesters to 3.0×10^6 semesters. In summary, on 16 October 1987 we simply did not have the data to say anything meaningful about an event of this magnitude. This illustrates the inherent difficulties of attempting to quantify events beyond our empirical experience.

Notes and Comments

The main source for this chapter is Embrechts, Klüppelberg and Mikosch (1997) (EKM). Further important texts on EVT include Gumbel (1958), Leadbetter, Lindgren and Rootzén (1983), Galambos (1987), Resnick (2008), Falk, Hüsler and Reiss (1994), Reiss and Thomas (1997), de Haan and Ferreira (2000), Coles (2001), Beirlant et al. (2004) and Resnick (2007).

The forms of the limit law for maxima were first studied by Fisher and Tippett (1928). The subject was brought to full mathematical fruition in the fundamental papers of Gnedenko (1941, 1943). The concept of the extremal index, which appears in the theory of maxima of stationary series, has a long history. The first mathematically precise definition seems to have been given by Leadbetter (1983). See also Leadbetter, Lindgren and Rootzén (1983) and Smith and Weissman (1994) for more details. The theory required to calculate the extremal index of an ARCH(1) process (as in Table 5.1) is found in de Haan et al. (1989) and also in EKM (pp. 473–480). For the GARCH(1, 1) process, consult Mikosch and Stărică (2000).

A further difficult task is the statistical estimation of the extremal index from time-series data under the assumption that these data do indeed come from a process with an extremal index. Two general methods known as the *blocks* and *runs* methods are described in Section 8.1.3 of EKM; these methods go back to work of Hsing (1991) and Smith and Weissman (1994). Although the estimators have been used in real-world data analyses (see, for example, Davison and Smith 1990), it remains true that the extremal index is a very difficult parameter to estimate accurately.

The maximum likelihood fitting of the GEV distribution is described by Hosking (1985) and Hosking, Wallis and Wood (1985). Consistency and asymptotic normality can be demonstrated for the case $\xi > -0.5$ using results in Smith (1985). An alternative method known as probability-weighted moments (PWM) has been proposed by Hosking, Wallis and Wood (1985) (see also pp. 321–323 of EKM). The analysis of block maxima in Examples 5.12 and 5.15 is based on McNeil (1998). Analyses of financial data using the block maxima method may also be found in Longin (1996), one of the earliest papers to apply EVT methodology to financial data.

5.2 Threshold Exceedances

The block maxima method discussed in Section 5.1.4 has the major defect that it is very wasteful of data; to perform our analyses we retain only the maximum losses in large blocks. For this reason it has been largely superseded in practice by methods based on threshold exceedances, where we use all the data that exceed a particular designated high level.

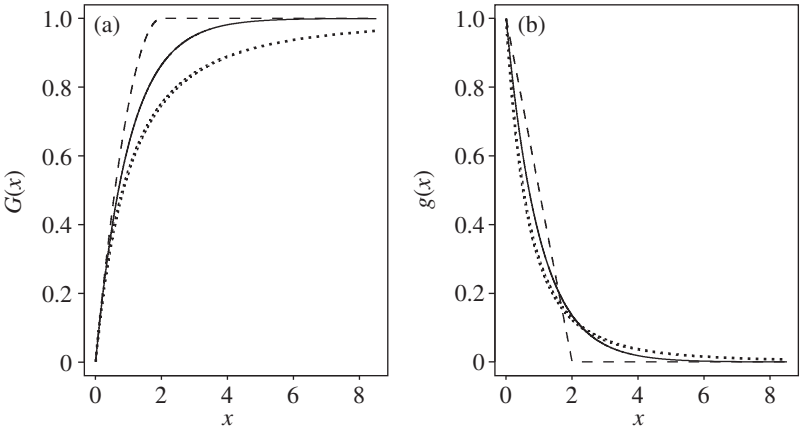

Figure 5.3. (a) Distribution function of the GPD in three cases: the solid line corresponds to $\xi = 0$ (exponential); the dotted line to $\xi = 0.5$ (a Pareto distribution); and the dashed line to $\xi = -0.5$ (Pareto type II). The scale parameter β is equal to 1 in all cases. (b) Corresponding densities.

5.2.1 Generalized Pareto Distribution

The main distributional model for exceedances over thresholds is the generalized Pareto distribution (GPD).

Definition 5.16 (GPD). The df of the GPD is given by

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp(-x/\beta), & \xi = 0, \end{cases} \tag{5.8}$$

where $\beta > 0$, and $x \geqslant 0$ when $\xi \geqslant 0$ and $0 \leqslant x \leqslant -\beta/\xi$ when $\xi < 0$. The parameters ξ and β are referred to, respectively, as the *shape* and *scale* parameters.

Like the GEV distribution in Definition 5.1, the GPD is generalized in the sense that it contains a number of special cases: when $\xi > 0$ the df $G_{\xi,\beta}$ is that of an ordinary Pareto distribution with $\alpha = 1/\xi$ and $\kappa = \beta/\xi$ (see Section A.2.8); when $\xi = 0$ we have an exponential distribution; when $\xi < 0$ we have a short-tailed, Pareto type II distribution. Moreover, as in the case of the GEV distribution, for fixed x the parametric form is continuous in ξ, so $\lim_{\xi \to 0} G_{\xi,\beta}(x) = G_{0,\beta}(x)$. The df and density of the GPD for various values of ξ and $\beta = 1$ are shown in Figure 5.3.

In terms of domains of attraction we have that $G_{\xi,\beta} \in \mathrm{MDA}(H_\xi)$ for all $\xi \in \mathbb{R}$. Note that, for $\xi > 0$ and $\xi < 0$, this assertion follows easily from the characterizations in Theorems 5.8 and 5.10. In the heavy-tailed case, $\xi > 0$, it may be easily verified that $E(X^k) = \infty$ for $k \geqslant 1/\xi$. The mean of the GPD is defined provided $\xi < 1$ and is

$$E(X) = \beta/(1 - \xi). \tag{5.9}$$

The role of the GPD in EVT is as a natural model for the *excess distribution* over a high threshold. We define this concept along with the *mean excess function*, which will also play an important role in the theory.

Definition 5.17 (excess distribution over threshold u). Let X be an rv with df F. The excess distribution over the threshold u has df

$$F_u(x) = P(X - u \leqslant x \mid X > u) = \frac{F(x+u) - F(u)}{1 - F(u)} \tag{5.10}$$

for $0 \leqslant x < x_F - u$, where $x_F \leqslant \infty$ is the right endpoint of F.

Definition 5.18 (mean excess function). The mean excess function of an rv X with finite mean is given by

$$e(u) = E(X - u \mid X > u). \tag{5.11}$$

The excess df F_u describes the distribution of the excess loss over the threshold u, given that u is exceeded. The mean excess function $e(u)$ expresses the mean of F_u as a function of u. In survival analysis the excess df is more commonly known as the residual life df—it expresses the probability that, say, an electrical component that has functioned for u units of time fails in the time period $(u, u + x]$. The mean excess function is known as the mean residual life function and gives the expected residual lifetime for components with different ages. For the special case of the GPD, the excess df and mean excess function are easily calculated.

Example 5.19 (excess distribution of exponential and GPD). If F is the df of an exponential rv, then it is easily verified that $F_u(x) = F(x)$ for all x, which is the famous *lack-of-memory* property of the exponential distribution—the residual lifetime of the aforementioned electrical component would be independent of the amount of time that component has already survived. More generally, if X has df $F = G_{\xi,\beta}$, then, using (5.10), the excess df is easily calculated to be

$$F_u(x) = G_{\xi,\beta(u)}(x), \quad \beta(u) = \beta + \xi u, \tag{5.12}$$

where $0 \leqslant x < \infty$ if $\xi \geqslant 0$ and $0 \leqslant x \leqslant -(\beta/\xi) - u$ if $\xi < 0$. The excess distribution remains a GPD with the same shape parameter ξ but with a scaling that grows linearly with the threshold u. The mean excess function of the GPD is easily calculated from (5.12) and (5.9) to be

$$e(u) = \frac{\beta(u)}{1-\xi} = \frac{\beta + \xi u}{1 - \xi}, \tag{5.13}$$

where $0 \leqslant u < \infty$ if $0 \leqslant \xi < 1$ and $0 \leqslant u \leqslant -\beta/\xi$ if $\xi < 0$. It may be observed that the mean excess function is *linear in the threshold* u, which is a characterizing property of the GPD.

Example 5.19 shows that the GPD has a kind of stability property under the operation of calculating excess distributions. We now give a mathematical result that shows that the GPD is, in fact, a natural limiting excess distribution for many underlying loss distributions. The result can also be viewed as a characterization theorem for the maximum domain of attraction of the GEV distribution. In Section 5.1.2 we looked separately at characterizations for each of the three cases $\xi > 0$, $\xi = 0$ and $\xi < 0$; the following result offers a global characterization of $\mathrm{MDA}(H_\xi)$ for all ξ in terms of the limiting behaviour of excess distributions over thresholds.

Theorem 5.20 (Pickands–Balkema–de Haan). *We can find a (positive-measurable function) $\beta(u)$ such that*

$$\lim_{u \to x_F} \sup_{0 \leqslant x < x_F - u} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0$$

if and only if $F \in \mathrm{MDA}(H_\xi)$, $\xi \in \mathbb{R}$.

Thus the distributions for which normalized maxima converge to a GEV distribution constitute a set of distributions for which the excess distribution converges to the GPD as the threshold is raised; moreover, the shape parameter of the limiting GPD for the excesses is the same as the shape parameter of the limiting GEV distribution for the maxima. We have already stated in Section 5.1.2 that essentially all the commonly used continuous distributions of statistics are in $\mathrm{MDA}(H_\xi)$ for some ξ, so Theorem 5.20 proves to be a very widely applicable result that essentially says that the GPD is *the canonical distribution* for modelling excess losses over high thresholds.

5.2.2 Modelling Excess Losses

We exploit Theorem 5.20 by assuming that we are dealing with a loss distribution $F \in \mathrm{MDA}(H_\xi)$ so that, for some suitably chosen high threshold u, we can model F_u by a generalized Pareto distribution. We formalize this with the following assumption.

Assumption 5.21. *Let F be a loss distribution with right endpoint x_F and assume that for some high threshold u we have $F_u(x) = G_{\xi,\beta}(x)$ for $0 \leqslant x < x_F - u$ and some $\xi \in \mathbb{R}$ and $\beta > 0$.*

This is clearly an idealization, since in practice the excess distribution will generally not be *exactly* GPD, but we use Assumption 5.21 to make a number of calculations in the following sections.

The method. Given loss data $X_1, \ldots, X_n$ from F, a random number N_u will exceed our threshold u; it will be convenient to relabel these data $\tilde{X}_1, \ldots, \tilde{X}_{N_u}$. For each of these exceedances we calculate the amount $Y_j = \tilde{X}_j - u$ of the excess loss. We wish to estimate the parameters of a GPD model by fitting this distribution to the N_u excess losses. There are various ways of fitting the GPD, including maximum likelihood (ML) and probability-weighted moments (PWM). The former method is more commonly used and is easy to implement if the excess data can be assumed to be realizations of independent rvs, since the joint density will then be a product of marginal GPD densities.

Writing $g_{\xi,\beta}$ for the density of the GPD, the log-likelihood may be easily calculated to be

$$\begin{aligned} \ln L(\xi, \beta; Y_1, \ldots, Y_{N_u}) &= \sum_{j=1}^{N_u} \ln g_{\xi,\beta}(Y_j) \\ &= -N_u \ln \beta - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^{N_u} \ln\left(1 + \xi \frac{Y_j}{\beta}\right), \end{aligned} \tag{5.14}$$

which must be maximized subject to the parameter constraints that $\beta > 0$ and $1 + \xi Y_j/\beta > 0$ for all j. Solving the maximization problem yields a GPD model $G_{\hat{\xi},\hat{\beta}}$ for the excess distribution F_u.

Non-iid data. For insurance or operational risk data the iid assumption is often unproblematic, but this is clearly not true for time series of financial returns. If the data are serially dependent but show no tendency to give clusters of extreme values, then this might suggest that the underlying process has extremal index $\theta = 1$. In this case, asymptotic theory that we summarize in Section 5.3 suggests a limiting model for high-level threshold exceedances, in which exceedances occur according to a Poisson process and the excess loss amounts are iid generalized Pareto distributed. If extremal clustering is present, suggesting an extremal index $\theta < 1$ (as would be consistent with an underlying GARCH process), the assumption of independent excess losses is less satisfactory. The easiest approach is to neglect this problem and to consider the ML method to be a QML method, where the likelihood is misspecified with respect to the serial dependence structure of the data; we follow this course in this section. The point estimates should still be reasonable, although standard errors may be too small. In Section 5.3 we discuss threshold exceedances in non-iid data in more detail.

Excesses over higher thresholds. From the model we have fitted to the excess distribution over u, we can easily infer a model for the excess distribution over any higher threshold. We have the following lemma.

Lemma 5.22. *Under Assumption 5.21 it follows that* $F_v(x) = G_{\xi,\beta+\xi(v-u)}(x)$ *for any higher threshold* $v \geqslant u$.

Proof. We use (5.10) and the df of the GPD in (5.8) to infer that

$$\begin{aligned}\bar{F}_v(x) &= \frac{\bar{F}(v+x)}{\bar{F}(v)} = \frac{\bar{F}(u+(x+v-u))}{\bar{F}(u)}\frac{\bar{F}(u)}{\bar{F}(u+(v-u))} \\ &= \frac{\bar{F}_u(x+v-u)}{\bar{F}_u(v-u)} = \frac{\bar{G}_{\xi,\beta}(x+v-u)}{\bar{G}_{\xi,\beta}(v-u)} \\ &= \bar{G}_{\xi,\beta+\xi(v-u)}(x).\end{aligned}$$

□

Thus the excess distribution over higher thresholds remains a GPD with the same ξ parameter but a scaling that grows linearly with the threshold v. Provided that $\xi < 1$, the mean excess function is given by

$$e(v) = \frac{\beta + \xi(v-u)}{1-\xi} = \frac{\xi v}{1-\xi} + \frac{\beta - \xi u}{1-\xi}, \tag{5.15}$$

where $u \leqslant v < \infty$ if $0 \leqslant \xi < 1$ and $u \leqslant v \leqslant u - \beta/\xi$ if $\xi < 0$.

The linearity of the mean excess function (5.15) in v is commonly used as a diagnostic for data admitting a GPD model for the excess distribution. It forms the basis for the following simple graphical method for choosing an appropriate threshold.

Sample mean excess plot. For positive-valued loss data $X_1, \ldots, X_n$ we define the *sample mean excess function* to be an empirical estimator of the mean excess function in Definition 5.18. The estimator is given by

$$e_n(v) = \frac{\sum_{i=1}^{n}(X_i - v)I_{\{X_i > v\}}}{\sum_{i=1}^{n} I_{\{X_i > v\}}}. \tag{5.16}$$

To study the sample mean excess function we generally construct the mean excess plot $\{(X_{i,n}, e_n(X_{i,n})) : 2 \leqslant i \leqslant n\}$, where $X_{i,n}$ denotes the upper (or descending) ith order statistic. If the data support a GPD model over a high threshold, then (5.15) suggests that this plot should become increasingly "linear" for higher values of v. A linear upward trend indicates a GPD model with positive shape parameter ξ; a plot tending towards the horizontal indicates a GPD with approximately zero shape parameter, or, in other words, an exponential excess distribution; a linear downward trend indicates a GPD with negative shape parameter.

These are the ideal situations, but in practice some experience is required to read mean excess plots. Even for data that are genuinely generalized Pareto distributed, the sample mean excess plot is seldom perfectly linear, particularly towards the right-hand end, where we are averaging a small number of large excesses. In fact, we often omit the final few points from consideration, as they can severely distort the picture. If we do see visual evidence that the mean excess plot becomes linear, then we might select as our threshold u a value towards the beginning of the linear section of the plot (see, in particular, Example 5.24).

Example 5.23 (Danish fire loss data). The Danish fire insurance data are a well-studied set of financial losses that neatly illustrate the basic ideas behind modelling observations that seem consistent with an iid model. The data set consists of 2156 fire insurance losses over 1 000 000 Danish kroner from 1980 to 1990 inclusive, expressed in units of 1 000 000 kroner. The loss figure represents a combined loss for a building and its contents, as well as in some cases a loss of business earnings; the losses are inflation adjusted to reflect 1985 values and are shown in Figure 5.4 (a).

The sample mean excess plot in Figure 5.4 (b) is in fact fairly "linear" over the entire range of the losses, and its upward slope leads us to expect that a GPD with positive shape parameter ξ could be fitted to the entire data set. However, there is some evidence of a "kink" in the plot below the value 10 and a "straightening out" of the plot above this value, so we have chosen to set our threshold at $u = 10$ and fit a GPD to excess losses above this threshold, in the hope of obtaining a model that is a good fit to the largest of the losses. The ML parameter estimates are $\hat{\xi} = 0.50$ and $\hat{\beta} = 7.0$ with standard errors 0.14 and 1.1, respectively. Thus the model we have fitted is essentially a very heavy-tailed, infinite-variance model. A picture of the fitted GPD model for the excess distribution $\hat{F}_u(x - u)$ is also given in Figure 5.4 (c), superimposed on points plotted at empirical estimates of the excess probabilities for each loss; note the good correspondence between the empirical estimates and the GPD curve.

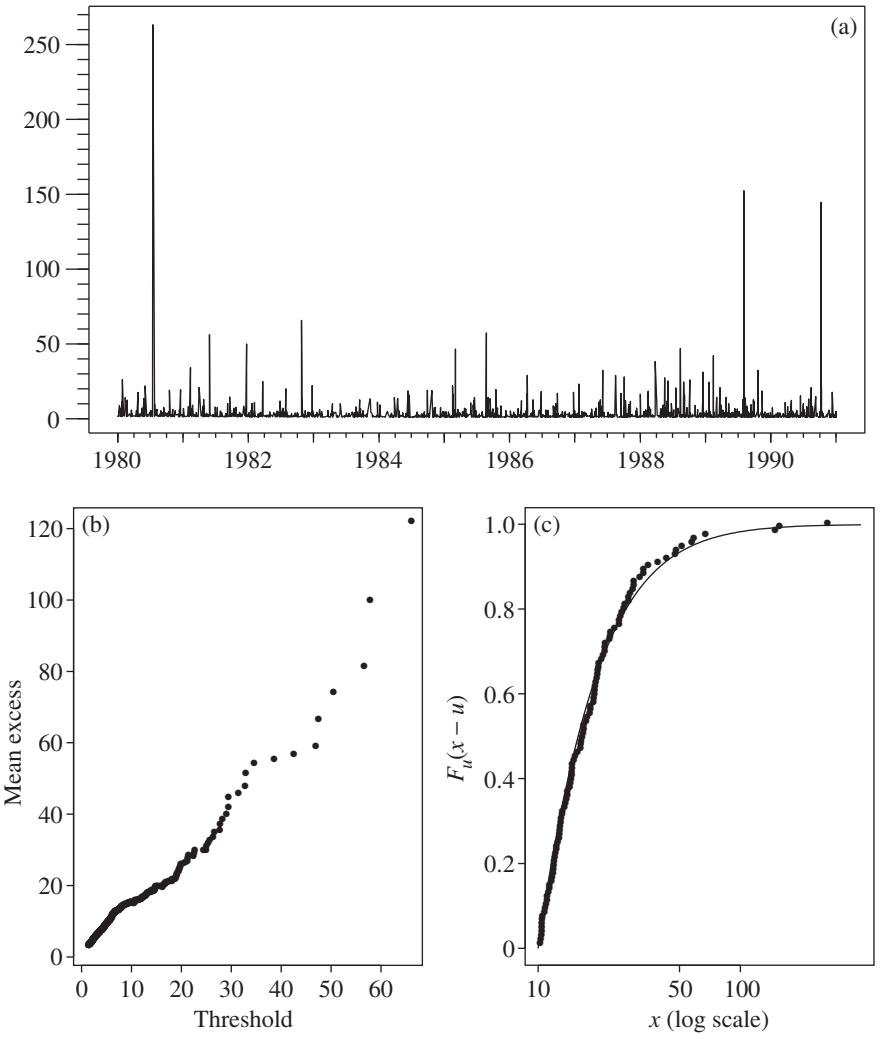

Figure 5.4. (a) Time-series plot of the Danish data. (b) Sample mean excess plot. (c) Empirical distribution of excesses and fitted GPD. See Example 5.23 for full details.

In insurance we might use the model to estimate the expected size of the insurance loss, given that it enters a given insurance *layer*. Thus we can estimate the expected loss size given exceedance of the threshold of 10 000 000 kroner or of any other higher threshold by using (5.15) with the appropriate parameter estimates.

Example 5.24 (AT&T weekly loss data). Suppose we have an investment in AT&T stock and want to model weekly losses in value using an unconditional approach. If X_t denotes the weekly log-return, then the percentage loss in value of our position over a week is given by $L_t = 100(1 - \exp(X_t))$, and data on this loss for the 521 complete weeks in the period 1991–2000 are shown in Figure 5.5 (a).

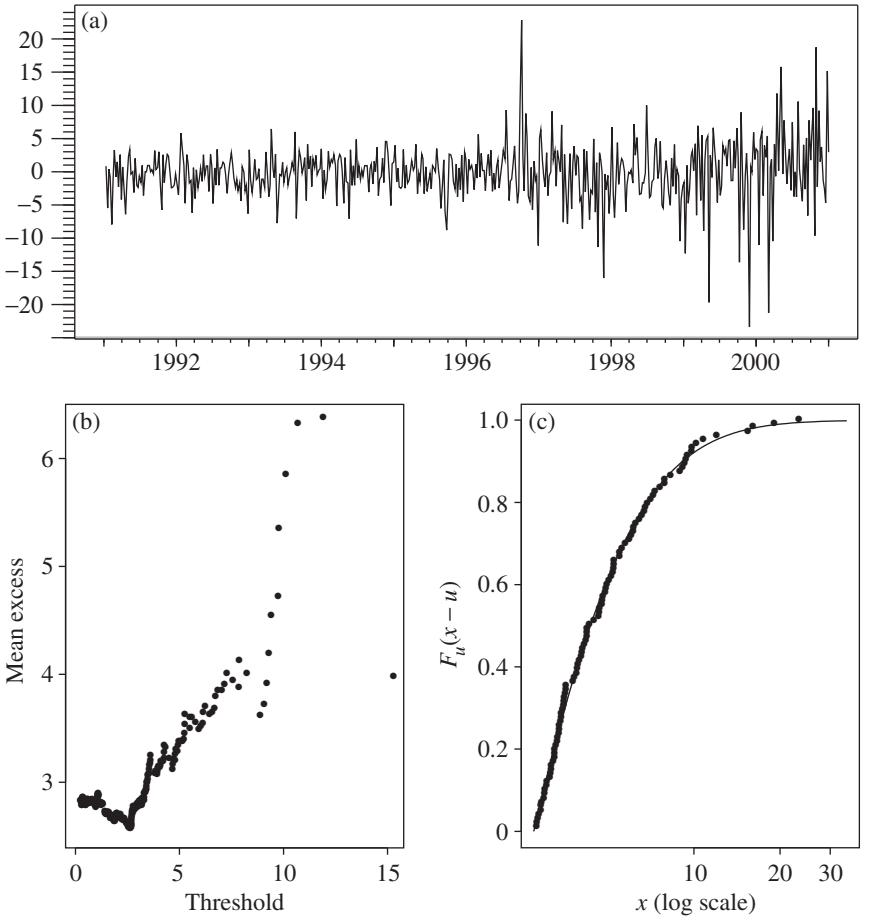

Figure 5.5. (a) Time-series plot of AT&T weekly percentage loss data. (b) Sample mean excess plot. (c) Empirical distribution of excesses and fitted GPD. See Example 5.24 for full details.

A sample mean excess plot of the positive loss values is shown in Figure 5.5 (b) and this suggests that a threshold can be found above which a GPD approximation to the excess distribution should be possible. We have chosen to position the threshold at a loss value of 2.75% and this gives 102 exceedances.

We observed in Section 3.1 that monthly AT&T return data over the period 1993–2000 do not appear consistent with a strict white noise hypothesis, so the issue of whether excess losses can be modelled as independent is relevant. This issue is taken up in Section 5.3 but for the time being we ignore it and implement a standard ML approach to estimating the parameters of a GPD model for the excess distribution; we obtain the estimates $\hat{\xi} = 0.22$ and $\hat{\beta} = 2.1$ with standard errors 0.13 and 0.34, respectively. Thus the model we have fitted is a model that is close to having an infinite fourth moment. A picture of the fitted GPD model for the excess distribution $\hat{F}_u(x-u)$ is also given in Figure 5.5 (c), superimposed on points plotted at empirical estimates of the excess probabilities for each loss.

5.2.3 Modelling Tails and Measures of Tail Risk

In this section we describe how the GPD model for the excess losses is used to estimate the tail of the underlying loss distribution F and associated risk measures. To make the necessary theoretical calculations we again make Assumption 5.21.

Tail probabilities and risk measures. We observe firstly that under Assumption 5.21 we have, for $x \geqslant u$,

$$\begin{aligned}\bar{F}(x) &= P(X > u)P(X > x \mid X > u)\\ &= \bar{F}(u)P(X - u > x - u \mid X > u)\\ &= \bar{F}(u)\bar{F}_u(x - u)\\ &= \bar{F}(u)\left(1 + \xi\frac{x-u}{\beta}\right)^{-1/\xi}, \end{aligned} \tag{5.17}$$

which, if we know $F(u)$, gives us a formula for tail probabilities. This formula may be inverted to obtain a high quantile of the underlying distribution, which we interpret as a VaR. For $\alpha \geqslant F(u)$ we have that VaR is equal to

$$\mathrm{VaR}_\alpha = q_\alpha(F) = u + \frac{\beta}{\xi}\left(\left(\frac{1-\alpha}{\bar{F}(u)}\right)^{-\xi} - 1\right). \tag{5.18}$$

Assuming that $\xi < 1$, the associated expected shortfall can be calculated easily from (2.22) and (5.18). We obtain

$$\mathrm{ES}_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 q_x(F)\,\mathrm{d}x = \frac{\mathrm{VaR}_\alpha}{1-\xi} + \frac{\beta - \xi u}{1-\xi}. \tag{5.19}$$

Note that Assumption 5.21 and Lemma 5.22 imply that excess losses above VaR_α have a GPD distribution satisfying $F_{\mathrm{VaR}_\alpha} = G_{\xi,\beta+\xi(\mathrm{VaR}_\alpha - u)}$. The expected shortfall estimator in (5.19) can also be obtained by adding the mean of this distribution to VaR_α, i.e. $\mathrm{ES}_\alpha = \mathrm{VaR}_\alpha + e(\mathrm{VaR}_\alpha)$, where $e(\mathrm{VaR}_\alpha)$ is given in (5.15). It is interesting to look at how the *ratio* of the two risk measures behaves for large values of the quantile probability α. It is easily calculated from (5.18) and (5.19) that

$$\lim_{\alpha\to 1}\frac{\mathrm{ES}_\alpha}{\mathrm{VaR}_\alpha} = \begin{cases}(1-\xi)^{-1}, & 0 \leqslant \xi < 1,\\ 1, & \xi < 0,\end{cases} \tag{5.20}$$

so the shape parameter ξ of the GPD effectively determines the ratio when we go far enough out into the tail.

Estimation in practice. We note that, under Assumption 5.21, tail probabilities, VaRs and expected shortfalls are all given by formulas of the form $g(\xi, \beta, \bar{F}(u))$. Assuming that we have fitted a GPD to excess losses over a threshold u, as described in Section 5.2.2, we estimate these quantities by first replacing ξ and β in formulas (5.17)–(5.19) by their estimates. Of course, we also require an estimate of $\bar{F}(u)$ and here we take the simple empirical estimator N_u/n. In doing this, we are implicitly assuming that there is a sufficient proportion of sample values above the threshold u to estimate $\bar{F}(u)$ reliably. However, we hope to gain over the empirical method by

using a kind of extrapolation based on the GPD for more extreme tail probabilities and risk measures. For tail probabilities we obtain an estimator, first proposed by Smith (1987), of the form

$$\hat{\bar{F}}(x) = \frac{N_u}{n}\left(1 + \hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}}, \tag{5.21}$$

which we stress is only valid for $x \geqslant u$. For $\alpha \geqslant 1 - N_u/n$ we obtain analogous point estimators of VaR_α and ES_α from (5.18) and (5.19).

Of course, we would also like to obtain confidence intervals. If we have taken the likelihood approach to estimating ξ and β, then it is quite easy to give confidence intervals for $g(\hat{\xi}, \hat{\beta}, N_u/n)$ that take into account the uncertainty in $\hat{\xi}$ and $\hat{\beta}$, but neglect the uncertainty in N_u/n as an estimator of $\bar{F}(u)$. We use the approach described at the end of Section 5.1.4 for return levels, whereby the GPD model is reparametrized in terms of $\phi = g(\xi, \beta, N_u/n)$, and a confidence interval for $\hat{\phi}$ is constructed based on the likelihood ratio test as in Section A.3.5.

Example 5.25 (risk measures for AT&T loss data). Suppose we have fitted a GPD model to excess weekly losses above the threshold $u = 2.75\%$, as in Example 5.24. We use this model to obtain estimates of the 99% VaR and expected shortfall of the underlying weekly loss distribution. The essence of the method is displayed in Figure 5.6; this is a plot of estimated tail probabilities on logarithmic axes, with various dotted lines superimposed to indicate the estimation of risk measures and associated confidence intervals. The points on the graph are the 102 threshold exceedances and are plotted at y-values corresponding to the tail of the empirical distribution function; the smooth curve running through the points is the tail estimator (5.21).

Estimation of the 99% quantile amounts to determining the point of intersection of the tail estimation curve and the horizontal line $\bar{F}(x) = 0.01$ (not marked on the graph); the first vertical dotted line shows the quantile estimate. The horizontal dotted line aids in the visualization of a 95% confidence interval for the VaR estimate; the degree of confidence is shown on the alternative y-axis to the right of the plot. The boundaries of a 95% confidence interval are obtained by determining the two points of intersection of this horizontal line with the dotted curve, which is a profile likelihood curve for the VaR as a parameter of the GPD model and is constructed using likelihood ratio test arguments as in Section A.3.5. Dropping the horizontal line to the 99% mark would correspond to constructing a 99% confidence interval for the estimate of the 99% VaR. The point estimate and the 95% confidence interval for the 99% quantile are estimated to be 11.7% and (9.6, 16.1).

The second vertical line on the plot shows the point estimate of the 99% expected shortfall. A 95% confidence interval is determined from the dotted horizontal line and its points of intersection with the second dotted curve. The point estimate and the 95% confidence interval are 17.0% and (12.7, 33.6). Note that if we take the ratio of the point estimates of the shortfall and the VaR, we get $17/11.7 \approx 1.45$, which is larger than the asymptotic ratio $(1 - \hat{\xi})^{-1} = 1.29$ suggested by (5.20); this is generally the case at finite levels and is explained by the second term in (5.19) being a non-negligible positive quantity.

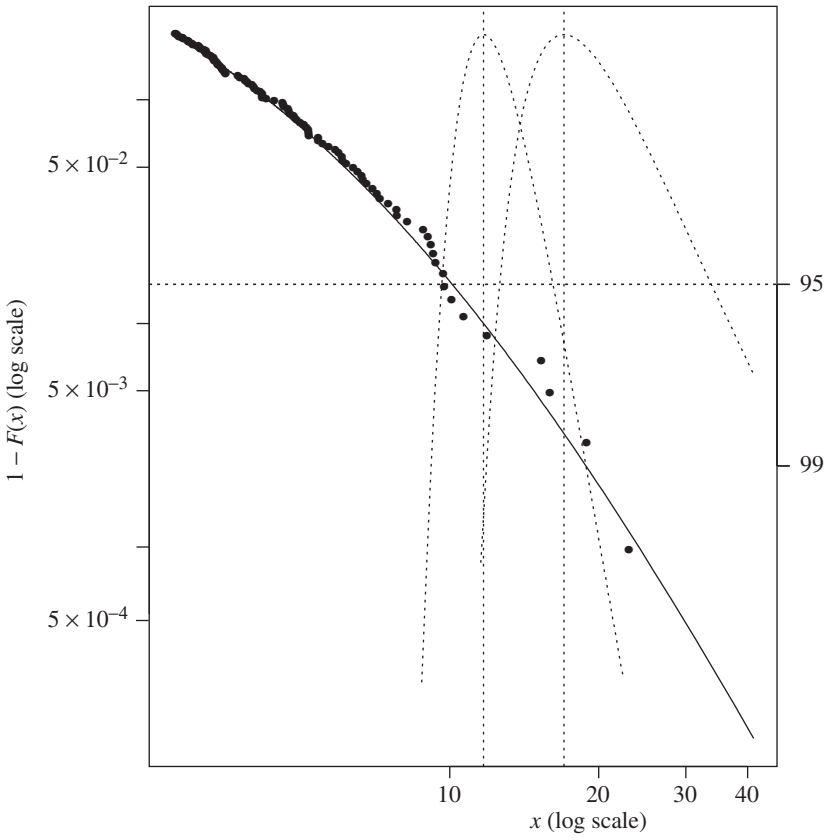

Figure 5.6. The smooth curve through the points shows the estimated tail of the AT&T weekly percentage loss data using the estimator (5.21). Points are plotted at empirical tail probabilities calculated from empirical df. The vertical dotted lines show estimates of 99% VaR and expected shortfall. The other curves are used in the construction of confidence intervals. See Example 5.25 for full details.

Before leaving the topic of GPD tail modelling it is clearly important to see how sensitive our risk-measure estimates are to the choice of the threshold. Hitherto, we have considered single choices of threshold u and looked at a series of incremental calculations that always build on the same GPD model for excesses over that threshold. We would hope that there is some robustness to our inference for different choices of threshold.

Example 5.26 (varying the threshold). In the case of the AT&T weekly loss data the influence of different thresholds is investigated in Figure 5.7. Given the importance of the ξ parameter in determining the weight of the tail and the relationship between quantiles and expected shortfalls, we first show how estimates of ξ vary as we consider a series of thresholds that give us between 20 and 150 exceedances. In fact, the estimates remain fairly constant around a value of approximately 0.2; a symmetric 95% confidence interval constructed from the standard error estimate is also shown, and it indicates how the uncertainty about the parameter

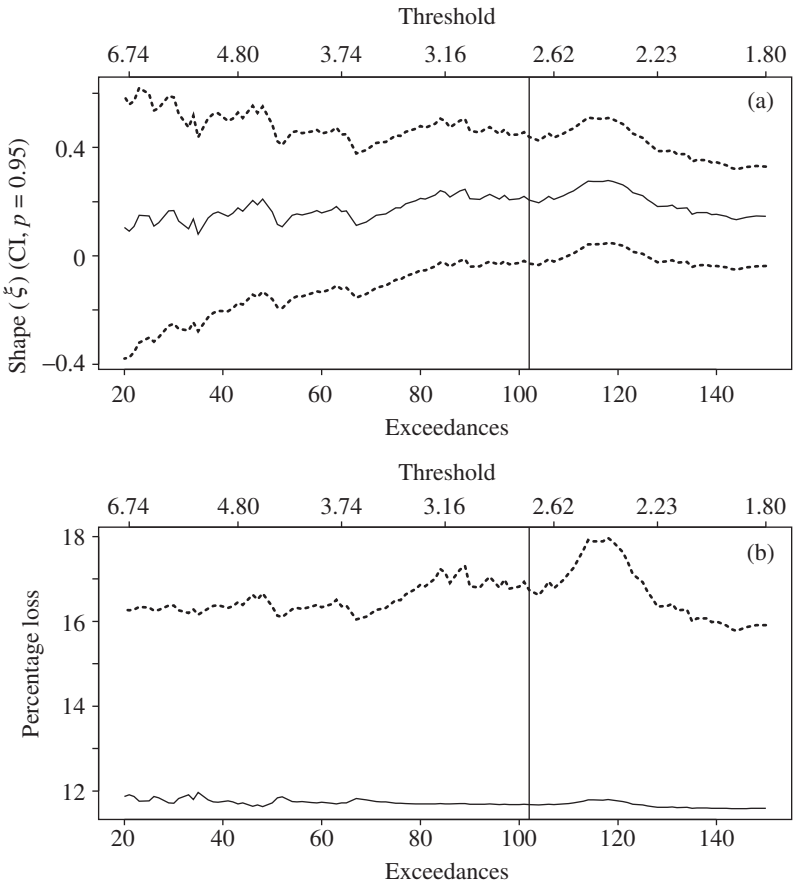

Figure 5.7. (a) Estimate of ξ for different thresholds u and numbers of exceedances N_u, together with a 95% confidence interval based on the standard error. (b) Associated point estimates of the 99% VaR (solid line) and the expected shortfall (dotted line). See Example 5.26 for commentary.

value decreases as the threshold is lowered or the number of threshold exceedances is increased.

Point estimates of the 99% VaR and expected shortfall estimates are also shown. The former remain remarkably constant around 12%, while the latter show modest variability that essentially tracks the variability of the ξ estimate. These pictures provide some reassurance that different thresholds do not lead to drastically different conclusions. We return to the issue of threshold choice again in Section 5.2.5.

5.2.4 The Hill Method

The GPD method is not the only way to estimate the tail of a distribution and, as an alternative, we describe in this section the well-known Hill approach to modelling the tails of heavy-tailed distributions.

Estimating the tail index. For this method we assume that the underlying loss distribution is in the maximum domain of attraction of the Fréchet distribution so that, by Theorem 5.8, it has a tail of the form

$$\bar{F}(x) = x^{-\alpha} L(x) \tag{5.22}$$

for a slowly varying function L (see Definition 5.7) and a positive parameter α. Traditionally, in the Hill approach, interest centres on the *tail index* α, rather than its reciprocal ξ, which appears in (5.3). The goal is to find an estimator of α based on identically distributed data $X_1, \ldots, X_n$.

The Hill estimator can be derived in various ways (see EKM, pp. 330–336). Perhaps the most elegant is to consider the mean excess function of the generic logarithmic loss $\ln X$, where X is an rv with df (5.22). Writing e^* for the mean excess function of $\ln X$ and using integration by parts we find that

$$\begin{aligned} e^*(\ln u) &= E(\ln X - \ln u \mid \ln X > \ln u) \\ &= \frac{1}{\bar{F}(u)} \int_u^\infty (\ln x - \ln u)\, \mathrm{d}F(x) \\ &= \frac{1}{\bar{F}(u)} \int_u^\infty \frac{\bar{F}(x)}{x}\, \mathrm{d}x \\ &= \frac{1}{\bar{F}(u)} \int_u^\infty L(x) x^{-(\alpha+1)}\, \mathrm{d}x. \end{aligned}$$

For u sufficiently large, the slowly varying function $L(x)$ for $x \geqslant u$ can essentially be treated as a constant and taken outside the integral. More formally, using Karamata's Theorem (see Section A.1.4), we get, for $u \to \infty$,

$$e^*(\ln u) \sim \frac{L(u) u^{-\alpha} \alpha^{-1}}{\bar{F}(u)} = \alpha^{-1},$$

so $\lim_{u\to\infty} \alpha e^*(\ln u) = 1$. We expect to see similar tail behaviour in the sample mean excess function e_n^* (see (5.16)) constructed from the log observations. That is, we expect that $e_n^*(\ln X_{k,n}) \approx \alpha^{-1}$ for n large and k sufficiently small, where $X_{n,n} \leqslant \cdots \leqslant X_{1,n}$ are the order statistics as usual. Evaluating $e_n^*(\ln X_{k,n})$ gives us the estimator $\hat{\alpha}^{-1} = ((k-1)^{-1} \sum_{j=1}^{k-1} \ln X_{j,n} - \ln X_{k,n})$. The standard form of the Hill estimator is obtained by a minor modification:

$$\hat{\alpha}_{k,n}^{(\mathrm{H})} = \left(\frac{1}{k} \sum_{j=1}^{k} \ln X_{j,n} - \ln X_{k,n} \right)^{-1}, \quad 2 \leqslant k \leqslant n. \tag{5.23}$$

The Hill estimator is one of the best-studied estimators in the EVT literature. The asymptotic properties (consistency, asymptotic normality) of this estimator (as sample size $n \to \infty$, number of extremes $k \to \infty$ and the so-called tail-fraction $k/n \to 0$) have been extensively investigated under various assumed models for the data, including ARCH and GARCH (see Notes and Comments). We concentrate on the use of the estimator in practice and, in particular, on its performance relative to the GPD estimation approach.

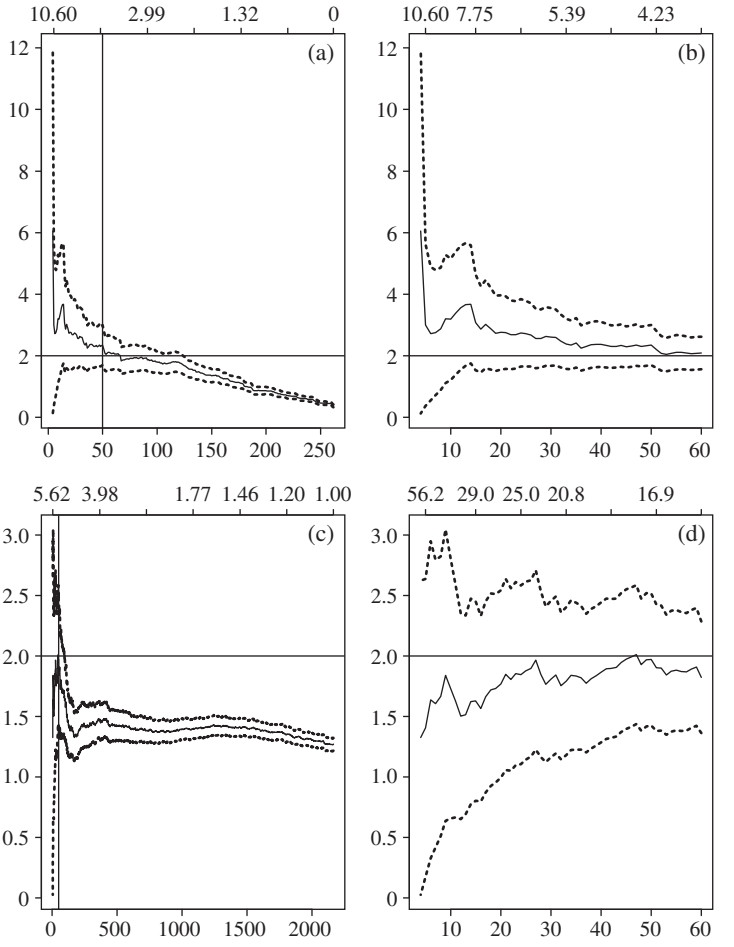

Figure 5.8. Hill plots showing estimates of the tail index $\alpha = 1/\xi$ for (a), (b) the AT&T weekly percentages losses and (c), (d) the Danish fire loss data. Parts (b) and (d) are expanded versions of sections of (a) and (c) showing Hill estimates based on up to 60 order statistics.

When the data are from a distribution with a tail that is close to a perfect power function, the Hill estimator is often a good estimator of α, or its reciprocal ξ. In practice, the general strategy is to plot Hill estimates for various values of k. This gives the Hill plot $\{(k, \hat{\alpha}_{k,n}^{(\mathrm{H})}) : k = 2, \ldots, n\}$. We hope to find a stable region in the Hill plot where estimates constructed from different numbers of order statistics are quite similar.

Example 5.27 (Hill plots). We construct Hill plots for the Danish fire data of Example 5.23 and the weekly percentage loss data (positive values only) of Example 5.24 (shown in Figure 5.8).

It is very easy to construct the Hill plot for all possible values of k, but it can be misleading to do so; practical experience (see Example 5.28) suggests that the best choices of k are relatively small: say, 10–50 order statistics in a sample of size 1000.

For this reason we have enlarged sections of the Hill plots showing the estimates obtained for values of k less than 60.

For the Danish data, the estimates of α obtained are between 1.5 and 2, suggesting ξ estimates between 0.5 and 0.67, all of which correspond to infinite-variance models for these data. Recall that the estimate derived from our GPD model in Example 5.23 was $\hat{\xi} = 0.50$. For the AT&T data, there is no particularly stable region in the plot. The α estimates based on $k = 2, \ldots, 60$ order statistics mostly range from 2 to 4, suggesting a ξ value in the range 0.25–0.5, which is larger than the values estimated in Example 5.26 with a GPD model.

Example 5.27 shows that the interpretation of Hill plots can be difficult. In practice, various deviations from the ideal situation can occur. If the data do not come from a distribution with a regularly varying tail, the Hill method is really not appropriate and Hill plots can be very misleading. Serial dependence in the data can also spoil the performance of the estimator, although this is also true for the GPD estimator. EKM contains a number of Hill "horror plots" based on simulated data illustrating the issues that arise (see Notes and Comments).

Hill-based tail estimates. For the risk-management applications of this book we are less concerned with estimating the tail index of heavy-tailed data and more concerned with tail and risk-measure estimates. We give a heuristic argument for a standard tail estimator based on the Hill approach. We assume a tail model of the form $\bar{F}(x) = Cx^{-\alpha}$, $x \geqslant u > 0$, for some high threshold u; in other words, we replace the slowly varying function by a constant for sufficiently large x. For an appropriate value of k the tail index α is estimated by $\hat{\alpha}^{(\mathrm{H})}_{k,n}$ and the threshold u is replaced by $X_{k,n}$ (or $X_{(k+1),n}$ in some versions); it remains to estimate C. Since C can be written as $C = u^{\alpha}\bar{F}(u)$, this is equivalent to estimating $\bar{F}(u)$, and the obvious empirical estimator is k/n (or $(k-1)/n$ in some versions). Putting these ideas together gives us the *Hill tail estimator* in its standard form:

$$\hat{\bar{F}}(x) = \frac{k}{n}\left(\frac{x}{X_{k,n}}\right)^{-\hat{\alpha}^{(\mathrm{H})}_{k,n}}, \quad x \geqslant X_{k,n}. \tag{5.24}$$

Writing the estimator in this way emphasizes the way it is treated mathematically. For any pair k and n, both the Hill estimator and the associated tail estimator are treated as functions of the k upper-order statistics from the sample of size n. Obviously, it is possible to invert this estimator to get a quantile estimator and it is also possible to devise an estimator of expected shortfall using arguments about regularly varying tails.

The GPD-based tail estimator (5.21) is usually treated as a function of a random number N_u of upper-order statistics for a fixed threshold u. The different presentation of these estimators in the literature is a matter of convention and we can easily recast both estimators in a similar form. Suppose we rewrite (5.24) in the notation of (5.21) by substituting $\hat{\xi}^{(\mathrm{H})}$, u and N_u for $1/\hat{\alpha}^{(\mathrm{H})}_{k,n}$, $X_{k,n}$ and k, respectively. We get

$$\hat{\bar{F}}(x) = \frac{N_u}{n}\left(1 + \hat{\xi}^{(\mathrm{H})}\frac{x-u}{\hat{\xi}^{(\mathrm{H})}u}\right)^{-1/\hat{\xi}^{(\mathrm{H})}}, \quad x \geqslant u.$$

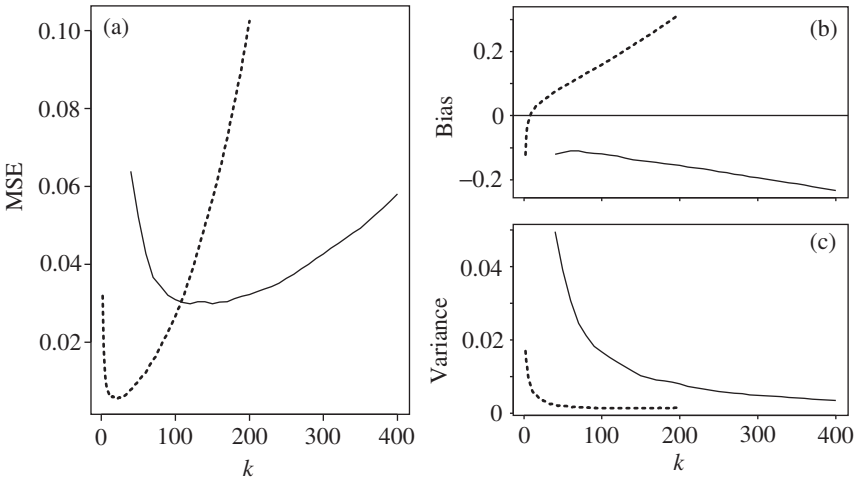

Figure 5.9. Comparison of (a) estimated MSE, (b) bias and (c) variance for the Hill (dotted line) and GPD (solid line) estimators of ξ, the reciprocal of the tail index, as a function of k (or N_u), the number of upper-order statistics from a sample of 1000 t-distributed data with four degrees of freedom. See Example 5.28 for details.

This estimator lacks the additional scaling parameter β in (5.21) and tends not to perform as well, as is shown in simulated examples in the next section.

5.2.5 *Simulation Study of EVT Quantile Estimators*

First we consider estimation of ξ and then estimation of the high quantile VaR_α. In both cases estimators are compared using mean squared errors (MSEs); we recall that the MSE of an estimator $\hat{\theta}$ of a parameter θ is given by $\text{MSE}(\hat{\theta}) = E(\hat{\theta}-\theta)^2 = (E(\hat{\theta}-\theta))^2 + \text{var}(\hat{\theta})$, and thus has the well-known decomposition into *squared bias* plus *variance*. A good estimator should keep both the bias term $E(\hat{\theta}-\theta)$ and the variance term $\text{var}(\hat{\theta})$ small.

Since analytical evaluation of bias and variance is not possible, we calculate Monte Carlo estimates by simulating 1000 data sets in each experiment. The parameters of the GPD are determined in all cases by ML; PWM, the main alternative, gives slightly different results, but the conclusions are similar.

We calculate estimates using the Hill method and the GPD method based on different numbers of upper-order statistics (or differing thresholds) and try to determine the choice of k (or N_u) that is most appropriate for a sample of size n. In the case of estimating VaR we also compare the EVT estimators with the simple empirical quantile estimator.

Example 5.28 (Monte Carlo experiment). We assume that we have a sample of 1000 iid data from a t distribution with four degrees of freedom and want to estimate ξ, the reciprocal of the tail index, which in this case has the true value 0.25. (This is demonstrated in Example 16.1.) The Hill estimate is constructed for k values in the range $\{2, \ldots, 200\}$, and the GPD estimate is constructed for k (or N_u) values in $\{30, 40, 50, \ldots, 400\}$. The results are shown in Figure 5.9.

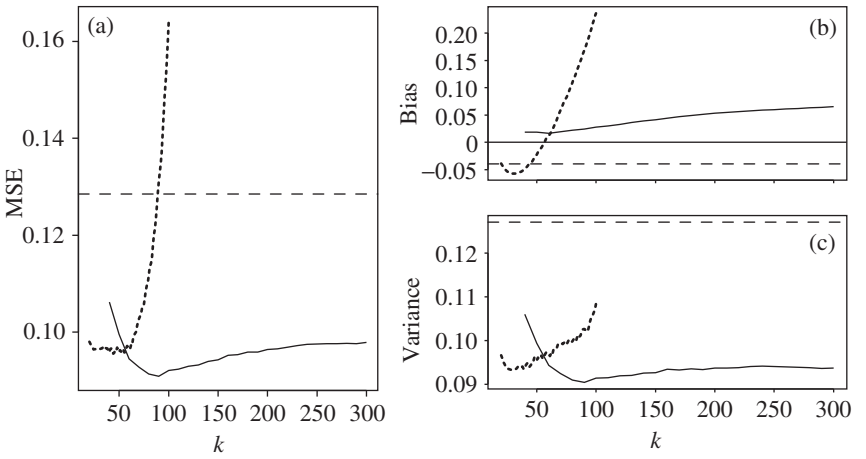

Figure 5.10. Comparison of (a) estimated MSE, (b) bias and (c) variance for the Hill (dotted line) and GPD (solid line) estimators of $\text{VaR}_{0.99}$, as a function of k (or N_u), the number of upper-order statistics from a sample of 1000 t-distributed data with four degrees of freedom. The dashed line also shows results for the (threshold-independent) empirical quantile estimator. See Example 5.28 for details.

The t distribution has a well-behaved regularly varying tail and the Hill estimator gives better estimates of ξ than the GPD method, with an optimal value of k around 20–30. The variance plot shows where the Hill method gains over the GPD method; the variance of the GPD-based estimator is much higher than that of the Hill estimator for small numbers of order statistics. The magnitudes of the biases are closer together, with the Hill method tending to overestimate ξ and the GPD method tending to underestimate it. If we were to use the GPD method, the optimal choice of threshold would be one giving 100–150 exceedances.

The conclusions change when we attempt to estimate the 99% VaR; the results are shown in Figure 5.10. The Hill method has a negative bias for low values of k but a rapidly growing positive bias for larger values of k; the GPD estimator has a positive bias that grows much more slowly; the empirical method has a negative bias. The GPD attains its lowest MSE value for a value of k around 100, but, more importantly, the MSE is very robust to the choice of k because of the slow growth of the bias. The Hill method performs well for $20 \leqslant k \leqslant 75$ (we only use k values that lead to a quantile estimate beyond the effective threshold $X_{k,n}$) but then deteriorates rapidly. Both EVT methods obviously outperform the empirical quantile estimator. Given the relative robustness of the GPD-based tail estimator to changes in k, the issue of threshold choice for this estimator seems less critical than for the Hill method.

5.2.6 Conditional EVT for Financial Time Series

The GPD method when applied to threshold exceedances in a financial return series (as in Examples 5.24 and 5.25) gives us risk-measure estimates for the stationary (or unconditional) distribution of the underlying time series. We now consider a simple adaptation of the GPD method that allows us to obtain risk-measure estimates for the

conditional (or one-step-ahead forecast) distribution of a time series. This adaptation uses the GARCH model and related ideas in Chapter 4 and will be applied in the market-risk context in Chapter 9.

We use the notation developed for prediction or forecasting in Sections 4.1.5 and 4.2.5. Let $X_{t-n+1}, \ldots, X_t$ denote a series of *negative* log-returns and assume that these come from a strictly stationary time-series process (X_t). Assume further that the process satisfies equations of the form $X_t = \mu_t + \sigma_t Z_t$, where μ_t and σ_t are $\mathcal{F}_{t-1}$-measurable and (Z_t) are iid innovations with some unknown df F_Z; an example would be an ARMA model with GARCH errors.

We want to obtain VaR and expected shortfall estimates for the conditional distribution $F_{X_{t+1}|\mathcal{F}_t}$, and in Section 4.2.5 we showed that these risk measures are given by the equations

$$\mathrm{VaR}_\alpha^t = \mu_{t+1} + \sigma_{t+1} q_\alpha(Z), \qquad \mathrm{ES}_\alpha^t = \mu_{t+1} + \sigma_{t+1}\,\mathrm{ES}_\alpha(Z),$$

where we write Z for a generic rv with df F_Z.

These equations suggest an estimation method as follows. We first fit an ARMA–GARCH model by the QML procedure of Section 4.2.4 (since we do not wish to assume a particular innovation distribution) and use this to estimate μ_{t+1} and σ_{t+1}. As an alternative we could use EWMA volatility forecasting. To estimate $q_\alpha(Z)$ and $\mathrm{ES}_\alpha(Z)$ we essentially apply the GPD tail estimation procedure to the innovation distribution F_Z. To get round the problem that we do not observe data directly from the innovation distribution, we treat the residuals from the GARCH analysis as our data and apply the GPD tail estimation method of Section 5.2.3 to the residuals. In particular, we estimate $q_\alpha(Z)$ and $\mathrm{ES}_\alpha(Z)$ using the VaR and expected shortfall formulas in (5.18) and (5.19).

Notes and Comments

The ideas behind the important Theorem 5.20, which underlies GPD modelling, may be found in Pickands (1975) and Balkema and de Haan (1974). Important papers developing the technique in the statistical literature are Davison (1984) and Davison and Smith (1990). The estimation of the parameters of the GPD, both by ML and by the method of probability-weighted moments, is discussed in Hosking and Wallis (1987). The tail estimation formula (5.21) was suggested by Smith (1987), and the theoretical properties of this estimator for iid data in the domain of attraction of an extreme value distribution are extensively investigated in that paper. The Danish fire loss example is taken from McNeil (1997).

The Hill estimator goes back to Hill (1975) (see also Hall 1982). The theoretical properties for dependent data, including linear processes with heavy-tailed innovations and ARCH and GARCH processes, were investigated by Resnick and Stărică (1995, 1996). The idea of smoothing the estimator is examined in Resnick and Stărică (1997) and Resnick (1997). For Hill "horror plots", showing situations when the Hill estimator delivers particularly poor estimates of the tail index, see pp. 194, 270 and 343 of EKM.

Alternative estimators based on order statistics include the estimator of Pickands (1975), which is also discussed in Dekkers and de Haan (1989), and the DEdH estimator of Dekkers, Einmahl and de Haan (1989). This latter estimator is used as the basis of a quantile estimator in de Haan and Rootzén (1993). Both the Pickands and DEdH estimators are designed to estimate general ξ in the extreme value limit (in contrast to the Hill estimator, which is designed for positive ξ); in empirical studies the DEdH estimator seems to work better than the Pickands estimator. The issue of the optimal number of order statistics in such estimators is taken up in a series of papers by Dekkers and de Haan (1993) and Daníelsson et al. (2001a). A method is proposed that is essentially based on the bootstrap approach to estimating mean squared error discussed in Hall (1990). A review paper relevant for applications to insurance and finance is Matthys and Beirlant (2000).

Analyses of the tails of financial data using methods based on the Hill estimator can be found in Koedijk, Schafgans and de Vries (1990), Lux (1996) and various papers by Daníelsson and de Vries (1997a,b,c). The conditional EVT method was developed in McNeil and Frey (2000); a Monte Carlo method using the GPD model to estimate risk measures for the h-day loss distribution is also described. See also Gençay, Selçuk and Ulugülyağci (2003), Gençay and Selçuk (2004) and Chavez-Demoulin, Embrechts and Sardy (2014) for interesting applications of EVT methodology to financial time series and VaR estimation.

5.3 Point Process Models

In our discussion of threshold models in Section 5.2 we considered only the magnitude of excess losses over high thresholds. In this section we consider exceedances of thresholds as events in time and use a point process approach to model the occurrence of these events. We begin by looking at the case of regularly spaced iid data and discuss the well-known POT model for the occurrence of extremes in such data; this model elegantly subsumes the models for maxima and the GPD models for excess losses that we have so far described.

However, the assumptions of the standard POT model are typically violated by financial return series, because of the kind of serial dependence that volatility clustering generates in such data. Our ultimate aim is to find more general point process models to describe the occurrence of extreme values in financial time series, and we find suitable candidates in the class of self-exciting point processes. These models are of a dynamic nature and can be used to estimate conditional VaRs; they offer an interesting alternative to the conditional EVT approach of Section 5.2.6, with the advantage that no pre-whitening of data with GARCH processes is required.

The following section gives an idea of the theory behind the POT model, but it may be skipped by readers who are content to go directly to a description of the standard POT model in Section 5.3.2.

5.3.1 Threshold Exceedances for Strict White Noise

Consider a strict white noise process $(X_i)_{i\in\mathbb{N}}$ representing financial losses. While we discuss the theory for iid variables for simplicity, the results we describe also hold for

dependent processes with extremal index $\theta = 1$, i.e. processes where extreme values show no tendency to cluster (see Section 5.1.3 for examples of such processes).

Throughout this section we assume that the common loss distribution is in the maximum domain of attraction of an extreme value distribution (MDA(H_ξ)) so that (5.1) holds for the non-degenerate limiting distribution H_ξ and normalizing sequences c_n and d_n. From (5.1) it follows, by taking logarithms, that for any fixed x we have

$$\lim_{n\to\infty} n \ln F(c_n x + d_n) = \ln H_\xi(x). \tag{5.25}$$

Throughout this section we also consider a sequence of thresholds $(u_n(x))$ defined by $u_n(x) := c_n x + d_n$ for some fixed value of x. By noting that $-\ln y \sim 1 - y$ as $y \to 1$, we can infer from (5.25) that $n\bar{F}(u_n(x)) \sim -n \ln F(u_n(x)) \to -\ln H_\xi(x)$ as $n \to \infty$ for this sequence of thresholds.

The number of losses in the sample $X_1, \ldots, X_n$ exceeding the threshold $u_n(x)$ is a binomial rv, $N_{u_n(x)} \sim B(n, \bar{F}(u_n(x)))$, with expectation $n\bar{F}(u_n(x))$. Since (5.25) holds, the standard Poisson limit result implies that, as $n \to \infty$, the number of exceedances $N_{u_n(x)}$ converges to a Poisson rv with mean $\lambda(x) = -\ln H_\xi(x)$, depending on the particular value of x chosen.

The theory goes further. Not only is the number of exceedances asymptotically Poisson, these exceedances occur according to a Poisson point process. To state the result it is useful to give a brief summary of some ideas concerning point processes.

On point processes. Suppose we have a sequence of rvs or vectors $Y_1, \ldots, Y_n$ taking values in some *state space* $\mathcal{X}$ (for example, $\mathbb{R}$ or $\mathbb{R}^2$) and we define, for any set $A \subset \mathcal{X}$, the rv

$$N(A) = \sum_{i=1}^{n} I_{\{Y_i \in A\}}, \tag{5.26}$$

which counts the random number of Y_i in the set A. Under some technical conditions (see EKM, pp. 220–223), (5.26) is said to define a point process $N(\cdot)$. An example of a point process is the Poisson point process.

Definition 5.29 (Poisson point process). The point process $N(\cdot)$ is called a *Poisson point process* (or Poisson random measure) on $\mathcal{X}$ with *intensity measure* Λ if the following two conditions are satisfied.

(a) For $A \subset \mathcal{X}$ and $k \geqslant 0$,

$$P(N(A) = k) = \begin{cases} \mathrm{e}^{-\Lambda(A)} \dfrac{\Lambda(A)^k}{k!}, & \Lambda(A) < \infty, \\ 0, & \Lambda(A) = \infty. \end{cases}$$

(b) For any $m \geqslant 1$, if $A_1, \ldots, A_m$ are mutually disjoint subsets of $\mathcal{X}$, then the rvs $N(A_1), \ldots, N(A_m)$ are independent.

The intensity measure $\Lambda(\cdot)$ of $N(\cdot)$ is also known as the *mean measure* because $E(N(A)) = \Lambda(A)$. We also speak of the intensity function (or simply intensity) of the process, which is the derivative $\lambda(\boldsymbol{x})$ of the measure satisfying $\Lambda(A) = \int_A \lambda(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$.

Asymptotic behaviour of the point process of exceedances. Consider again the strict white noise $(X_i)_{i\in\mathbb{N}}$ and the sequence of thresholds $u_n(x) = c_n x + d_n$ for some fixed x. For $n \in \mathbb{N}$ and $1 \leqslant i \leqslant n$ let $Y_{i,n} = (i/n)I_{\{X_i > u_n(x)\}}$ and observe that $Y_{i,n}$ can be thought of as returning either the normalized "time" i/n of an exceedance, or zero. The point process of exceedances of the threshold u_n is the process $N_n(\cdot)$ with state space $\mathcal{X} = (0, 1]$, which is given by

$$N_n(A) = \sum_{i=1}^{n} I_{\{Y_{i,n}\in A\}} \tag{5.27}$$

for $A \subset \mathcal{X}$. As the notation indicates, we consider this process to be an element in a sequence of point processes indexed by n. The point process (5.27) counts the exceedances with time of occurrence in the set A, and we are interested in the behaviour of this process as $n \to \infty$.

It may be shown (see Theorem 5.3.2 in EKM) that $N_n(\cdot)$ converges in distribution on $\mathcal{X}$ to a Poisson point process $N(\cdot)$ with intensity measure $\Lambda(\cdot)$ satisfying $\Lambda(A) = (t_2-t_1)\lambda(x)$ for $A = (t_1, t_2) \subset \mathcal{X}$, where $\lambda(x) = -\ln H_\xi(x)$ as before. This implies, in particular, that $E(N_n(A)) \to E(N(A)) = \Lambda(A) = (t_2 - t_1)\lambda(x)$. Clearly, the intensity does not depend on time and takes the constant value $\lambda := \lambda(x)$; we refer to the limiting process as a *homogeneous Poisson process* with intensity or rate λ.

Application of the result in practice. We give a heuristic argument explaining how this limiting result is used in practice. We consider a fixed large sample size n and a fixed high threshold u, which we assume satisfies $u = c_n y + d_n$ for some value y. We expect that the number of threshold exceedances can be approximated by a Poisson rv and that the point process of exceedances of u can be approximated by a homogeneous Poisson process with rate $\lambda = -\ln H_\xi(y) = -\ln H_\xi((u-d_n)/c_n)$. If we replace the normalizing constants c_n and d_n by $\sigma > 0$ and μ, we have a Poisson process with rate $-\ln H_{\xi,\mu,\sigma}(u)$. Clearly, we could repeat the same argument with any high threshold so that, for example, we would expect it to be approximately true that exceedances of the level $x \geqslant u$ occur according to a Poisson process with rate $-\ln H_{\xi,\mu,\sigma}(x)$.

We therefore have an intimate relationship between the GEV model for block maxima and a Poisson model for the occurrence in time of exceedances of a high threshold. The arguments of this section therefore provide theoretical support for the observation in Figure 3.3: that exceedances for simulated iid t data are separated by waiting times that behave like iid exponential observations.

5.3.2 The POT Model

The theory of the previous section combined with the theory of Section 5.2 suggests an asymptotic model for threshold exceedances in regularly spaced iid data (or data from a process with extremal index $\theta = 1$). The so-called POT model makes the following assumptions.

- Exceedances occur according to a homogeneous Poisson process in time.
- Excess amounts above the threshold are iid and independent of exceedance times.
- The distribution of excess amounts is generalized Pareto.

There are various alternative ways of describing this model. It might also be called a *marked Poisson* point process, where the exceedance times constitute the points and the GPD-distributed excesses are the marks. It can also be described as a (non-homogeneous) *two-dimensional Poisson* point process, where points (t, x) in two-dimensional space record times and magnitudes of exceedances. The latter representation is particularly powerful, as we now discuss.

Two-dimensional Poisson formulation of POT model. Assume that we have regularly spaced random losses $X_1, \ldots, X_n$ and that we set a high threshold u. We *assume* that, on the state space $\mathcal{X} = (0, 1] \times (u, \infty)$, the point process defined by $N(A) = \sum_{i=1}^{n} I_{\{(i/n, X_i) \in A\}}$ is a Poisson process with intensity at a point (t, x) given by

$$\lambda(t, x) = \frac{1}{\sigma}\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi - 1} \tag{5.28}$$

provided $(1+\xi(x-\mu)/\sigma) > 0$, and by $\lambda(t, x) = 0$ otherwise. Note that this intensity does not depend on t but does depend on x, and hence the two-dimensional Poisson process is non-homogeneous; we simplify the notation to $\lambda(x) := \lambda(t, x)$. For a set of the form $A = (t_1, t_2) \times (x, \infty) \subset \mathcal{X}$, the intensity measure is

$$\Lambda(A) = \int_{t_1}^{t_2} \int_{x}^{\infty} \lambda(y)\, \mathrm{d}y\, \mathrm{d}t = -(t_2 - t_1) \ln H_{\xi,\mu,\sigma}(x). \tag{5.29}$$

It follows from (5.29) that for any $x \geqslant u$, the implied one-dimensional process of exceedances of the level x is a homogeneous Poisson process with rate $\tau(x) := -\ln H_{\xi,\mu,\sigma}(x)$. Now consider the excess amounts over the threshold u. The tail of the excess df over the threshold u, denoted by $\bar{F}_u(x)$ before, can be calculated as the ratio of the rates of exceeding the levels $u + x$ and u. We obtain

$$\bar{F}_u(x) = \frac{\tau(u + x)}{\tau(u)} = \left(1 + \frac{\xi x}{\sigma + \xi(u - \mu)}\right)^{-1/\xi} = \bar{G}_{\xi,\beta}(x)$$

for a positive scaling parameter $\beta = \sigma + \xi(u - \mu)$. This is precisely the tail of the GPD model for excesses over the threshold u used in Section 5.2.2. Thus this seemingly complicated model is indeed the POT model described informally at the beginning of this section.

Note also that the model implies the GEV distributional model for maxima. To see this, consider the event that $\{M_n \leqslant x\}$ for some value $x \geqslant u$. This may be expressed in point process language as the event that there are no points in the set $A = (0, 1] \times (x, \infty)$. The probability of this event is calculated to be $P(M_n \leqslant x) = P(N(A) = 0) = \mathrm{e}^{-\Lambda(A)} = H_{\xi,\mu,\sigma}(x)$, $x \geqslant u$, which is precisely the GEV model for maxima of n-blocks used in Section 5.1.4.

Statistical estimation of the POT model. The most elegant way of fitting the POT model to data is to fit the point process with intensity (5.28) to the exceedance data in one step. Given the exceedance data $\{\tilde{X}_j : j = 1, \ldots, N_u\}$, the likelihood can be written as

$$L(\xi, \sigma, \mu; \tilde{X}_1, \ldots, \tilde{X}_{N_u}) = \mathrm{e}^{-\tau(u)} \prod_{j=1}^{N_u} \lambda(\tilde{X}_j). \tag{5.30}$$

Parameter estimates of ξ, σ and μ are obtained by maximizing this expression, which is easily accomplished by numerical means. For literature on the derivation of this likelihood, see Notes and Comments.

There are, however, simpler ways of getting the same parameter estimates. Suppose we reparametrize the POT model in terms of $\tau := \tau(u) = -\ln H_{\xi,\mu,\sigma}(u)$, the rate of the one-dimensional Poisson process of exceedances of the level u, and $\beta = \sigma + \xi(u - \mu)$, the scaling parameter of the implied GPD for the excess losses over u. The intensity in (5.28) can then be rewritten as

$$\lambda(x) = \lambda(t, x) = \frac{\tau}{\beta}\left(1 + \xi\frac{x - u}{\beta}\right)^{-1/\xi - 1}, \tag{5.31}$$

where $\xi \in \mathbb{R}$ and $\tau, \beta > 0$. Using this parametrization it is easily verified that the log of the likelihood in (5.30) becomes

$$\ln L(\xi, \sigma, \mu; \tilde{X}_1, \ldots, \tilde{X}_{N_u}) = \ln L_1(\xi, \beta; \tilde{X}_1 - u, \ldots, \tilde{X}_{N_u} - u) + \ln L_2(\tau; N_u),$$

where L_1 is precisely the likelihood for fitting a GPD to excess losses given in (5.14), and $\ln L_2(\tau; N_u) = -\tau + N_u \ln \tau$, which is the log-likelihood for a one-dimensional homogeneous Poisson process with rate τ. Such a partition of a log-likelihood into a sum of two terms involving two different sets of parameters means that we can make separate inferences about the two sets of parameters; we can estimate ξ and β in a GPD analysis and then estimate τ by its MLE N_u and use these to infer estimates of μ and σ.

Advantages of the POT model formulation. One might ask what the advantages of approaching the modelling of extremes through the two-dimensional Poisson point process model described by the intensity (5.28) could be? One advantage is the fact that the parameters ξ, μ and σ in the Poisson point process model do not have any theoretical dependence on the threshold chosen, unlike the parameter β in the GPD model, which appears in the theory as a function of the threshold u. In practice, we would expect the estimated parameters of the Poisson model to be roughly stable over a range of high thresholds, whereas the estimated β parameter varies with threshold choice.

For this reason the intensity (5.28) is a framework that is often used to introduce covariate effects into extreme value modelling. One method of doing this is to replace the parameters μ and σ in (5.28) by parameters that vary over time as a function of deterministic covariates. For example, we might have $\mu(t) = \alpha + \boldsymbol{\gamma}'\boldsymbol{y}(t)$, where $\boldsymbol{y}(t)$ represents a vector of covariate values at time t. This would give us Poisson processes that are also non-homogeneous in time.

Applicability of the POT model to return series data. We now turn to the use of the POT model with financial return data. An initial comment is that returns do not really form genuine point events in time, in contrast to recorded water levels or wind speeds, for example. Returns are discrete-time measurements that describe changes in value taking place over the course of, say, a day or a week. Nonetheless, we assume that if we take a longer-term perspective, such data can be approximated by point events in time.

In Section 3.1 and in Figure 3.3 in particular, we saw evidence that, in contrast to iid data, exceedances of a high threshold for daily financial return series do not necessarily occur according to a homogeneous Poisson process. They tend instead to form clusters corresponding to episodes of high volatility. Thus the standard POT model is not directly applicable to financial return data.

Theory suggests that for stochastic processes with extremal index $\theta < 1$, such as GARCH processes, the extremal clusters themselves should occur according to a homogeneous Poisson process in time, so that the individual exceedances occur according to a *Poisson cluster process* (see, for example, Leadbetter 1991). A suitable model for the occurrence and magnitude of exceedances in a financial return series might therefore be some form of marked Poisson cluster process.

Rather than attempting to specify the mechanics of cluster formation, it is quite common to try to circumvent the problem by *declustering* financial return data: we attempt to formally identify clusters of exceedances and then we apply the POT model to cluster maxima only. This method is obviously somewhat ad hoc, as there is usually no clear way of deciding where one cluster ends and another begins. A possible declustering algorithm is given by the *runs method*. In this method a run size r is fixed and two successive exceedances are said to belong to two different clusters if they are separated by a run of at least r values below the threshold (see EKM, pp. 422–424). In Figure 5.11 the DAX daily negative returns of Figure 3.3 have been declustered with a run length of ten trading days; this reduces the 100 exceedances to 42 cluster maxima.

However, it is not clear that applying the POT model to declustered data gives us a particularly useful model. We can estimate the rate of occurrence of clusters of extremes and say something about average cluster size; we can also derive a GPD model for excess losses over thresholds for cluster maxima (where standard errors for parameters may be more realistic than if we fitted the GPD to the dependent sample of all threshold exceedances). However, by neglecting the modelling of cluster formation, we cannot make more dynamic statements about the intensity of occurrence of threshold exceedances at any point in time. In Section 16.2 we describe self-exciting point process models, which do attempt to model the dynamics of cluster formation.

Example 5.30 (POT analysis of AT&T weekly losses). We close this section with an example of a standard POT model applied to extremes in financial return data. To mitigate the clustering phenomenon discussed above we use weekly return data, as previously analysed in Examples 5.24 and 5.25. Recall that these yield 102 weekly percentage losses for the AT&T stock price exceeding a threshold of 2.75%. The

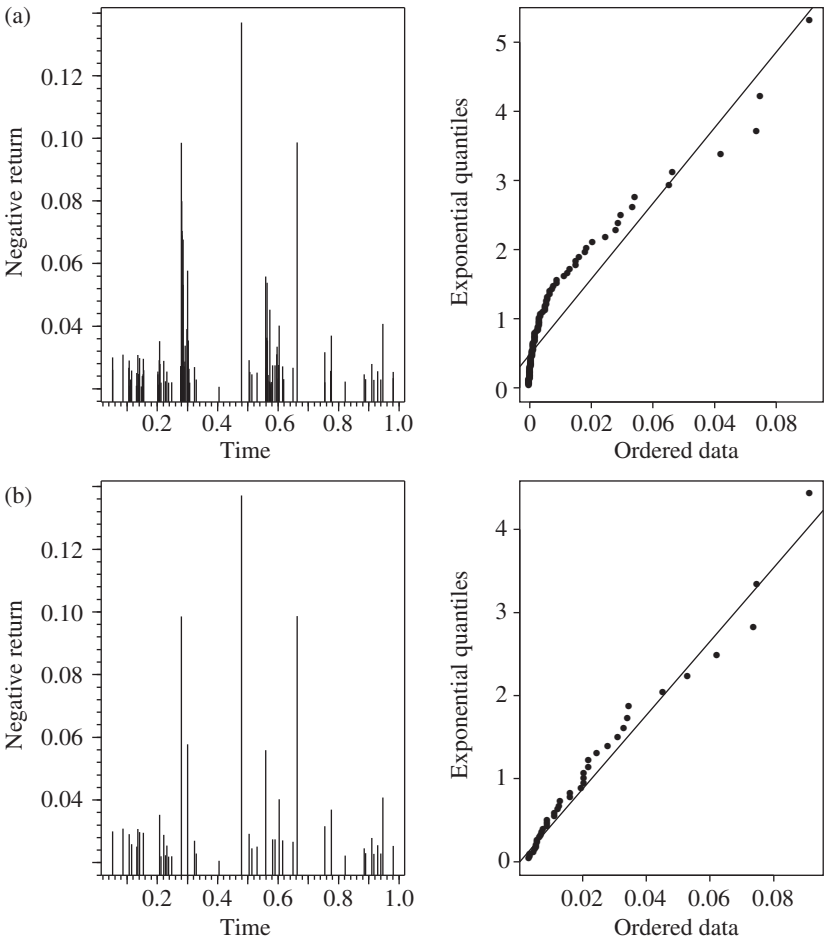

Figure 5.11. (a) DAX daily negative returns and a Q–Q plot of their spacings as in Figure 3.3. (b) Data have been declustered with the runs method using a run length of ten trading days. The spacings of the 42 cluster maxima are more consistent with a Poisson model.

data are shown in Figure 5.12, where we observe that the inter-exceedance times seem to have a roughly exponential distribution, although the discrete nature of the times and the relatively low value of n means that there are some tied values for the spacings, which makes the plot look a little granular. Another noticeable feature is that the exceedances of the threshold appear to become more frequent over time, which might be taken as evidence against the homogeneous Poisson assumption for threshold exceedances and against the implicit assumption that the underlying data form a realization from a stationary time series. It would be possible to consider a POT model incorporating a trend of increasingly frequent exceedances, but we will not go this far.

We fit the standard two-dimensional Poisson model to the 102 exceedances of the threshold 2.75% using the likelihood in (5.30) and obtain parameter estimates

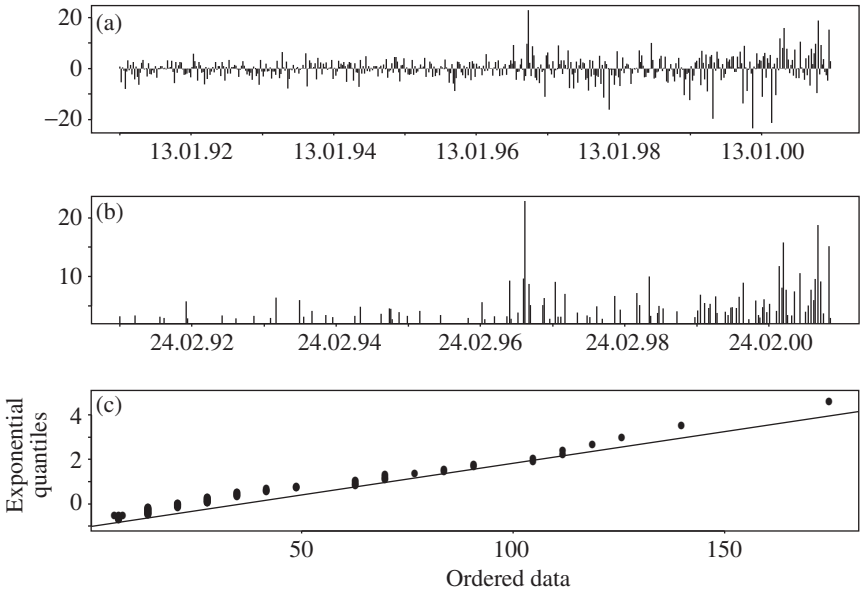

Figure 5.12. (a) Time series of AT&T weekly percentage losses from 1991 to 2000. (b) Corresponding realization of the marked point process of exceedances of the threshold 2.75%. (c) Q–Q plot of inter-exceedance times against an exponential reference distribution. See Example 5.30 for details.

$\hat{\xi} = 0.22$, $\hat{\mu} = 19.9$ and $\hat{\sigma} = 5.95$. The implied GPD scale parameter for the distribution of excess losses over the threshold u is $\hat{\beta} = \hat{\sigma} + \hat{\xi}(u - \hat{\mu}) = 2.1$, so we have exactly the same estimates of ξ and β as in Example 5.24.

The estimated exceedance rate for the threshold $u = 2.75$ is given by $\hat{\tau}(u) = -\ln H_{\hat{\xi},\hat{\mu},\hat{\sigma}}(u) = 102$, which is precisely the number of exceedances of that threshold, as theory suggests. It is of more interest to look at estimated exceedance rates for higher thresholds. For example, we get $\hat{\tau}(15) = 2.50$, which implies that losses exceeding 15% occur as a Poisson process with rate 2.5 losses per ten-year period, so that such a loss is, roughly speaking, a four-year event. Thus the Poisson model gives us an alternative method of defining the return period of a stress event and a more powerful way of calculating such a risk measure. Similarly, we can invert the problem to estimate return levels: suppose we define the ten-year return level as that level which is exceeded according to a Poisson process with rate one loss per ten years, then we can easily estimate the level in our model by calculating

$$H^{-1}_{\hat{\xi},\hat{\mu},\hat{\sigma}}(\mathrm{e}^{-1}) = 19.9,$$

so the ten-year event is a weekly loss of roughly 20%. Using the profile likelihood method in Section A.3.5 we could also give confidence intervals for such estimates.

Notes and Comments

For more information about point processes consult EKM, Cox and Isham (1980), Kallenberg (1983) and Resnick (2008). The point process approach to extremes

dates back to Pickands (1971) and is also discussed in Leadbetter, Lindgren and Rootzén (1983), Leadbetter (1991) and Falk, Hüsler and Reiss (1994).

The two-dimensional Poisson point process model was first used in practice by Smith (1989) and may also be found in Smith and Shively (1995); both these papers discuss the adaptation of the point process model to incorporate covariates or time trends in the context of environmental data. An insurance application is treated in Smith and Goodman (2000), which also treats the point process model from a Bayesian perspective. An interesting application to wind storm losses is Rootzén and Tajvidi (1997). A further application of the bivariate point process framework to the modeling of insurance loss data, showing trends in both intensity and severity of occurrence, is found in McNeil and Saladin (2000). For further applications to insurance and finance, see Chavez-Demoulin and Embrechts (2004) and Chavez-Demoulin, Embrechts and Hofert (2014). An excellent overview of statistical approaches to the GPD and point process models is found in Coles (2001).

The derivation of likelihoods for point process models is beyond the scope of this book and we have simply recorded the likelihoods to be maximized without further justification. See Daley and Vere-Jones (2003, Chapter 7) for more details on this subject; see also Coles (2001, p. 127) for a good intuitive account in the Poisson case.

6
Multivariate Models

Financial risk models, whether for market or credit risks, are inherently multivariate. The value change of a portfolio of traded instruments over a fixed time horizon depends on a random vector of risk-factor changes or returns. The loss incurred by a credit portfolio depends on a random vector of losses for the individual counterparties in the portfolio.

This chapter is the first of two successive ones that focus on models for random vectors. The emphasis in this chapter is on tractable models that describe both the individual behaviour of the components of a random vector and their joint behaviour or dependence structure. We consider a number of distributions that extend the multivariate normal but provide more realistic models for many kinds of financial data.

In Chapter 7 we focus explicitly on modelling the dependence structure of a random vector and largely ignore marginal behaviour. We introduce copula models of dependence and study a number of dependence measures and concepts related to copulas. Both Chapter 6 and Chapter 7 take a static, distributional view of multivariate modelling; for multivariate time-series models, see Chapter 14.

Section 6.1 reviews basic ideas in multivariate statistics and discusses the multivariate normal (or Gaussian) distribution and its deficiencies as a model for empirical return data.

In Section 6.2 we consider a generalization of the multivariate normal distribution known as a multivariate normal mixture distribution, which shares much of the structure of the multivariate normal and retains many of its properties. We treat both variance mixtures, which belong to the wider class of elliptical distributions, and mean–variance mixtures, which allow asymmetry. Concrete examples include t distributions and generalized hyperbolic distributions, and we show in empirical examples that these models provide a better fit than a Gaussian distribution to asset return data. In some cases, multivariate return data are not strongly asymmetric and models from the class of elliptical distributions are good enough; in Section 6.3 we investigate the elegant properties of these distributions.

In Section 6.4 we discuss the important issue of dimension-reduction techniques for reducing large sets of risk factors to smaller subsets of essential risk drivers. The key idea here is that of a factor model, and we also review the principal components method of constructing factors.

6.1 Basics of Multivariate Modelling

This first section reviews important basic material from multivariate statistics, which will be known to many readers. The main topic of the section is the multivariate normal distribution and its properties; this distribution is central to much of classical multivariate analysis and was the starting point for attempts to model market risk (the variance–covariance method of Section 9.2.2).

6.1.1 Random Vectors and Their Distributions

Joint and marginal distributions. Consider a general d-dimensional random vector of risk-factor changes (or so-called returns) $\boldsymbol{X} = (X_1, \ldots, X_d)'$. The distribution of $\boldsymbol{X}$ is completely described by the joint distribution function (df)

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = F_{\boldsymbol{X}}(x_1, \ldots, x_d) = P(\boldsymbol{X} \leqslant \boldsymbol{x}) = P(X_1 \leqslant x_1, \ldots, X_d \leqslant x_d).$$

Where no ambiguity arises we simply write F, omitting the subscript.

The *marginal* df of X_i, written F_{X_i} or often simply F_i, is the df of that risk factor considered individually and is easily calculated from the joint df. For all i we have

$$F_i(x_i) = P(X_i \leqslant x_i) = F(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty). \tag{6.1}$$

If the marginal df $F_i(x)$ is absolutely continuous, then we refer to its derivative $f_i(x)$ as the marginal density of X_i. It is also possible to define k-dimensional marginal distributions of $\boldsymbol{X}$ for $2 \leqslant k \leqslant d-1$. Suppose we partition $\boldsymbol{X}$ into $(\boldsymbol{X}_1', \boldsymbol{X}_2')'$, where $\boldsymbol{X}_1 = (X_1, \ldots, X_k)'$ and $\boldsymbol{X}_2 = (X_{k+1}, \ldots, X_d)'$, then the marginal df of $\boldsymbol{X}_1$ is

$$F_{\boldsymbol{X}_1}(\boldsymbol{x}_1) = P(\boldsymbol{X}_1 \leqslant \boldsymbol{x}_1) = F(x_1, \ldots, x_k, \infty, \ldots, \infty).$$

For bivariate and other low-dimensional margins it is convenient to have a simpler alternative notation in which, for example, $F_{ij}(x_i, x_j)$ stands for the marginal distribution of the components X_i and X_j.

The df of a random vector $\boldsymbol{X}$ is said to be absolutely continuous if

$$F(x_1, \ldots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(u_1, \ldots, u_d)\, \mathrm{d}u_1 \cdots \mathrm{d}u_d$$

for some non-negative function f, which is then known as the *joint density* of $\boldsymbol{X}$. Note that the existence of a joint density implies the existence of marginal densities for all k-dimensional marginals. However, the existence of a joint density is not necessarily implied by the existence of marginal densities (counterexamples can be found in Chapter 7 on copulas).

In some situations it is convenient to work with the *survival function* of $\boldsymbol{X}$, defined by

$$\bar{F}_{\boldsymbol{X}}(\boldsymbol{x}) = \bar{F}_{\boldsymbol{X}}(x_1, \ldots, x_d) = P(\boldsymbol{X} > \boldsymbol{x}) = P(X_1 > x_1, \ldots, X_d > x_d)$$

and written simply as $\bar{F}$ when no ambiguity arises. The marginal survival function of X_i, written $\bar{F}_{X_i}$ or often simply $\bar{F}_i$, is given by

$$\bar{F}_i(x_i) = P(X_i > x_i) = \bar{F}(-\infty, \ldots, -\infty, x_i, -\infty, \ldots, -\infty).$$

Conditional distributions and independence. If we have a multivariate model for risks in the form of a joint df, survival function or density, then we have implicitly described the *dependence structure* of the risks. We can make conditional probability statements about the probability that certain components take certain values given that other components take other values. For example, consider again our partition of $\boldsymbol{X}$ into $(\boldsymbol{X}_1', \boldsymbol{X}_2')'$ and assume absolute continuity of the df of $\boldsymbol{X}$. Let $f_{\boldsymbol{X}_1}$ denote the joint density of the k-dimensional marginal distribution $F_{\boldsymbol{X}_1}$. Then the conditional distribution of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1 = \boldsymbol{x}_1$ has density

$$f_{\boldsymbol{X}_2|\boldsymbol{X}_1}(\boldsymbol{x}_2 \mid \boldsymbol{x}_1) = \frac{f(\boldsymbol{x}_1, \boldsymbol{x}_2)}{f_{\boldsymbol{X}_1}(\boldsymbol{x}_1)}, \tag{6.2}$$

and the corresponding df is

$$F_{\boldsymbol{X}_2|\boldsymbol{X}_1}(\boldsymbol{x}_2 \mid \boldsymbol{x}_1) = \int_{u_{k+1}=-\infty}^{x_{k+1}} \cdots \int_{u_d=-\infty}^{x_d} \frac{f(x_1, \ldots, x_k, u_{k+1}, \ldots, u_d)}{f_{\boldsymbol{X}_1}(\boldsymbol{x}_1)} \, \mathrm{d}u_{k+1} \cdots \mathrm{d}u_d.$$

If the joint density of $\boldsymbol{X}$ factorizes into $f(\boldsymbol{x}) = f_{\boldsymbol{X}_1}(\boldsymbol{x}_1) f_{\boldsymbol{X}_2}(\boldsymbol{x}_2)$, then the conditional distribution and density of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1 = \boldsymbol{x}_1$ are identical to the marginal distribution and density of $\boldsymbol{X}_2$: in other words, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent. We recall that $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent if and only if

$$F(\boldsymbol{x}) = F_{\boldsymbol{X}_1}(\boldsymbol{x}_1) F_{\boldsymbol{X}_2}(\boldsymbol{x}_2), \quad \forall \boldsymbol{x},$$

or, in the case where $\boldsymbol{X}$ possesses a joint density, $f(\boldsymbol{x}) = f_{\boldsymbol{X}_1}(\boldsymbol{x}_1) f_{\boldsymbol{X}_2}(\boldsymbol{x}_2)$.

The components of $\boldsymbol{X}$ are *mutually* independent if and only if $F(\boldsymbol{x}) = \prod_{i=1}^d F_i(x_i)$ for all $\boldsymbol{x} \in \mathbb{R}^d$ or, in the case where $\boldsymbol{X}$ possesses a density, $f(\boldsymbol{x}) = \prod_{i=1}^d f_i(x_i)$.

Moments and characteristic function. The *mean vector* of $\boldsymbol{X}$, when it exists, is given by

$$E(\boldsymbol{X}) := (E(X_1), \ldots, E(X_d))'.$$

The *covariance matrix*, when it exists, is the matrix $\mathrm{cov}(\boldsymbol{X})$ defined by

$$\mathrm{cov}(\boldsymbol{X}) := E((\boldsymbol{X} - E(\boldsymbol{X}))(\boldsymbol{X} - E(\boldsymbol{X}))'),$$

where the expectation operator acts componentwise on matrices. If we write Σ for $\mathrm{cov}(\boldsymbol{X})$, then the (i, j)th element of this matrix is

$$\sigma_{ij} = \mathrm{cov}(X_i, X_j) = E(X_i X_j) - E(X_i) E(X_j),$$

the ordinary pairwise covariance between X_i and X_j. The diagonal elements $\sigma_{11}, \ldots, \sigma_{dd}$ are the variances of the components of $\boldsymbol{X}$.

The *correlation matrix* of $\boldsymbol{X}$, denoted by $\rho(\boldsymbol{X})$, can be defined by introducing a standardized vector $\boldsymbol{Y}$ such that $Y_i = X_i / \sqrt{\mathrm{var}(X_i)}$ for all i and taking $\rho(\boldsymbol{X}) := \mathrm{cov}(\boldsymbol{Y})$. If we write P for $\rho(\boldsymbol{X})$, then the (i, j)th element of this matrix is

$$\rho_{ij} = \rho(X_i, X_j) = \frac{\mathrm{cov}(X_i, X_j)}{\sqrt{\mathrm{var}(X_i)\,\mathrm{var}(X_j)}}, \tag{6.3}$$

the ordinary pairwise linear correlation of X_i and X_j. To express the relationship between correlation and covariance matrices in matrix form, it is useful to introduce operators on a covariance matrix Σ as follows:

$$\Delta(\Sigma) := \operatorname{diag}(\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{dd}}), \tag{6.4}$$

$$\wp(\Sigma) := (\Delta(\Sigma))^{-1}\Sigma(\Delta(\Sigma))^{-1}. \tag{6.5}$$

Thus $\Delta(\Sigma)$ extracts from Σ a diagonal matrix of standard deviations, and $\wp(\Sigma)$ extracts a correlation matrix. The covariance and correlation matrices Σ and P of $\boldsymbol{X}$ are related by

$$P = \wp(\Sigma). \tag{6.6}$$

Mean vectors and covariance matrices are manipulated extremely easily under linear operations on the vector $\boldsymbol{X}$. For any matrix $B \in \mathbb{R}^{k\times d}$ and vector $\boldsymbol{b} \in \mathbb{R}^k$ we have

$$E(B\boldsymbol{X} + \boldsymbol{b}) = BE(\boldsymbol{X}) + \boldsymbol{b}, \tag{6.7}$$

$$\operatorname{cov}(B\boldsymbol{X} + \boldsymbol{b}) = B\operatorname{cov}(\boldsymbol{X})B'. \tag{6.8}$$

Covariance matrices (and hence correlation matrices) are therefore *positive semi-definite*; writing Σ for $\operatorname{cov}(\boldsymbol{X})$ we see that (6.8) implies that $\operatorname{var}(\boldsymbol{a}'\boldsymbol{X}) = \boldsymbol{a}'\Sigma\boldsymbol{a} \geqslant 0$ for any $\boldsymbol{a} \in \mathbb{R}^d$. If we have that $\boldsymbol{a}'\Sigma\boldsymbol{a} > 0$ for any $\boldsymbol{a} \in \mathbb{R}^d \setminus \{\boldsymbol{0}\}$, we say that Σ is *positive definite*; in this case the matrix is invertible. We will make use of the well-known *Cholesky factorization* of positive-definite covariance matrices at many points; it is well known that such a matrix can be written as $\Sigma = AA'$ for a lower triangular matrix A with positive diagonal elements. The matrix A is known as the Cholesky factor. It will be convenient to denote this factor by $\Sigma^{1/2}$ and its inverse by $\Sigma^{-1/2}$. Note that there are other ways of defining the "square root" of a symmetric positive-definite matrix (such as the symmetric decomposition), but we will always use $\Sigma^{1/2}$ to denote the Cholesky factor.

In this chapter many properties of the multivariate distribution of a vector $\boldsymbol{X}$ are demonstrated using the characteristic function, which is given by

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = E(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{X}}) = E(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{X}}), \quad \boldsymbol{t} \in \mathbb{R}^d.$$

6.1.2 Standard Estimators of Covariance and Correlation

Suppose we have n observations of a d-dimensional risk-factor return vector denoted by $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. Typically, these would be daily, weekly, monthly or yearly observations forming a multivariate time series. We will assume throughout this chapter that the observations are *identically distributed* in the window of observation and either independent or at least serially *uncorrelated* (also known as multivariate white noise). As discussed in Chapter 3, the assumption of independence may be roughly tenable for longer time intervals such as months or years. For shorter time intervals independence may be a less appropriate assumption (due to a phenomenon known as volatility clustering, discussed in Section 3.1.1), but serial correlation of returns is often quite weak.

We assume that the observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ come from a distribution with mean vector $\boldsymbol{\mu}$, finite covariance matrix Σ and correlation matrix P. We now briefly review the standard estimators of these vector and matrix parameters.

Standard method-of-moments estimators of $\boldsymbol{\mu}$ and Σ are given by the *sample mean vector* $\bar{\boldsymbol{X}}$ and the *sample covariance matrix* S. These are defined by

$$\bar{\boldsymbol{X}} := \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i, \qquad S := \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})', \tag{6.9}$$

where arithmetic operations on vectors and matrices are performed componentwise. $\bar{\boldsymbol{X}}$ is an unbiased estimator but S is biased; an unbiased version may be obtained by taking $S_u := nS/(n-1)$, as may be seen by calculating

$$\begin{aligned} nE(S) &= E\bigg(\sum_{i=1}^{n} (\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_i - \boldsymbol{\mu})' - n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\bigg) \\ &= \sum_{i=1}^{n} \operatorname{cov}(\boldsymbol{X}_i) - n\operatorname{cov}(\bar{\boldsymbol{X}}) = n\Sigma - \Sigma, \end{aligned}$$

since $\operatorname{cov}(\bar{\boldsymbol{X}}) = n^{-1}\Sigma$ when the data vectors are iid, or identically distributed and uncorrelated.

The *sample correlation matrix* R may be easily calculated from the sample covariance matrix; its (j,k)th element is given by $r_{jk} = s_{jk}/\sqrt{s_{jj}s_{kk}}$, where s_{jk} denotes the (j,k)th element of S. Or, using the notation introduced in (6.5), we have

$$R = \wp(S),$$

which is the analogous equation to (6.6) for estimators.

Further properties of the estimators $\bar{\boldsymbol{X}}$, S and R will very much depend on the *true multivariate distribution* of the observations. These quantities are not necessarily the best estimators of the corresponding theoretical quantities in all situations. This point is often forgotten in financial risk management, where sample covariance and correlation matrices are routinely calculated and interpreted with little critical consideration of underlying models.

If our data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are iid multivariate normal, then $\bar{\boldsymbol{X}}$ and S are the *maximum likelihood estimators* (MLEs) of the mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. Their behaviour as estimators is well understood, and statistical inference for the model parameters is described in all standard texts on multivariate analysis.

However, the multivariate normal is certainly not a good description of financial risk-factor returns over short time intervals, such as daily data, and is often not good over longer time intervals either. Under these circumstances the behaviour of the standard estimators in (6.9) is often less well understood, and other estimators of the true mean vector $\boldsymbol{\mu}$ and covariance matrix Σ may perform better in terms of *efficiency* and *robustness*. Roughly speaking, by a more efficient estimator we mean an estimator with a smaller expected estimation error; by a more robust estimator we mean an estimator whose performance is not so susceptible to the presence of outlying data values.

6.1.3 The Multivariate Normal Distribution

Definition 6.1. $\boldsymbol{X} = (X_1, \ldots, X_d)'$ has a multivariate normal or *Gaussian* distribution if

$$\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + A\boldsymbol{Z},$$

where $\boldsymbol{Z} = (Z_1, \ldots, Z_k)'$ is a vector of iid univariate *standard* normal rvs (mean 0 and variance 1), and $A \in \mathbb{R}^{d\times k}$ and $\boldsymbol{\mu} \in \mathbb{R}^d$ are a matrix and a vector of constants, respectively.

It is easy to verify, using (6.7) and (6.8), that the mean vector of this distribution is $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and the covariance matrix is $\mathrm{cov}(\boldsymbol{X}) = \Sigma$, where $\Sigma = AA'$ is a positive-semidefinite matrix. Moreover, using the fact that the characteristic function of a standard univariate normal variate Z is $\phi_Z(t) = \mathrm{e}^{-t^2/2}$, the characteristic function of $\boldsymbol{X}$ may be calculated to be

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = E(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{X}}) = \exp(\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu} - \tfrac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}), \quad \boldsymbol{t} \in \mathbb{R}^d. \tag{6.10}$$

Clearly, the distribution is characterized by its mean vector and covariance matrix, and hence a standard notation is $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$. Note that the components of $\boldsymbol{X}$ are mutually independent if and only if Σ is diagonal. For example, $\boldsymbol{X} \sim N_d(\boldsymbol{0}, I_d)$ if and only if $X_1, \ldots, X_d$ are iid $N(0, 1)$, the standard univariate normal distribution.

We concentrate on the *non-singular case* of the multivariate normal when $\mathrm{rank}(A) = d \leqslant k$. In this case the covariance matrix Σ has full rank d and is therefore invertible (non-singular) and positive definite. Moreover, $\boldsymbol{X}$ has an absolutely continuous distribution function with joint density given by

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\{-\tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\}, \quad \boldsymbol{x} \in \mathbb{R}^d, \tag{6.11}$$

where $|\Sigma|$ denotes the determinant of Σ.

The form of the density clearly shows that points with equal density lie on *ellipsoids* determined by equations of the form $(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) = c$, for constants $c > 0$. In two dimensions the contours of equal density are ellipses, as illustrated in Figure 6.1. Whenever a multivariate density $f(\boldsymbol{x})$ depends on $\boldsymbol{x}$ only through the quadratic form $(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$, it is the density of a so-called elliptical distribution, as discussed in more detail in Section 6.3.

Definition 6.1 is essentially a simulation recipe for the multivariate normal distribution. To be explicit, if we wished to generate a vector $\boldsymbol{X}$ with distribution $N_d(\boldsymbol{\mu}, \Sigma)$, where Σ is positive definite, we would use the following algorithm.

Algorithm 6.2 (simulation of multivariate normal distribution).

(1) Perform a Cholesky decomposition of Σ (see, for example, Press et al. 1992) to obtain the Cholesky factor $\Sigma^{1/2}$.

(2) Generate a vector $\boldsymbol{Z} = (Z_1, \ldots, Z_d)'$ of independent standard normal variates.

(3) Set $\boldsymbol{X} = \boldsymbol{\mu} + \Sigma^{1/2}\boldsymbol{Z}$.

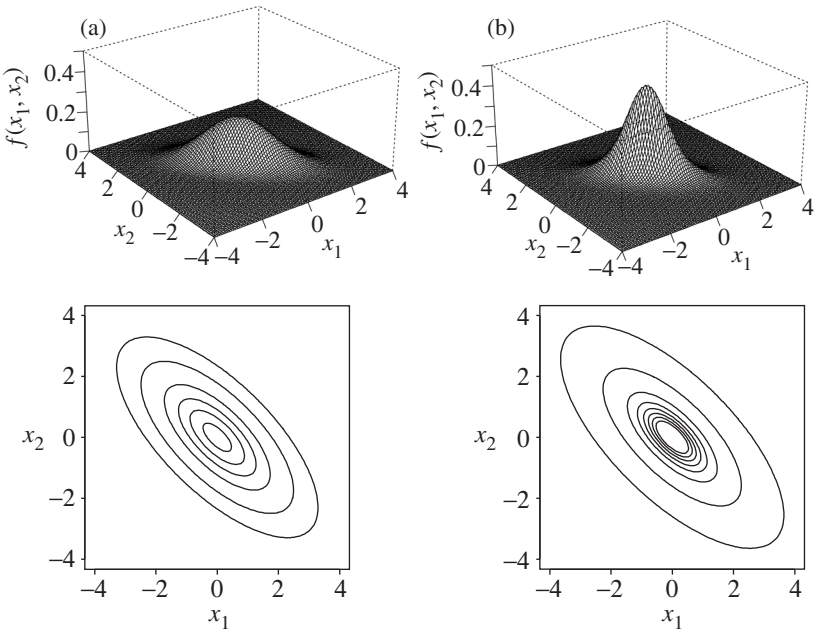

Figure 6.1. (a) Perspective and contour plots for the density of a bivariate normal distribution with standard normal margins and correlation −70%. (b) Corresponding plots for a bivariate t density with four degrees of freedom (see Example 6.7 for details) and the *same mean vector and covariance matrix* as the normal distribution. Contour lines are plotted at the same heights for both densities.

We now summarize further useful properties of the multivariate normal. These properties underline the attractiveness of the multivariate normal for computational work in risk management. Note, however, that many of them are in fact shared by the broader classes of normal mixture distributions and elliptical distributions (see Section 6.3.3 for properties of the latter).

Linear combinations. If we take linear combinations of multivariate normal random vectors, then these remain multivariate normal. Let $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$ and take any $B \in \mathbb{R}^{k \times d}$ and $\boldsymbol{b} \in \mathbb{R}^k$. Then it is easily shown (e.g. using the characteristic function (6.10)) that

$$B\boldsymbol{X} + \boldsymbol{b} \sim N_k(B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B'). \tag{6.12}$$

As a special case, if $\boldsymbol{a} \in \mathbb{R}^d$, then

$$\boldsymbol{a}'\boldsymbol{X} \sim N(\boldsymbol{a}'\boldsymbol{\mu}, \boldsymbol{a}'\Sigma\boldsymbol{a}), \tag{6.13}$$

and this fact is used routinely in the variance–covariance approach to risk management, as discussed in Section 9.2.2.

In this context it is interesting to note the following elegant characterization of multivariate normality. It is easily shown using characteristic functions that $\boldsymbol{X}$ is multivariate normal *if and only if* $\boldsymbol{a}'\boldsymbol{X}$ is univariate normal for all vectors $\boldsymbol{a} \in \mathbb{R}^d \setminus \{\boldsymbol{0}\}$.

Marginal distributions. It is clear from (6.13) that the univariate marginal distributions of $\boldsymbol{X}$ must be univariate normal. More generally, using the $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2')'$ notation from Section 6.1.1 and extending this notation naturally to $\boldsymbol{\mu}$ and Σ,

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

property (6.12) implies that the marginal distributions of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are also multivariate normal and are given by $\boldsymbol{X}_1 \sim N_k(\boldsymbol{\mu}_1, \Sigma_{11})$ and $\boldsymbol{X}_2 \sim N_{d-k}(\boldsymbol{\mu}_2, \Sigma_{22})$.

Conditional distributions. Assuming that Σ is positive definite, the conditional distributions of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1$ and of $\boldsymbol{X}_1$ given $\boldsymbol{X}_2$ may also be shown to be multivariate normal. For example, $\boldsymbol{X}_2 \mid \boldsymbol{X}_1 = \boldsymbol{x}_1 \sim N_{d-k}(\boldsymbol{\mu}_{2.1}, \Sigma_{22.1})$, where

$$\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1) \quad \text{and} \quad \Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

are the conditional mean vector and covariance matrix.

Quadratic forms. If $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$ with Σ positive definite, then

$$(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi_d^2, \tag{6.14}$$

a chi-squared distribution with d degrees of freedom. This is seen by observing that $\boldsymbol{Z} = \Sigma^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}) \sim N_d(\boldsymbol{0}, I_d)$ and $(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) = \boldsymbol{Z}'\boldsymbol{Z} \sim \chi_d^2$. This property (6.14) is useful for checking multivariate normality (see Section 6.1.4).

Convolutions. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent d-dimensional random vectors satisfying $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$ and $\boldsymbol{Y} \sim N_d(\tilde{\boldsymbol{\mu}}, \tilde{\Sigma})$, then we may take the product of characteristic functions to show that $\boldsymbol{X} + \boldsymbol{Y} \sim N_d(\boldsymbol{\mu} + \tilde{\boldsymbol{\mu}}, \Sigma + \tilde{\Sigma})$.

6.1.4 Testing Multivariate Normality

We now consider the issue of testing whether the data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are observations from a multivariate normal distribution.

Univariate tests. If $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are iid multivariate normal, then for $1 \leqslant j \leqslant d$ the univariate sample $X_{1,j}, \ldots, X_{n,j}$ consisting of the observations of the jth component must be iid univariate normal; in fact, any univariate sample constructed from a linear combination of the data of the form $\boldsymbol{a}'\boldsymbol{X}_1, \ldots, \boldsymbol{a}'\boldsymbol{X}_n$ must be iid univariate normal. This can be assessed graphically with a Q–Q plot against a standard normal reference distribution, or it can be tested formally using one of the many numerical tests of normality (see Section 3.1.2 for more details of univariate tests of normality).

Multivariate tests. To test for multivariate normality it is not sufficient to test that the univariate margins of the distribution are normal. We will see in Chapter 7 that it is possible to have multivariate distributions with normal margins that are not themselves multivariate normal distributions. Thus we also need to be able to test *joint normality*, and a simple way of doing this is to exploit the fact that the quadratic

form in (6.14) has a chi-squared distribution. Suppose we estimate $\boldsymbol{\mu}$ and Σ using the standard estimators in (6.9) and construct the data

$$\{D_i^2 = (\boldsymbol{X}_i - \bar{\boldsymbol{X}})' S^{-1} (\boldsymbol{X}_i - \bar{\boldsymbol{X}}) : i = 1, \ldots, n\}. \tag{6.15}$$

Because the estimates of the mean vector and the covariance matrix are used in the construction of each D_i^2, these data are not independent, even if the original $\boldsymbol{X}_i$ data were. Moreover, the marginal distribution of D_i^2 under the null hypothesis is not exactly chi-squared; in fact, we have that $n(n-1)^{-2} D_i^2 \sim \text{Beta}(\frac{1}{2}d, \frac{1}{2}(n-d-1))$, so that the true distribution is a scaled beta distribution, although it turns out to be very close to chi-squared for large n. We expect $D_1^2, \ldots, D_n^2$ to behave roughly like an iid sample from a χ_d^2 distribution, and for simplicity we construct Q–Q plots against this distribution. (It is also possible to make Q–Q plots against the beta reference distribution, and these look very similar.)

Numerical tests of multivariate normality based on multivariate measures of skewness and kurtosis are also possible. Suppose we define, in analogy to (3.1),

$$b_d = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} D_{ij}^3, \qquad k_d = \frac{1}{n} \sum_{i=1}^{n} D_i^4, \tag{6.16}$$

where D_i is given in (6.15) and is known as the *Mahalanobis distance* between $\boldsymbol{X}_i$ and $\bar{\boldsymbol{X}}$, and $D_{ij} = (\boldsymbol{X}_i - \bar{\boldsymbol{X}}) S^{-1} (\boldsymbol{X}_j - \bar{\boldsymbol{X}})$ is known as the *Mahalanobis angle* between $\boldsymbol{X}_i - \bar{\boldsymbol{X}}$ and $\boldsymbol{X}_j - \bar{\boldsymbol{X}}$. Under the null hypothesis of multivariate normality the asymptotic distributions of these statistics as $n \to \infty$ are

$$\tfrac{1}{6} n b_d \sim \chi^2_{d(d+1)(d+2)/6}, \qquad \frac{k_d - d(d+2)}{\sqrt{8d(d+2)/n}} \sim N(0, 1). \tag{6.17}$$

Mardia's test of multinormality involves comparing the skewness and kurtosis statistics with the above theoretical reference distributions. Since large values of the statistics cast doubt on the multivariate normal model, one-sided tests are generally performed. Usually, the tests of kurtosis and skewness are performed separately, although there are also a number of joint (or so-called omnibus) tests (see Notes and Comments).

Example 6.3 (on the normality of returns on Dow Jones 30 stocks). In Section 3.1.2 we applied univariate tests of normality to an arbitrary subgroup of ten stocks from the Dow Jones index. We took eight years of data spanning the period 1993–2000 and formed daily, weekly, monthly and quarterly logarithmic returns. In this example we apply Mardia's tests of multinormality based on both multivariate skewness and kurtosis to the multivariate data for all ten stocks. The results are shown in Table 6.1. We also compare the D_i^2 data (6.15) to a χ_{10}^2 distribution using a Q–Q plot (see Figure 6.2).

The daily, weekly and monthly return data fail the multivariate tests of normality. For quarterly return data the multivariate kurtosis test does not reject the null hypothesis but the skewness test does; the Q–Q plot in Figure 6.2 (d) looks slightly

Table 6.1. Mardia's tests of multivariate normality based on the multivariate measures of skewness and kurtosis in (6.16) and the asymptotic distributions in (6.17) (see Example 6.3 for details).

		Daily	Weekly	Monthly	Quarterly
	n	2020	416	96	32
	b_{10}	9.31	9.91	21.10	50.10
	p-value	0.00	0.00	0.00	0.02
	k_{10}	242.45	177.04	142.65	120.83
	p-value	0.00	0.00	0.00	0.44

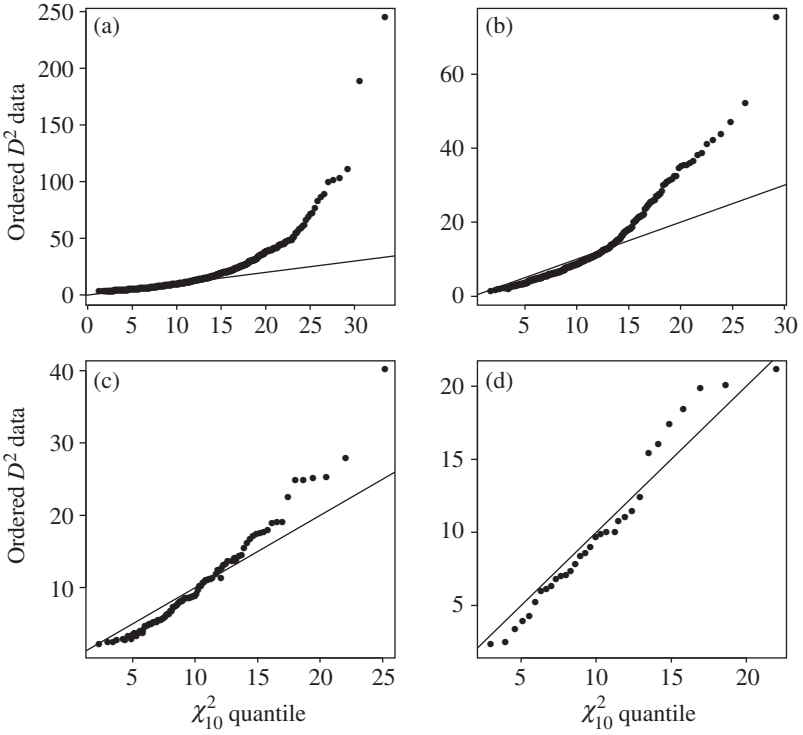

Figure 6.2. Q–Q plot of the D_i^2 data in (6.15) against a χ^2_{10} distribution for the data sets of Example 6.3: (a) daily analysis, (b) weekly analysis, (c) monthly analysis and (d) quarterly analysis. Under the null hypothesis of *multivariate* normality these should be roughly linear.

more linear. There is therefore some evidence that returns over a quarter year are close to being normally distributed, which might indicate a central limit theorem effect taking place, although the sample size is too small to reach any more reliable conclusion. The evidence against the multivariate normal distribution is certainly overwhelming for daily, weekly and monthly data.

The results in Example 6.3 are fairly typical for financial return data. This suggests that in many risk-management applications the multivariate normal distribution is

not a good description of reality. It has three main defects, all of which are discussed at various points in this book.

(1) The tails of its univariate marginal distributions are too thin; they do not assign enough weight to *extreme* events (see also Section 3.1.2).

(2) The joint tails of the distribution do not assign enough weight to *joint extreme* outcomes (see also Section 7.3.1).

(3) The distribution has a strong form of symmetry, known as elliptical symmetry.

In the next section we look at models that address some of these defects. We consider normal variance mixture models, which share the elliptical symmetry of the multivariate normal but have the flexibility to address (1) and (2) above; we also look at normal mean–variance mixture models, which introduce some asymmetry and thus address (3).

Notes and Comments

Much of the material covered briefly in Section 6.1 can be found in greater detail in standard texts on multivariate statistical analysis such as Mardia, Kent and Bibby (1979), Seber (1984), Giri (1996) and Johnson and Wichern (2002).

The true distribution of $D_i^2 = (\boldsymbol{X}_i - \bar{\boldsymbol{X}})S^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})$ for iid Gaussian data was shown by Gnanadesikan and Kettenring (1972) to be a scaled beta distribution (see also Gnanadesikan 1997). The implications of this fact for the construction of Q–Q plots in small samples are considered by Small (1978). References for multivariate measures of skewness and kurtosis and Mardia's test of multinormality are Mardia (1970, 1974, 1975). See also Mardia, Kent and Bibby (1979), the entry on "multivariate normality, testing for" in Volume 6 of the *Encyclopedia of Statistical Sciences* (Kotz, Johnson and Read 1985), and the entry on "Mardia's test of multinormality" in Volume 5 of the same publication. A paper that compares the performance of different goodness-of-fit tests for the multivariate normal distribution implemented in R is Joenssen and Vogel (2014).

6.2 Normal Mixture Distributions

In this section we generalize the multivariate normal to obtain multivariate normal mixture distributions. The crucial idea is the introduction of randomness into first the covariance matrix and then the mean vector of a multivariate normal distribution via a positive mixing variable, which will be known throughout as W.

6.2.1 Normal Variance Mixtures

Definition 6.4. The random vector $\boldsymbol{X}$ is said to have a (multivariate) normal variance mixture distribution if

$$\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \sqrt{W}A\boldsymbol{Z}, \tag{6.18}$$

where

(i) $\boldsymbol{Z} \sim N_k(\mathbf{0}, I_k)$;

(ii) $W \geqslant 0$ is a non-negative, scalar-valued rv that is independent of $\boldsymbol{Z}$, and

(iii) $A \in \mathbb{R}^{d\times k}$ and $\boldsymbol{\mu} \in \mathbb{R}^d$ are a matrix and a vector of constants, respectively.

Such distributions are known as variance mixtures, since if we condition on the rv W, we observe that $\boldsymbol{X} \mid W = w \sim N_d(\boldsymbol{\mu}, w\Sigma)$, where $\Sigma = AA'$. The distribution of $\boldsymbol{X}$ can be thought of as a composite distribution constructed by taking a set of multivariate normal distributions with the same mean vector and with the same covariance matrix up to a multiplicative constant w. The mixture distribution is constructed by drawing randomly from this set of component multivariate normals according to a set of "weights" determined by the distribution of W; the resulting mixture is not itself a multivariate normal distribution. In the context of modelling risk-factor returns, the mixing variable W could be interpreted as a *shock* that arises from new information and impacts the volatilities of all risk factors.

As for the multivariate normal, we are most interested in the case where $\operatorname{rank}(A) = d \leqslant k$ and Σ is a full-rank, positive-definite matrix; this will give us a non-singular normal variance mixture.

Provided that W has a finite expectation, we may easily calculate that

$$E(\boldsymbol{X}) = E(\boldsymbol{\mu} + \sqrt{W}A\boldsymbol{Z}) = \boldsymbol{\mu} + E(\sqrt{W})AE(\boldsymbol{Z}) = \boldsymbol{\mu}$$

and that

$$\operatorname{cov}(\boldsymbol{X}) = E((\sqrt{W}A\boldsymbol{Z})(\sqrt{W}A\boldsymbol{Z})') = E(W)AE(\boldsymbol{Z}\boldsymbol{Z}')A' = E(W)\Sigma. \quad (6.19)$$

We generally refer to $\boldsymbol{\mu}$ and Σ as the *location vector* and the *dispersion matrix* of the distribution. Note that Σ (the covariance matrix of $A\boldsymbol{Z}$) is only the covariance matrix of $\boldsymbol{X}$ if $E(W) = 1$, and that $\boldsymbol{\mu}$ is only the mean vector when $E(\boldsymbol{X})$ is defined, which requires $E(W^{1/2}) < \infty$. The correlation matrices of $\boldsymbol{X}$ and $A\boldsymbol{Z}$ are the same when $E(W) < \infty$. Note also that these distributions provide good examples of models where a lack of correlation does not necessarily imply independence of the components of $\boldsymbol{X}$; indeed, we have the following simple result.

Lemma 6.5. *Let (X_1, X_2) have a normal mixture distribution with $A = I_2$ and $E(W) < \infty$ so that $\operatorname{cov}(X_1, X_2) = 0$. Then X_1 and X_2 are independent if and only if W is almost surely constant, i.e. (X_1, X_2) are normally distributed.*

Proof. If W is almost surely a constant, then (X_1, X_2) have a bivariate normal distribution and are independent. Conversely, if (X_1, X_2) are independent, then we must have $E(|X_1||X_2|) = E(|X_1|)E(|X_2|)$. We calculate that

$$\begin{aligned} E(|X_1||X_2|) &= E(W|Z_1||Z_2|) = E(W)E(|Z_1|)E(|Z_2|) \\ &\geqslant (E(\sqrt{W}))^2 E(|Z_1|)E(|Z_2|) = E(|X_1|)E(|X_2|), \end{aligned}$$

and we can only have equality throughout when W is a constant. □

Using (6.10), we can calculate that the characteristic function of a normal variance mixture is given by

$$\begin{aligned}\phi_{\boldsymbol{X}}(\boldsymbol{t}) = E(E(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{X}} \mid W)) &= E(\exp(\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu} - \tfrac{1}{2}W\boldsymbol{t}'\Sigma\boldsymbol{t})) \\ &= \mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu}}\hat{H}(\tfrac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}), \end{aligned} \tag{6.20}$$

where $\hat{H}(\theta) = \int_0^\infty \mathrm{e}^{-\theta v}\,\mathrm{d}H(v)$ is the Laplace–Stieltjes transform of the df H of W. Based on (6.20) we use the notation $\boldsymbol{X} \sim M_d(\boldsymbol{\mu}, \Sigma, \hat{H})$ for normal variance mixtures.

Assuming that Σ is positive definite and that the distribution of W has no point mass at zero, we may derive the joint density of a normal variance mixture distribution. Writing $f_{\boldsymbol{X}|W}$ for the (Gaussian) conditional density of $\boldsymbol{X}$ given W, the density of $\boldsymbol{X}$ is given by

$$\begin{aligned} f(\boldsymbol{x}) &= \int f_{\boldsymbol{X}|W}(\boldsymbol{x} \mid w)\,\mathrm{d}H(w) \\ &= \int \frac{w^{-d/2}}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2w}\right\}\mathrm{d}H(w), \end{aligned} \tag{6.21}$$

in terms of the Lebesgue–Stieltjes integral; when H has density h we simply mean the Riemann integral $\int_0^\infty f_{\boldsymbol{X}|W}(\boldsymbol{x} \mid w)h(w)\,\mathrm{d}w$. All such densities will depend on $\boldsymbol{x}$ only through the quadratic form $(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$, and this means they are the densities of elliptical distributions, as will be discussed in Section 6.3.

Example 6.6 (multivariate two-point normal mixture distribution). Simple examples of normal mixtures are obtained when W is a discrete rv. For example, the two-point normal mixture model is obtained by taking W in (6.18) to be a discrete rv that assumes the distinct positive values k_1 and k_2 with probabilities p and $1-p$, respectively. By setting k_2 large relative to k_1 and choosing p large, this distribution might be used to define two regimes: an *ordinary* regime that holds most of the time and a *stress* regime that occurs with small probability $1-p$. Obviously this idea extends to k-point mixture models.

Example 6.7 (multivariate t distribution). If we take W in (6.18) to be an rv with an inverse gamma distribution $W \sim \mathrm{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$ (which is equivalent to saying that $\nu/W \sim \chi^2_\nu$), then $\boldsymbol{X}$ has a multivariate t distribution with ν degrees of freedom (see Section A.2.6 for more details concerning the inverse gamma distribution). Our notation for the multivariate t is $\boldsymbol{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$, and we note that Σ is not the covariance matrix of $\boldsymbol{X}$ in this definition of the multivariate t. Since $E(W) = \nu/(\nu-2)$ we have $\mathrm{cov}(\boldsymbol{X}) = (\nu/(\nu-2))\Sigma$, and the covariance matrix (and correlation matrix) of this distribution is only defined if $\nu > 2$.

Using (6.21), the density can be calculated to be

$$f(\boldsymbol{x}) = \frac{\Gamma(\frac{1}{2}(\nu+d))}{\Gamma(\frac{1}{2}\nu)(\pi\nu)^{d/2}|\Sigma|^{1/2}}\left(1 + \frac{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{\nu}\right)^{-(\nu+d)/2}. \tag{6.22}$$

Clearly, the locus of points with equal density is again an ellipsoid with equation $(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = c$ for some $c > 0$. A bivariate example with four degrees of freedom is given in Figure 6.1. In comparison with the multivariate normal, the contours of equal density rise more quickly in the centre of the distribution and decay more gradually on the "lower slopes" of the distribution. In comparison with the multivariate normal, the multivariate t has heavier marginal tails (as discussed in Section 5.1.2) and a more pronounced tendency to generate simultaneous extreme values (see also Section 7.3.1).

Example 6.8 (symmetric generalized hyperbolic distribution). A flexible family of normal variance mixtures is obtained by taking W in (6.18) to have a generalized inverse Gaussian (GIG) distribution, $W \sim N^-(\lambda, \chi, \psi)$ (see Section A.2.5). Using (6.21), it can be shown that a normal variance mixture constructed with this mixing density has the joint density

$$f(\boldsymbol{x}) = \frac{(\sqrt{\chi\psi})^{-\lambda}\psi^{d/2}}{(2\pi)^{d/2}|\Sigma|^{1/2}K_\lambda(\sqrt{\chi\psi})} \frac{K_{\lambda-(d/2)}(\sqrt{(\chi+(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))\psi})}{(\sqrt{(\chi+(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))\psi})^{(d/2)-\lambda}}, \tag{6.23}$$

where K_λ denotes a modified Bessel function of the third kind (see Section A.2.5 for more details). This distribution is a special case of the more general family of multivariate generalized hyperbolic distributions, which we will discuss in greater detail in Section 6.2.2. The more general family can be obtained as *mean–variance* mixtures of normals, which are not necessarily elliptical distributions.

The GIG mixing distribution is very flexible and contains the gamma and inverse gamma distributions as special boundary cases (corresponding, respectively, to $\lambda > 0$, $\chi = 0$ and to $\lambda < 0$, $\psi = 0$). In these cases the density in (6.23) should be interpreted as a limit as $\chi \to 0$ or as $\psi \to 0$. (Information on the limits of Bessel functions is found in Section A.2.5.) The gamma mixing distribution yields Laplace distributions or so-called symmetric variance-gamma (VG) models, and the inverse gamma yields the t as in Example 6.7; to be precise, the t corresponds to the case when $\lambda = -\nu/2$ and $\chi = \nu$. The special cases $\lambda = -0.5$ and $\lambda = 1$ have also attracted attention in financial modelling. The former gives rise to the symmetric normal inverse Gaussian (NIG) distribution; the latter gives rise to a symmetric multivariate distribution whose one-dimensional margins are known simply as hyperbolic distributions.

To calculate the covariance matrix of distributions in the symmetric generalized hyperbolic family, we require the mean of the GIG distribution, which is given in (A.15) for the case $\chi > 0$ and $\psi > 0$. The covariance matrix of the multivariate distribution in (6.23) follows from (6.19).

Normal variance mixture distributions are easy to work with under linear operations, as shown in the following simple proposition.

Proposition 6.9. *If* $\boldsymbol{X} \sim M_d(\boldsymbol{\mu}, \Sigma, \hat{H})$ *and* $\boldsymbol{Y} = B\boldsymbol{X} + \boldsymbol{b}$, *where* $B \in \mathbb{R}^{k\times d}$ *and* $\boldsymbol{b} \in \mathbb{R}^k$, *then* $\boldsymbol{Y} \sim M_k(B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B', \hat{H})$.

Proof. The characteristic function in (6.20) may be used to show that

$$\phi_{\boldsymbol{Y}}(\boldsymbol{t}) = E(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'(B\boldsymbol{X}+\boldsymbol{b})}) = \mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{b}}\phi_{\boldsymbol{X}}(B'\boldsymbol{t}) = \mathrm{e}^{\mathrm{i}\boldsymbol{t}'(B\boldsymbol{\mu}+\boldsymbol{b})}\hat{H}(\tfrac{1}{2}\boldsymbol{t}'B\Sigma B'\boldsymbol{t}).$$

□

The subclass of mixture distributions specified by $\hat{H}$ is therefore closed under linear transformations. For example, if $\boldsymbol{X}$ has a multivariate t distribution with ν degrees of freedom, then so does any linear transformation of $\boldsymbol{X}$; the linear combination $\boldsymbol{a}'\boldsymbol{X}$ would have a univariate t distribution with ν degrees of freedom (more precisely, the distribution $\boldsymbol{a}'\boldsymbol{X} \sim t_1(\nu, \boldsymbol{a}'\boldsymbol{\mu}, \boldsymbol{a}'\Sigma\boldsymbol{a})$).

Normal variance mixture distributions (and the mean–variance mixtures considered later in Section 6.2.2) are easily simulated, the method being obvious from Definition 6.4. To generate a variate $\boldsymbol{X} \sim M_d(\boldsymbol{\mu}, \Sigma, \hat{H})$ with Σ positive definite, we use the following algorithm.

Algorithm 6.10 (simulation of normal variance mixtures).

(1) Generate $\boldsymbol{Z} \sim N_d(\boldsymbol{0}, \Sigma)$ using Algorithm 6.2.

(2) Generate independently a positive mixing variable W with df H (corresponding to the Laplace–Stieltjes transform $\hat{H}$).

(3) Set $\boldsymbol{X} = \boldsymbol{\mu} + \sqrt{W}\boldsymbol{Z}$.

To generate $\boldsymbol{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$, the mixing variable W should have an $\mathrm{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$ distribution; it is helpful to note that in this case $\nu/W \sim \chi^2_\nu$, a chi-squared distribution with ν degrees of freedom. Sampling from a generalized hyperbolic distribution with density (6.23) requires us to generate $W \sim N^-(\lambda, \chi, \psi)$. Sampling from the GIG distribution can be accomplished using a rejection algorithm proposed by Atkinson (1982).

6.2.2 Normal Mean–Variance Mixtures

All of the multivariate distributions we have considered so far have elliptical symmetry (see Section 6.3.2 for explanation) and this may well be an oversimplified model for real risk-factor return data. Among other things, elliptical symmetry implies that all one-dimensional marginal distributions are rigidly symmetric, which contradicts the frequent observation for stock returns that negative returns (losses) have heavier tails than positive returns (gains). The models we now introduce attempt to add some asymmetry to the class of normal mixtures by mixing normal distributions with different means as well as different variances; this yields the class of multivariate normal mean–variance mixtures.

Definition 6.11. The random vector $\boldsymbol{X}$ is said to have a (multivariate) normal mean–variance mixture distribution if

$$\boldsymbol{X} \overset{\mathrm{d}}{=} \boldsymbol{m}(W) + \sqrt{W}A\boldsymbol{Z}, \tag{6.24}$$

where

(i) $\boldsymbol{Z} \sim N_k(\mathbf{0}, I_k)$,

(ii) $W \geqslant 0$ is a non-negative, scalar-valued rv which is independent of $\boldsymbol{Z}$,

(iii) $A \in \mathbb{R}^{d \times k}$ is a matrix, and

(iv) $\boldsymbol{m}\colon [0, \infty) \to \mathbb{R}^d$ is a measurable function.

In this case we have that

$$\boldsymbol{X} \mid W = w \sim N_d(\boldsymbol{m}(w), w\Sigma), \tag{6.25}$$

where $\Sigma = AA'$ and it is clear why such distributions are known as mean–variance mixtures of normals. In general, such distributions are not elliptical.

A possible concrete specification for the function $\boldsymbol{m}(W)$ in (6.25) is

$$\boldsymbol{m}(W) = \boldsymbol{\mu} + W\boldsymbol{\gamma}, \tag{6.26}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ are parameter vectors in $\mathbb{R}^d$. Since $E(\boldsymbol{X} \mid W) = \boldsymbol{\mu} + W\boldsymbol{\gamma}$ and $\operatorname{cov}(\boldsymbol{X} \mid W) = W\Sigma$, it follows in this case by simple calculations that

$$E(\boldsymbol{X}) = E(E(\boldsymbol{X} \mid W)) = \boldsymbol{\mu} + E(W)\boldsymbol{\gamma}, \tag{6.27}$$

$$\begin{aligned} \operatorname{cov}(\boldsymbol{X}) &= E(\operatorname{cov}(\boldsymbol{X} \mid W)) + \operatorname{cov}(E(\boldsymbol{X} \mid W)) \\ &= E(W)\Sigma + \operatorname{var}(W)\boldsymbol{\gamma}\boldsymbol{\gamma}' \end{aligned} \tag{6.28}$$

when the mixing variable W has finite variance. We observe from (6.27) and (6.28) that the parameters $\boldsymbol{\mu}$ and Σ are not, in general, the mean vector and covariance matrix of $\boldsymbol{X}$ (or a multiple thereof). This is only the case when $\boldsymbol{\gamma} = \mathbf{0}$, so that the distribution is a normal variance mixture and the simpler moment formulas given in (6.19) apply.

6.2.3 Generalized Hyperbolic Distributions

In Example 6.8 we looked at the special subclass of the generalized hyperbolic (GH) distributions consisting of the elliptically symmetric normal variance mixture distributions. The full GH family is obtained using the mean–variance mixture construction (6.24) and the conditional mean specification (6.26). For the mixing distribution we assume that $W \sim N^-(\lambda, \chi, \psi)$, a GIG distribution with density (A.14).

Remark 6.12. This class of distributions has received a lot of attention in the financial-modelling literature, particularly in the univariate case. An important reason for this attention is their link to Lévy processes, i.e. processes with independent and stationary increments (like Brownian motion or the compound Poisson distribution) that are used to model price processes in continuous time. For every GH distribution it is possible to construct a Lévy process so that the value of the increment of the process over a fixed time interval has that distribution; this is only possible because the GH law is a so-called infinitely divisible distribution, a property that it inherits from the GIG mixing distribution of W.

The joint density in the non-singular case (Σ has rank d) is

$$f(\boldsymbol{x}) = \int_0^\infty \frac{\mathrm{e}^{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma}}}{(2\pi)^{d/2}|\Sigma|^{1/2}w^{d/2}} \times \exp\left\{-\frac{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2w} - \frac{\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma}}{2/w}\right\} h(w)\,\mathrm{d}w,$$

where $h(w)$ is the density of W. Evaluation of this integral gives the GH density

$$f(\boldsymbol{x}) = c\frac{K_{\lambda-(d/2)}(\sqrt{(\chi+(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))(\psi+\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})})\mathrm{e}^{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma}}}{(\sqrt{(\chi+(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))(\psi+\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})})^{(d/2)-\lambda}}, \tag{6.29}$$

where the normalizing constant is

$$c = \frac{(\sqrt{\chi\psi})^{-\lambda}\psi^{\lambda}(\psi+\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})^{(d/2)-\lambda}}{(2\pi)^{d/2}|\Sigma|^{1/2}K_{\lambda}(\sqrt{\chi\psi})}.$$

Clearly, if $\boldsymbol{\gamma} = \mathbf{0}$, the distribution reduces to the symmetric GH special case of Example 6.8. In general, we have a non-elliptical distribution with asymmetric margins. The mean vector and covariance matrix of the distribution are easily calculated from (6.27) and (6.28) using the information on the GIG and its moments given in Section A.2.5. The characteristic function of the GH distribution may be calculated using the same approach as in (6.20) to yield

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = E(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{X}}) = \mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu}}\hat{H}(\tfrac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t} - \mathrm{i}\boldsymbol{t}'\boldsymbol{\gamma}), \tag{6.30}$$

where $\hat{H}$ is the Laplace–Stieltjes transform of the GIG distribution.

We adopt the notation $\boldsymbol{X} \sim \mathrm{GH}_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$. Note that the distributions $\mathrm{GH}_d(\lambda, \chi/k, k\psi, \boldsymbol{\mu}, k\Sigma, k\boldsymbol{\gamma})$ and $\mathrm{GH}_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$ are identical for any $k > 0$, which causes an *identifiability problem* when we attempt to estimate the parameters in practice. This can be solved by constraining the determinant $|\Sigma|$ to be a particular value (such as one) when fitting. Note that, while such a constraint will have an effect on the values of χ and ψ that we estimate, it will not have an effect on the value of $\chi\psi$, so this product is a useful summary parameter for the GH distribution.

Linear combinations. The GH class is closed under linear operations.

Proposition 6.13. *If $\boldsymbol{X} \sim \mathrm{GH}_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$ and $\boldsymbol{Y} = B\boldsymbol{X} + \boldsymbol{b}$, where $B \in \mathbb{R}^{k\times d}$ and $\boldsymbol{b} \in \mathbb{R}^k$, then $\boldsymbol{Y} \sim \mathrm{GH}_k(\lambda, \chi, \psi, B\boldsymbol{\mu}+\boldsymbol{b}, B\Sigma B', B\boldsymbol{\gamma})$.*

Proof. We calculate, using (6.30) and a similar method to Proposition 6.9, that

$$\phi_{\boldsymbol{Y}}(\boldsymbol{t}) = \mathrm{e}^{\mathrm{i}\boldsymbol{t}'(B\boldsymbol{\mu}+\boldsymbol{b})}\hat{H}(\tfrac{1}{2}\boldsymbol{t}'B\Sigma B'\boldsymbol{t} - \mathrm{i}\boldsymbol{t}'B\boldsymbol{\gamma}).$$

□

The parameters inherited from the GIG mixing distribution therefore remain unchanged under linear operations. This means, for example, that margins of $\boldsymbol{X}$ are easy to calculate; we have that $X_i \sim \mathrm{GH}_1(\lambda, \chi, \psi, \mu_i, \Sigma_{ii}, \gamma_i)$.

Parametrizations. There is a bewildering array of alternative parametrizations for the GH distribution in the literature, and it is more common to meet this distribution in a reparametrized form. In one common version the dispersion matrix we call Σ is renamed Δ and the constraint that $|\Delta| = 1$ is imposed; this addresses the identifiability problem mentioned above. The skewness parameters $\boldsymbol{\gamma}$ are replaced by parameters $\boldsymbol{\beta}$, and the non-negative parameters χ and ψ are replaced by the non-negative parameters δ and α according to

$$\boldsymbol{\beta} = \Delta^{-1}\boldsymbol{\gamma}, \qquad \delta = \sqrt{\chi}, \qquad \alpha = \sqrt{\psi + \boldsymbol{\gamma}'\Delta^{-1}\boldsymbol{\gamma}}.$$

These parameters must satisfy the constraints $\delta \geqslant 0$, $\alpha^2 > \boldsymbol{\beta}'\Delta\boldsymbol{\beta}$ if $\lambda > 0$; $\delta > 0$, $\alpha^2 > \boldsymbol{\beta}'\Delta\boldsymbol{\beta}$ if $\lambda = 0$; and $\delta > 0$, $\alpha^2 \geqslant \boldsymbol{\beta}'\Delta\boldsymbol{\beta}$ if $\lambda < 0$. Blæsild (1981) uses this parametrization to show that GH distributions form a closed class of distributions under linear operations and conditioning. However, the parametrization does have the problem that the important parameters α and δ are not generally invariant under either of these operations.

It is useful to be able to move easily between our χ–ψ–Σ–$\boldsymbol{\gamma}$ parametrization, as in (6.29), and the α–δ–Δ–$\boldsymbol{\beta}$ parametrization; λ and $\boldsymbol{\mu}$ are common to both parametrizations. If the χ–ψ–Σ–$\boldsymbol{\gamma}$ parametrization is used, then the formulas for obtaining the other parametrization are

$$\Delta = |\Sigma|^{-1/d}\Sigma, \qquad \boldsymbol{\beta} = \Sigma^{-1}\boldsymbol{\gamma},$$
$$\delta = \sqrt{\chi|\Sigma|^{1/d}}, \qquad \alpha = \sqrt{|\Sigma|^{-1/d}(\psi + \boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})}.$$

If the α–δ–Δ–$\boldsymbol{\beta}$ form is used, then we can obtain our parametrization by setting

$$\Sigma = \Delta, \qquad \gamma = \Delta\boldsymbol{\beta}, \qquad \chi = \delta^2, \qquad \psi = \alpha^2 - \boldsymbol{\beta}'\Delta\boldsymbol{\beta}.$$

Special cases. The multivariate GH family is extremely flexible and, as we have mentioned, contains many special cases known by alternative names.

- If $\lambda = \frac{1}{2}(d+1)$, we drop the word "generalized" and refer to the distribution as a d-dimensional hyperbolic distribution. Note that the univariate margins of this distribution also have $\lambda = \frac{1}{2}(d+1)$ and are not one-dimensional hyperbolic distributions.
- If $\lambda = 1$, we get a multivariate distribution whose univariate margins are one-dimensional hyperbolic distributions. The one-dimensional hyperbolic distribution has been widely used in univariate analyses of financial return data (see Notes and Comments).
- If $\lambda = -\frac{1}{2}$, then the distribution is known as an NIG distribution. In the univariate case this model has also been used in analyses of return data; its functional form is similar to the hyperbolic distribution but with a slightly heavier tail. (Note that the NIG and the GIG are different distributions!)
- If $\lambda > 0$ and $\chi = 0$, we get a limiting case of the distribution known variously as a generalized Laplace, Bessel function or VG distribution.

- If $\lambda = -\frac{1}{2}\nu$, $\chi = \nu$ and $\psi = 0$, we get another limiting case that seems to have been less well studied; it could be called an asymmetric or skewed t distribution. Evaluating the limit of (6.29) as $\psi \to 0$ yields the multivariate density

$$f(\boldsymbol{x}) = c\frac{K_{(\nu+d)/2}(\sqrt{(\nu + Q(\boldsymbol{x}))\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma}})\exp((\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma})}{(\sqrt{(\nu + Q(\boldsymbol{x}))\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma}})^{-(\nu+d)/2}(1 + (Q(\boldsymbol{x})/\nu))^{(\nu+d)/2}}, \quad (6.31)$$

  where $Q(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ and the normalizing constant is

$$c = \frac{2^{1-(\nu+d)/2}}{\Gamma(\frac{1}{2}\nu)(\pi\nu)^{d/2}|\Sigma|^{1/2}}.$$

  This density reduces to the standard multivariate t density in (6.22) as $\boldsymbol{\gamma} \to \mathbf{0}$.

6.2.4 Empirical Examples

In this section we fit the multivariate GH distribution to real data and examine which of the subclasses—such as t, hyperbolic or NIG—are most useful; we also explore whether the general mean–variance mixture models can be replaced by (elliptically symmetric) variance mixtures. Our first example prepares the ground for multivariate examples by looking briefly at univariate models. The univariate distributions are fitted by straightforward numerical maximization of the log-likelihood. The multivariate distributions are fitted by using a variant of the EM algorithm, as described in Section 15.1.1.

Example 6.14 (univariate stock returns). In the literature, the NIG, hyperbolic and t models have been particularly popular special cases. We fit symmetric and asymmetric cases of these distributions to the data used in Example 6.3, restricting attention to daily and weekly returns, where the data are more plentiful ($n = 2020$ and $n = 468$, respectively). Models are fitted using maximum likelihood under the simplifying assumption that returns form iid samples; a simple quasi-Newton method provides a viable alternative to the EM algorithm in the univariate case.

In the upper two panels of Table 6.2 we show results for symmetric models. The t, NIG and hyperbolic models may be compared directly using the log-likelihood at the maximum, since all have the same number of parameters: for daily data we find that eight out of ten stocks prefer the t distribution to the hyperbolic and NIG distributions; for weekly returns the t distribution is favoured in six out of ten cases. Overall, the second best model appears to be the NIG distribution. The mixture models fit much better than the Gaussian model in all cases, and it may be easily verified using the Akaike information criterion (AIC) that they are preferred to the Gaussian model in a formal comparison (see Section A.3.6 for more on the AIC).

For the asymmetric models, we only show cases where at least one of the asymmetric t, NIG or hyperbolic models offered a significant improvement ($p < 0.05$) on the corresponding symmetric model according to a likelihood ratio test. This occurred for weekly returns on Citigroup (C) and Intel (INTC) but for no daily returns. For Citigroup the p-values of the tests were, respectively, 0.06, 0.04 and

Table 6.2. Comparison of univariate models in the GH family, showing estimates of selected parameters and the value of the log-likelihood at the maximum; bold numbers indicate the models that give the largest values of the log-likelihood. See Example 6.14 for commentary.

	Gauss	t model		NIG model		Hyperbolic model	
Stock	$\ln L$	ν	$\ln L$	$\sqrt{\chi\psi}$	$\ln L$	$\sqrt{\chi\psi}$	$\ln L$
Daily returns: symmetric models							
AXP	4945.7	5.8	5001.8	1.6	**5002.4**	1.3	5002.1
EK	5112.9	3.8	**5396.2**	0.8	5382.5	0.6	5366.0
BA	5054.9	3.8	**5233.5**	0.8	5229.1	0.5	5221.2
C	4746.6	6.3	**4809.5**	1.9	4806.8	1.7	4805.0
KO	5319.6	5.1	**5411.0**	1.4	5407.3	1.3	5403.3
MSFT	4724.3	5.8	**4814.6**	1.6	4809.5	1.5	4806.4
HWP	4480.1	4.5	**4588.8**	1.1	4587.2	0.9	4583.4
INTC	4392.3	5.4	**4492.2**	1.5	4486.7	1.4	4482.4
JPM	4898.3	5.1	4967.8	1.3	4969.5	0.9	**4969.7**
DIS	5047.2	4.4	**5188.3**	1	5183.8	0.8	5177.6
Weekly returns: symmetric models							
AXP	719.9	8.8	724.2	3.0	**724.3**	2.8	724.3
EK	718.7	3.6	**765.6**	0.7	764.0	0.5	761.3
BA	732.4	4.4	**759.2**	1.0	758.3	0.8	757.2
C	656.0	5.7	**669.6**	1.6	669.3	1.3	669
KO	757.1	6.0	765.7	1.7	766.2	1.3	**766.3**
MSFT	671.5	6.3	**683.9**	1.9	683.2	1.8	682.9
HWP	627.1	6.0	637.3	1.8	**637.3**	1.5	637.1
INTC	595.8	5.2	**611.0**	1.5	610.6	1.3	610
JPM	681.7	5.9	**693.0**	1.7	692.9	1.5	692.6
DIS	734.1	6.4	742.7	1.9	**742.8**	1.7	742.7
Weekly returns: asymmetric models							
C	NA	6.1	**671.4**	1.7	671.3	1.3	671.2
INTC	NA	6.3	**614.2**	1.8	613.9	1.7	613.3

0.04 for the t, NIG and hyperbolic cases; for Intel the p-values were 0.01 in all cases, indicating quite strong asymmetry.

In the case of Intel we have superimposed the densities of various fitted asymmetric distributions on a histogram of the data in Figure 6.3. A plot of the log densities shown alongside reveals the differences between the distributions in the tail area. The left tail (corresponding to losses) appears to be heavier for these data, and the best-fitting distribution according to the likelihood comparison is the asymmetric t distribution.

Example 6.15 (multivariate stock returns). We fitted multivariate models to the full ten-dimensional data set of log-returns used in the previous example. The resulting values of the maximized log-likelihood are shown in Table 6.3 along with p-values for a likelihood ratio test of all special cases against the (asymmetric) GH model. The number of parameters in each model is also given; note that the general

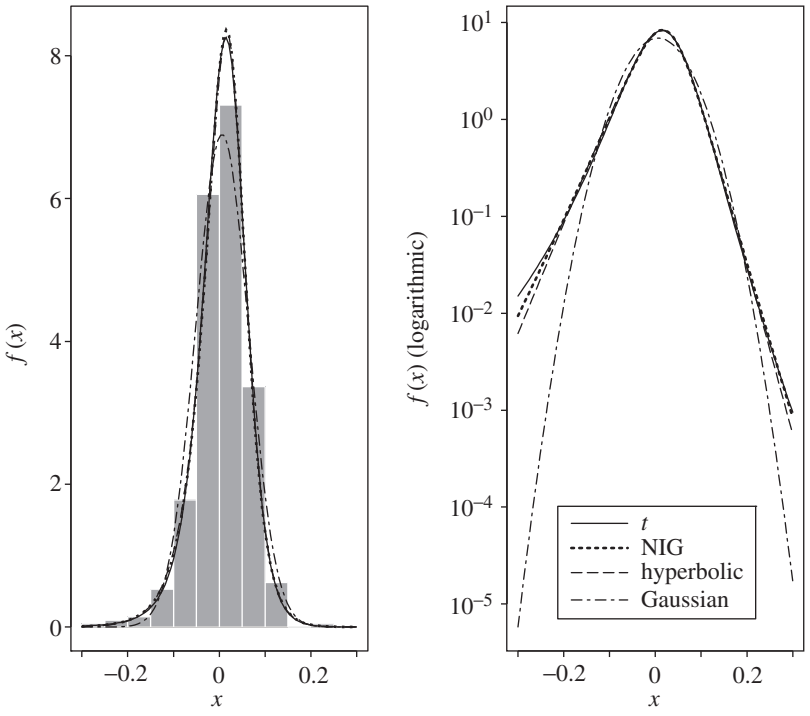

Figure 6.3. Models for weekly returns on Intel (INTC).

d-dimensional GH model has $\frac{1}{2}d(d+1)$ dispersion parameters, d location parameters, d skewness parameters and three parameters coming from the GIG mixing distribution, but is subject to one identifiability constraint; this gives $\frac{1}{2}(d(d+5)+4)$ free parameters.

For the daily data the best of the special cases is the skewed t distribution, which gives a value for the maximized likelihood that cannot be discernibly improved by the more general model with its additional parameter. All other non-elliptically symmetric submodels are rejected in a likelihood ratio test. Note, however, that the elliptically symmetric t distribution cannot be rejected when compared with the most general model, so that this seems to offer a simple parsimonious model for these data (the estimated degree of freedom is 6.0).

For the weekly data the best special case is the NIG distribution, followed closely by the skewed t; the hyperbolic and VG are rejected. The best elliptically symmetric special case seems to be the t distribution (the estimated degree of freedom this time being 6.2).

Example 6.16 (multivariate exchange-rate returns). We fitted the same multivariate models to a four-dimensional data set of exchange-rate log-returns, these being sterling, the euro, Japanese yen and Swiss franc against the US dollar for the period January 2000 to the end of March 2004 (1067 daily returns and 222 weekly returns). The resulting values of the maximized log-likelihood are shown in Table 6.4.

Table 6.3. A comparison of models in the GH family for ten-dimensional stock-return data. For each model, the table shows the value of the log-likelihood at the maximum ($\ln L$), the numbers of parameters ("# par.") and the p-value for a likelihood ratio test against the general GH model. The log-likelihood values for the general model, the best special case and the best elliptically symmetric special case are in bold type. See Example 6.15 for details.

	GH	NIG	Hyperbolic	t	VG	Gauss
Daily returns: asymmetric models						
$\ln L$	**52 174.62**	52 141.45	52 111.65	**52 174.62**	52 063.44	
# par.	77	76	76	76	76	
p-value		0.00	0.00	1.00	0.00	
Daily returns: symmetric models						
$\ln L$	52 170.14	52 136.55	52 106.34	**52 170.14**	52 057.38	50 805.28
# par.	67	66	66	66	66	65
p-value	0.54	0.00	0.00	0.63	0.00	0.00
Weekly returns: asymmetric models						
$\ln L$	**7 639.32**	**7 638.59**	7 636.49	7 638.56	7 631.33	
p-value		0.23	0.02	0.22	0.00	
Weekly returns: symmetric models						
$\ln L$	7 633.65	7 632.68	7 630.44	**7 633.11**	7 625.4	7 433.77
p-value	0.33	0.27	0.09	0.33	0.00	0.00

Table 6.4. A comparison of models in the GH family for four-dimensional exchange-rate return data. For each model, the table shows the value of the log-likelihood at the maximum ($\ln L$), the numbers of parameters ("# par.") and the p-value for a likelihood ratio test against the general GH model. The log-likelihood values for the general model, the best special case and the best elliptically symmetric special case are in bold type. See Example 6.16 for details.

	GH	NIG	Hyperbolic	t	VG	Gauss
Daily returns: asymmetric models						
$\ln L$	**17 306.44**	**17 306.43**	17 305.61	17 304.97	17 302.5	
# par.	20	19	19	19	19	
p-value		0.85	0.20	0.09	0.00	
Daily returns: symmetric models						
$\ln L$	17 303.10	**17 303.06**	17 302.15	17 301.85	17 299.15	17 144.38
# par.	16	15	15	15	15	14
p-value	0.15	0.24	0.13	0.10	0.01	0.00
Weekly returns: asymmetric models						
$\ln L$	**2 890.65**	2 889.90	2 889.65	**2 890.65**	2 888.98	
p-value		0.22	0.16	1.00	0.07	
Weekly returns: symmetric models						
$\ln L$	2 887.52	2 886.74	2 886.48	**2 887.52**	2 885.86	2 872.36
p-value	0.18	0.17	0.14	0.28	0.09	0.00

For the daily data the best of the special cases (both in general and if we restrict ourselves to symmetric models) is the NIG distribution, followed by the hyperbolic, t and VG distributions in that order. In a likelihood ratio test of the special cases against the general GH distribution, only the VG model is rejected at the 5% level; the skewed t model is rejected at the 10% level. When tested against the full model, certain elliptical models could not be rejected, the best of these being the NIG.

For the weekly data the best special case is the t distribution, followed by the NIG, hyperbolic and VG; none of the special cases can be rejected in a test at the 5% level, although the VG model is rejected at the 10% level. Among the elliptically symmetric distributions the Gauss distribution is clearly rejected, and the VG is again rejected at the 10% level, but otherwise the elliptical special cases are accepted; the best of these seems to be the t distribution, which has an estimated degrees-of-freedom parameter of 5.99.

Notes and Comments

Important early papers on multivariate normal mixtures are Kelker (1970) and Cambanis, Huang and Simons (1981). See also Bingham and Kiesel (2002), which contains an overview of the connections between the normal mixture, elliptical and hyperbolic models, and discusses their role in financial modelling. Fang, Kotz and Ng (1990) discuss the symmetric normal mixture models as special cases in their account of the more general family of spherical and elliptical distributions.

The GH distributions (univariate and multivariate) were introduced in Barndorff-Nielsen (1978) and further explored in Barndorff-Nielsen and Blæsild (1981). Useful references on the multivariate distribution are Blæsild (1981) and Blæsild and Jensen (1981). Generalized hyperbolic distributions (particularly in the univariate case) have been popularized as models for financial returns in recent papers by Eberlein and Keller (1995) and Eberlein, Keller and Prause (1998) (see also Bibby and Sørensen 2003). The PhD thesis of Prause (1999) is also a compendium of useful information in this context.

The reasons for their popularity in financial applications are both empirical and theoretical: they appear to provide a good fit to financial return data (again mostly in univariate investigations); they are consistent with continuous-time models, where logarithmic asset prices follow univariate or multivariate Lévy processes (thus generalizing the Black–Scholes model, where logarithmic prices follow Brownian motion); see Eberlein and Keller (1995) and Schoutens (2003).

For the NIG special case see Barndorff-Nielsen (1997), who discusses both univariate and multivariate cases and argues that the NIG is slightly superior to the hyperbolic as a univariate model for return data, a claim that our analyses support for stock-return data. Kotz, Kozubowski and Podgórski (2001) is a useful reference for the VG special case; the distribution appears here under the name generalized Laplace distribution and a (univariate or multivariate) Lévy process with VG-distributed increments is called a Laplace motion. The univariate Laplace motion is essentially the model proposed by Madan and Seneta (1990), who derived it as a Brownian motion under a stochastic time change and referred to it as the VG model

(see also Madan, Carr and Chang 1998). The multivariate t distribution is discussed in Kotz and Nadarajah (2004); the asymmetric or skewed t distribution presented in this chapter is also discussed in Bibby and Sørensen (2003). For alternative skewed extensions of the multivariate t, see Kotz and Nadarajah (2004) and Genton (2004).

6.3 Spherical and Elliptical Distributions

In the previous section we observed that normal variance mixture distributions—particularly the multivariate t and symmetric multivariate NIG—provided models that were far superior to the multivariate normal for daily and weekly US stock-return data. The more general asymmetric mean–variance mixture distributions did not seem to offer much of an improvement on the symmetric variance mixture models. While this was a single example, other investigations suggest that multivariate return data for groups of returns of a similar type often show similar behaviour.

The normal variance mixture distributions are so-called elliptical distributions, and in this section we look more closely at the theory of elliptical distributions. To do this we begin with the special case of spherical distributions.

6.3.1 Spherical Distributions

The spherical family constitutes a large class of distributions for random vectors with *uncorrelated* components and identical, symmetric marginal distributions. It is important to note that within this class, $N_d(\mathbf{0}, I_d)$ is the only model for a vector of mutually independent components. Many of the properties of elliptical distributions can best be understood by beginning with spherical distributions.

Definition 6.17. A random vector $\boldsymbol{X} = (X_1, \ldots, X_d)'$ has a spherical distribution if, for every orthogonal map $U \in \mathbb{R}^{d\times d}$ (i.e. maps satisfying $UU' = U'U = I_d$),

$$U\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{X}.$$

Thus spherical random vectors are distributionally invariant under rotations. There are a number of different ways of defining distributions with this property, as we demonstrate below.

Theorem 6.18. *The following are equivalent.*

(1) $\boldsymbol{X}$ *is spherical.*

(2) *There exists a function* ψ *of a scalar variable such that, for all* $\boldsymbol{t} \in \mathbb{R}^d$,

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = E(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{X}}) = \psi(\boldsymbol{t}'\boldsymbol{t}) = \psi(t_1^2 + \cdots + t_d^2). \tag{6.32}$$

(3) *For every* $\boldsymbol{a} \in \mathbb{R}^d$,

$$\boldsymbol{a}'\boldsymbol{X} \stackrel{\mathrm{d}}{=} \|\boldsymbol{a}\| X_1, \tag{6.33}$$

where $\|\boldsymbol{a}\|^2 = \boldsymbol{a}'\boldsymbol{a} = a_1^2 + \cdots + a_d^2$.

Proof. (1) $\Rightarrow$ (2). If $\boldsymbol{X}$ is spherical, then for any orthogonal matrix U we have

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \phi_{U\boldsymbol{X}}(\boldsymbol{t}) = E(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'U\boldsymbol{X}}) = \phi_{\boldsymbol{X}}(U'\boldsymbol{t}).$$

This can only be true if $\phi_{\boldsymbol{X}}(\boldsymbol{t})$ only depends on the length of $\boldsymbol{t}$, i.e. if $\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \psi(\boldsymbol{t}'\boldsymbol{t})$ for some function ψ of a non-negative scalar variable.

(2) $\Rightarrow$ (3). First observe that $\phi_{X_1}(t) = E(\mathrm{e}^{\mathrm{i}tX_1}) = \phi_{\boldsymbol{X}}(t\boldsymbol{e}_1) = \psi(t^2)$, where $\boldsymbol{e}_1$ denotes the first unit vector in $\mathbb{R}^d$. It follows that for any $\boldsymbol{a} \in \mathbb{R}^d$,

$$\phi_{\boldsymbol{a}'\boldsymbol{X}}(t) = \phi_{\boldsymbol{X}}(t\boldsymbol{a}) = \psi(t^2\boldsymbol{a}'\boldsymbol{a}) = \psi(t^2\|\boldsymbol{a}\|^2) = \phi_{X_1}(t\|\boldsymbol{a}\|) = \phi_{\|\boldsymbol{a}\|X_1}(t).$$

(3) $\Rightarrow$ (1). For any orthogonal matrix U we have

$$\phi_{U\boldsymbol{X}}(\boldsymbol{t}) = E(\mathrm{e}^{\mathrm{i}(U'\boldsymbol{t})'\boldsymbol{X}}) = E(\mathrm{e}^{\mathrm{i}\|U'\boldsymbol{t}\|X_1}) = E(\mathrm{e}^{\mathrm{i}\|\boldsymbol{t}\|X_1}) = E(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{X}}) = \phi_{\boldsymbol{X}}(\boldsymbol{t}).$$

□

Part (2) of Theorem 6.18 shows that the characteristic function of a spherically distributed random vector is fully described by a function ψ of a scalar variable. For this reason ψ is known as the *characteristic generator* of the spherical distribution and the notation $\boldsymbol{X} \sim S_d(\psi)$ is used. Part (3) of Theorem 6.18 shows that linear combinations of spherical random vectors always have a distribution of the same *type*, so that they have the same distribution up to changes of location and scale (see Section A.1.1). This important property will be used in Chapter 8 to prove the subadditivity of value-at-risk for linear portfolios of elliptically distributed risk factors. We now give examples of spherical distributions.

Example 6.19 (multivariate normal). A random vector $\boldsymbol{X}$ with the standard uncorrelated normal distribution $N_d(\boldsymbol{0}, I_d)$ is clearly spherical. The characteristic function is

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = E(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{X}}) = \mathrm{e}^{-\boldsymbol{t}'\boldsymbol{t}/2},$$

so that, using part (2) of Theorem 6.18, $\boldsymbol{X} \sim S_d(\psi)$ with characteristic generator $\psi(t) = \mathrm{e}^{-t/2}$.

Example 6.20 (normal variance mixtures). A random vector $\boldsymbol{X}$ with a standardized, uncorrelated normal variance mixture distribution $M_d(\boldsymbol{0}, I_d, \hat{H})$ also has a spherical distribution. Using (6.20), we see that $\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \hat{H}(\frac{1}{2}\boldsymbol{t}'\boldsymbol{t})$, which obviously satisfies (6.32), and the characteristic generator of the spherical distribution is related to the Laplace–Stieltjes transform of the mixture distribution function of W by $\psi(t) = \hat{H}(\frac{1}{2}t)$. Thus $\boldsymbol{X} \sim M_d(\boldsymbol{0}, I_d, \hat{H}(\cdot))$ and $\boldsymbol{X} \sim S_d(\hat{H}(\frac{1}{2}\cdot))$ are two ways of writing the same mixture distribution.

A further, extremely important, way of characterizing spherical distributions is given by the following result.

Theorem 6.21. *$\boldsymbol{X}$ has a spherical distribution if and only if it has the stochastic representation*

$$\boldsymbol{X} \stackrel{\mathrm{d}}{=} R\boldsymbol{S}, \tag{6.34}$$

where $\boldsymbol{S}$ is uniformly distributed on the unit sphere $\mathcal{S}^{d-1} = \{\boldsymbol{s} \in \mathbb{R}^d : \boldsymbol{s}'\boldsymbol{s} = 1\}$ and $R \geqslant 0$ is a radial rv, independent of $\boldsymbol{S}$.

Proof. First we prove that if $\boldsymbol{S}$ is uniformly distributed on the unit sphere and $R \geqslant 0$ is an independent scalar variable, then $R\boldsymbol{S}$ has a spherical distribution. This is seen by considering the characteristic function

$$\phi_{R\boldsymbol{S}}(\boldsymbol{t}) = E(\mathrm{e}^{\mathrm{i}R\boldsymbol{t}'\boldsymbol{S}}) = E(E(\mathrm{e}^{\mathrm{i}R\boldsymbol{t}'\boldsymbol{S}} \mid R)).$$

Since $\boldsymbol{S}$ is itself spherically distributed, its characteristic function has a characteristic generator, which is usually given the special notation Ω_d. Thus, by Theorem 6.18 (2) we have that

$$\phi_{R\boldsymbol{S}}(\boldsymbol{t}) = E(\Omega_d(R^2\boldsymbol{t}'\boldsymbol{t})) = \int \Omega_d(r^2\boldsymbol{t}'\boldsymbol{t})\,\mathrm{d}F(r), \tag{6.35}$$

where F is the df of R. Since this is a function of $\boldsymbol{t}'\boldsymbol{t}$, it follows, again from Theorem 6.18 (2), that $R\boldsymbol{S}$ has a spherical distribution.

We now prove that if the random vector $\boldsymbol{X}$ is spherical, then it has the representation (6.34). For any arbitrary $\boldsymbol{s} \in \mathcal{S}^{d-1}$, the characteristic generator ψ of $\boldsymbol{X}$ must satisfy $\psi(\boldsymbol{t}'\boldsymbol{t}) = \phi_{\boldsymbol{X}}(\boldsymbol{t}) = \phi_{\boldsymbol{X}}(\|\boldsymbol{t}\|\boldsymbol{s})$. It follows that, if we introduce a random vector $\boldsymbol{S}$ that is uniformly distributed on the sphere $\mathcal{S}^{d-1}$, we can write

$$\psi(\boldsymbol{t}'\boldsymbol{t}) = \int_{\mathcal{S}^{d-1}} \phi_{\boldsymbol{X}}(\|\boldsymbol{t}\|\boldsymbol{s})\,\mathrm{d}F_{\boldsymbol{S}}(\boldsymbol{s}) = \int_{\mathcal{S}^{d-1}} E(\mathrm{e}^{\mathrm{i}\|\boldsymbol{t}\|\boldsymbol{s}'\boldsymbol{X}})\,\mathrm{d}F_{\boldsymbol{S}}(\boldsymbol{s}).$$

Interchanging the order of integration and using the Ω_d notation for the characteristic generator of $\boldsymbol{S}$ we have

$$\psi(\boldsymbol{t}'\boldsymbol{t}) = E(\Omega_d(\|\boldsymbol{t}\|^2\|\boldsymbol{X}\|^2)) = \int \Omega_d(\boldsymbol{t}'\boldsymbol{t}r^2)\,\mathrm{d}F_{\|\boldsymbol{X}\|}(r), \tag{6.36}$$

where $F_{\|\boldsymbol{X}\|}$ is the df of $\|\boldsymbol{X}\|$. By comparison with (6.35) we see that (6.36) is the characteristic function of $R\boldsymbol{S}$, where R is an rv with df $F_{\|\boldsymbol{X}\|}$ that is independent of $\boldsymbol{S}$. □

We often exclude from consideration distributions that place point mass at the origin; that is, we consider spherical rvs $\boldsymbol{X}$ in the subclass $S_d^+(\psi)$ for which $P(\boldsymbol{X} = \boldsymbol{0}) = 0$. A particularly useful corollary of Theorem 6.21 is then the following result, which is used in Section 15.1.2 to devise tests for spherical and elliptical symmetry.

Corollary 6.22. *Suppose* $\boldsymbol{X} \stackrel{\mathrm{d}}{=} R\boldsymbol{S} \sim S_d^+(\psi)$. *Then*

$$\left(\|\boldsymbol{X}\|, \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}\right) \stackrel{\mathrm{d}}{=} (R, \boldsymbol{S}). \tag{6.37}$$

Proof. Let $f_1(\boldsymbol{x}) = \|\boldsymbol{x}\|$ and $f_2(\boldsymbol{x}) = \boldsymbol{x}/\|\boldsymbol{x}\|$. It follows from (6.34) that

$$\left(\|\boldsymbol{X}\|, \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}\right) = (f_1(\boldsymbol{X}), f_2(\boldsymbol{X})) \stackrel{\mathrm{d}}{=} (f_1(R\boldsymbol{S}), f_2(R\boldsymbol{S})) = (R, \boldsymbol{S}).$$

□

Example 6.23 (working with R and S). Suppose $\boldsymbol{X} \sim N_d(\mathbf{0}, I_d)$. Since $\boldsymbol{X}'\boldsymbol{X} \sim \chi^2_d$, a chi-squared distribution with d degrees of freedom, it follows from (6.37) that $R^2 \sim \chi^2_d$.

We can use this fact to calculate $E(\boldsymbol{S})$ and $\mathrm{cov}(\boldsymbol{S})$, the first two moments of a uniform distribution on the unit sphere. We have that

$$\begin{aligned} \mathbf{0} = E(\boldsymbol{X}) = E(R)E(\boldsymbol{S}) &\Rightarrow E(\boldsymbol{S}) = \mathbf{0}, \\ I_d = \mathrm{cov}(\boldsymbol{X}) = E(R^2)\,\mathrm{cov}(\boldsymbol{S}) &\Rightarrow \mathrm{cov}(\boldsymbol{S}) = I_d/d, \end{aligned} \tag{6.38}$$

since $E(R^2) = d$ when $R^2 \sim \chi^2_d$.

Now suppose that $\boldsymbol{X}$ has a spherical normal variance mixture distribution $\boldsymbol{X} \sim M_d(\mathbf{0}, I_d, \hat{H})$ and we wish to calculate the distribution of $R^2 \stackrel{\mathrm{d}}{=} \boldsymbol{X}'\boldsymbol{X}$ in this case. Since $\boldsymbol{X} \stackrel{\mathrm{d}}{=} \sqrt{W}\boldsymbol{Y}$, where $\boldsymbol{Y} \sim N_d(\mathbf{0}, I_d)$ and W is independent of $\boldsymbol{Y}$, it follows that $R^2 \stackrel{\mathrm{d}}{=} W\tilde{R}^2$, where $\tilde{R}^2 \sim \chi^2_d$ and W and $\tilde{R}$ are independent. If we can calculate the distribution of the product of W and an independent chi-squared variate, then we have the distribution of R^2.

For a concrete example suppose that $\boldsymbol{X} \sim t_d(\nu, \mathbf{0}, I_d)$. For a multivariate t distribution we know from Example 6.7 that $W \sim \mathrm{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$, which means that $\nu/W \sim \chi^2_\nu$. Using the fact that the ratio of independent chi-squared rvs divided by their degrees of freedom is F-distributed, it may be calculated that $R^2/d \sim F(d, \nu)$, the F distribution on d with ν degrees of freedom (see Section A.2.3). Since an $F(d, \nu)$ distribution has mean $\nu/(\nu - 2)$, it follows from (6.38) that

$$\mathrm{cov}(\boldsymbol{X}) = E(\mathrm{cov}(R\boldsymbol{S} \mid R)) = E(R^2 I_d/d) = (\nu/(\nu - 2))I_d.$$

The normal mixtures with $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I_d$ represent an easily understood subgroup of the spherical distributions. There are other spherical distributions that cannot be represented as normal variance mixtures; an example is the distribution of the uniform vector $\boldsymbol{S}$ on $\mathcal{S}^{d-1}$ itself. However, the normal mixtures have a special role in the spherical world, as summarized by the following theorem.

Theorem 6.24. *Denote by Ψ_∞ the set of characteristic generators that generate a d-dimensional spherical distribution for arbitrary $d \geqslant 1$. Then $\boldsymbol{X} \sim S_d(\psi)$ with $\psi \in \Psi_\infty$ if and only if $\boldsymbol{X} \stackrel{\mathrm{d}}{=} \sqrt{W}\boldsymbol{Z}$, where $\boldsymbol{Z} \sim N_d(\mathbf{0}, I_d)$ is independent of $W \geqslant 0$.*

Proof. This is proved in Fang, Kotz and Ng (1990, pp. 48–51). □

Thus, the characteristic generators of normal mixtures generate spherical distributions in arbitrary dimensions, while other spherical generators may only be used in certain dimensions. A concrete example is given by the uniform distribution on the unit sphere. Let Ω_d denote the characteristic generator of the uniform vector $\boldsymbol{S} = (S_1, \ldots, S_d)'$ on $\mathcal{S}_{d-1}$. It can be shown that $\Omega_d((t_1, \ldots, t_{d+1})'(t_1, \ldots, t_{d+1}))$ is not the characteristic function of a spherical distribution in $\mathbb{R}^{d+1}$ (for more details see Fang, Kotz and Ng (1990, pp. 70–72)).

If a spherical distribution has a density f, then, by using the inversion formula

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^d} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\boldsymbol{t}'\boldsymbol{x}} \phi_{\boldsymbol{X}}(\boldsymbol{t})\, \mathrm{d}t_1 \cdots \mathrm{d}t_d,$$

it is easily inferred from Theorem 6.18 that $f(\boldsymbol{x}) = f(U\boldsymbol{x})$ for any orthogonal matrix U, so that the density must be of the form

$$f(\boldsymbol{x}) = g(\boldsymbol{x}'\boldsymbol{x}) = g(x_1^2 + \cdots + x_d^2) \tag{6.39}$$

for some function g of a scalar variable, which is referred to as the *density generator*. Clearly, the joint density is constant on hyperspheres $\{\boldsymbol{x} : x_1^2 + \cdots + x_d^2 = c\}$ in $\mathbb{R}^d$. To give a single example, the density generator of the multivariate t (i.e. the model $\boldsymbol{X} \sim t_d(\nu, \mathbf{0}, I_d)$ of Example 6.7) is

$$g(x) = \frac{\Gamma(\frac{1}{2}(\nu + d))}{\Gamma(\frac{1}{2}\nu)(\pi\nu)^{d/2}} \left(1 + \frac{x}{\nu}\right)^{-(\nu+d)/2}.$$

6.3.2 Elliptical Distributions

Definition 6.25. $\boldsymbol{X}$ has an elliptical distribution if

$$\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + A\boldsymbol{Y},$$

where $\boldsymbol{Y} \sim S_k(\psi)$ and $A \in \mathbb{R}^{d\times k}$ and $\boldsymbol{\mu} \in \mathbb{R}^d$ are a matrix and vector of constants, respectively.

In other words, elliptical distributions are obtained by multivariate *affine* transformations of spherical distributions. Since the characteristic function is

$$\phi_X(\boldsymbol{t}) = E(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{X}}) = E(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'(\boldsymbol{\mu}+A\boldsymbol{Y})}) = \mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu}} E(\mathrm{e}^{\mathrm{i}(A'\boldsymbol{t})'\boldsymbol{Y}}) = \mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu}}\psi(\boldsymbol{t}'\Sigma\boldsymbol{t}),$$

where $\Sigma = AA'$, we denote the elliptical distributions by

$$\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$$

and refer to $\boldsymbol{\mu}$ as the location vector, Σ as the dispersion matrix and ψ as the characteristic generator of the distribution.

Remark 6.26. Knowledge of $\boldsymbol{X}$ does not uniquely determine its elliptical representation $E_d(\boldsymbol{\mu}, \Sigma, \psi)$. Although $\boldsymbol{\mu}$ is uniquely determined, Σ and ψ are only determined up to a positive constant. For example, the multivariate normal distribution $N_d(\boldsymbol{\mu}, \Sigma)$ can be written as $E_d(\boldsymbol{\mu}, \Sigma, \psi(\cdot))$ or $E_d(\boldsymbol{\mu}, c\Sigma, \psi(\cdot/c))$ for $\psi(u) = \mathrm{e}^{-u/2}$ and any $c > 0$. Provided that variances are finite, then an elliptical distribution is fully specified by its mean vector, covariance matrix and characteristic generator, and it is possible to find an elliptical representation $E_d(\boldsymbol{\mu}, \Sigma, \psi)$ such that Σ is the covariance matrix of $\boldsymbol{X}$, although this is not always the standard representation of the distribution.

We now give an alternative stochastic representation for the elliptical distributions that follows directly from Definition 6.25 and Theorem 6.21.

Proposition 6.27. *$\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ if and only if there exist $\boldsymbol{S}$, R and A satisfying*

$$\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + RA\boldsymbol{S}, \tag{6.40}$$

with

(i) $\boldsymbol{S}$ *uniformly distributed on the unit sphere* $\mathcal{S}^{k-1} = \{\boldsymbol{s} \in \mathbb{R}^k : \boldsymbol{s}'\boldsymbol{s} = 1\}$,

(ii) $R \geqslant 0$, *a radial rv, independent of* $\boldsymbol{S}$, *and*

(iii) $A \in \mathbb{R}^{d\times k}$ *with* $AA' = \Sigma$.

For practical examples we are most interested in the case where Σ is positive definite. The relation between the elliptical and spherical cases is then clearly

$$\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi) \iff \Sigma^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}) \sim S_d(\psi). \tag{6.41}$$

In this case, if the spherical vector $\boldsymbol{Y}$ has density generator g, then $\boldsymbol{X} = \boldsymbol{\mu} + \Sigma^{1/2}\boldsymbol{Y}$ has density

$$f(\boldsymbol{x}) = \frac{1}{|\Sigma|^{1/2}} g((\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})).$$

The joint density is always constant on sets of the form $\{\boldsymbol{x} : (\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = c\}$, which are ellipsoids in $\mathbb{R}^d$. Clearly, the full family of multivariate normal variance mixtures with general location and dispersion parameters $\boldsymbol{\mu}$ and Σ are elliptical, since they are obtained by affine transformations of the spherical special cases considered in the previous section.

It follows from (6.37) and (6.41) that for a non-singular elliptical variate $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ with no point mass at $\boldsymbol{\mu}$, we have

$$\left(\sqrt{(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu})}, \frac{\Sigma^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu})}{\sqrt{(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu})}}\right) \stackrel{\mathrm{d}}{=} (R, \boldsymbol{S}), \tag{6.42}$$

where $\boldsymbol{S}$ is uniformly distributed on $\mathcal{S}^{d-1}$ and R is an independent scalar rv. This forms the basis of a test of elliptical symmetry described in Section 15.1.2.

The following proposition shows that a particular conditional distribution of an elliptically distributed random vector $\boldsymbol{X}$ has the same correlation matrix as $\boldsymbol{X}$ and can also be used to test for elliptical symmetry.

Proposition 6.28. *Let* $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ *and assume that* Σ *is positive definite and* $\mathrm{cov}(\boldsymbol{X})$ *is finite. For any* $c \geqslant 0$ *such that* $P((\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \geqslant c) > 0$, *we have*

$$\rho(\boldsymbol{X} \mid (\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \geqslant c) = \rho(\boldsymbol{X}). \tag{6.43}$$

Proof. It follows easily from (6.42) that

$$\boldsymbol{X} \mid (\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \geqslant c \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + R\Sigma^{1/2}\boldsymbol{S} \mid R^2 \geqslant c,$$

where $R \stackrel{\mathrm{d}}{=} \sqrt{(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu})}$ and $\boldsymbol{S}$ is independent of R and uniformly distributed on $\mathcal{S}^{d-1}$. Thus we have

$$\boldsymbol{X} \mid (\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \geqslant c \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \tilde{R}\Sigma^{1/2}\boldsymbol{S},$$

where $\tilde{R} \stackrel{\mathrm{d}}{=} R \mid R^2 \geqslant c$. It follows from Proposition 6.27 that the conditional distribution remains elliptical with dispersion matrix Σ and that (6.43) holds. □

6.3.3 Properties of Elliptical Distributions

We now summarize some of the properties of elliptical distributions in a format that allows their comparison with the properties of multivariate normal distributions in Section 6.1.3. Many properties carry over directly and others only need to be modified slightly. These parallels emphasize that it would be fairly easy to base many standard procedures in risk management on an assumption that risk-factor changes have an approximately elliptical distribution, rather than the patently false assumption that they are multivariate normal.

Linear combinations. If we take linear combinations of elliptical random vectors, then these remain elliptical with the same characteristic generator ψ. Let $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ and take any $B \in \mathbb{R}^{k\times d}$ and $\boldsymbol{b} \in \mathbb{R}^k$. Using a similar argument to that in Proposition 6.9 it is then easily shown that

$$B\boldsymbol{X} + \boldsymbol{b} \sim E_k(B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B', \psi). \tag{6.44}$$

As a special case, if $\boldsymbol{a} \in \mathbb{R}^d$, then

$$\boldsymbol{a}'\boldsymbol{X} \sim E_1(\boldsymbol{a}'\boldsymbol{\mu}, \boldsymbol{a}'\Sigma\boldsymbol{a}, \psi). \tag{6.45}$$

Marginal distributions. It follows from (6.45) that marginal distributions of $\boldsymbol{X}$ must be elliptical distributions with the same characteristic generator. Using the $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2')'$ notation from Section 6.1.3 and again extending this notation naturally to $\boldsymbol{\mu}$ and Σ,

$$\boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}, \qquad \Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix},$$

we have that $\boldsymbol{X}_1 \sim E_k(\boldsymbol{\mu}_1, \Sigma_{11}, \psi)$ and $\boldsymbol{X}_2 \sim E_{d-k}(\boldsymbol{\mu}_2, \Sigma_{22}, \psi)$.

Conditional distributions. The conditional distribution of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1$ may also be shown to be elliptical, although in general it will have a *different* characteristic generator $\tilde{\psi}$. For details of how the generator changes see Fang, Kotz and Ng (1990, pp. 45, 46). In the special case of multivariate normality the generator remains the same.

Quadratic forms. If $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ with Σ non-singular, then we observed in (6.42) that

$$Q := (\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \overset{\mathrm{d}}{=} R^2, \tag{6.46}$$

where R is the radial rv in the stochastic representation (6.40). As we have seen in Example 6.23, for some particular cases the distribution of R^2 is well known: if $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$, then $R^2 \sim \chi^2_d$; if $\boldsymbol{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$, then $R^2/d \sim F(d, \nu)$. For all elliptical distributions, Q must be independent of $\Sigma^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu})/\sqrt{Q}$.

Convolutions. The convolution of two independent elliptical vectors with the *same dispersion matrix* Σ is also elliptical. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent d-dimensional random vectors satisfying $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ and $\boldsymbol{Y} \sim E_d(\tilde{\boldsymbol{\mu}}, \Sigma, \tilde{\psi})$, then we may take the product of characteristic functions to show that

$$\boldsymbol{X} + \boldsymbol{Y} \sim E_d(\boldsymbol{\mu} + \tilde{\boldsymbol{\mu}}, \Sigma, \bar{\psi}), \tag{6.47}$$

where $\bar{\psi}(u) = \psi(u)\tilde{\psi}(u)$.

If the dispersion matrices of $\boldsymbol{X}$ and $\boldsymbol{Y}$ differ by more than a constant factor, then the convolution will not necessarily remain elliptical, even when the two generators ψ and $\tilde{\psi}$ are identical.

6.3.4 Estimating Dispersion and Correlation

Suppose we have risk-factor return data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ that we believe come from some elliptical distribution $E_d(\boldsymbol{\mu}, \Sigma, \psi)$ with heavier tails than the multivariate normal. We recall from Remark 6.26 that the dispersion matrix Σ is not uniquely determined, but rather is only fixed up to a constant of proportionality; when covariances are finite, the covariance matrix is proportional to Σ.

In this section we briefly consider the problem of estimating the location parameter $\boldsymbol{\mu}$, a dispersion matrix Σ and the correlation matrix P, assuming finiteness of second moments. We could use the standard estimators of Section 6.1.2. Under an assumption of iid or uncorrelated vector observations we observed that $\bar{\boldsymbol{X}}$ and S in (6.9) are unbiased estimators of the mean vector and the covariance matrix, respectively. They will also be consistent under quite weak assumptions. However, this does not necessarily mean they are the best estimators of location and dispersion for any given finite sample of elliptical data. There are many alternative estimators that may be more efficient for heavy-tailed data and may enjoy better robustness properties for contaminated data.

One strategy would be to fit a number of normal variance mixture models, such as the t and NIG, using the approach of Section 6.2.4. From the best-fitting model we would obtain an estimate of the mean vector and could easily calculate the implied estimates of the covariance and correlation matrices. In this section we give simpler, alternative methods that do not require a full fitting of a multivariate distribution; consult Notes and Comments for further references to robust dispersion estimation.

M-estimators. Maronna's M-estimators (Maronna 1976) of location and dispersion are a relatively old idea in robust statistics, but they have the virtue of being particularly simple to implement. Let $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ denote estimates of the mean vector and the dispersion matrix. Suppose for every observation $\boldsymbol{X}_i$ we calculate $D_i^2 = (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}})' \hat{\Sigma}^{-1} (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}})$. If we wanted to calculate improved estimates of location and dispersion, particularly for heavy-tailed data, it might be expected that this could be achieved by reducing the influence of observations for which D_i is large, since these are the observations that might tend to distort the parameter estimates most. M-estimation uses decreasing weight functions $w_j : \mathbb{R}^+ \to \mathbb{R}^+$, $j = 1, 2$, to reduce the weight of observations with large D_i values. This can be turned into an iterative procedure that converges to so-called M-estimates of location and dispersion; the dispersion matrix estimate is generally a biased estimate of the true covariance matrix.

Algorithm 6.29 (M-estimators of location and dispersion).

(1) As starting estimates take $\hat{\boldsymbol{\mu}}^{[1]} = \bar{\boldsymbol{X}}$ and $\hat{\Sigma}^{[1]} = S$, the standard estimators in (6.9). Set iteration count $k = 1$.

(2) For $i = 1, \ldots, n$ set $D_i^2 = (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}^{[k]})' \hat{\Sigma}^{[k]\,-1} (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}^{[k]})$.

(3) Update the location estimate using

$$\hat{\boldsymbol{\mu}}^{[k+1]} = \frac{\sum_{i=1}^{n} w_1(D_i)\boldsymbol{X}_i}{\sum_{i=1}^{n} w_1(D_i)},$$

where w_1 is a weight function, as discussed below.

(4) Update the dispersion matrix estimate using

$$\hat{\Sigma}^{[k+1]} = \frac{1}{n}\sum_{i=1}^{n} w_2(D_i^2)(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}^{[k]})(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}^{[k]})',$$

where w_2 is a weight function.

(5) Set $k = k + 1$ and repeat steps (2)–(4) until estimates converge.

Popular choices for the weight functions w_1 and w_2 are the decreasing functions $w_1(x) = (d + \nu)/(x^2 + \nu) = w_2(x^2)$ for some positive constant ν. Interestingly, use of these weight functions in Algorithm 6.29 exactly corresponds to fitting a multivariate $t_d(\nu, \boldsymbol{\mu}, \Sigma)$ distribution with known degrees of freedom ν using the EM algorithm (see, for example, Meng and van Dyk 1997).

There are many other possibilities for the weight functions. For example, the observations in the central part of the distribution could be given full weight and only the more outlying observations downweighted. This can be achieved by setting $w_1(x) = 1$ for $x \leqslant a$, $w_1(x) = a/x$ for $x > a$, for some value a, and $w_2(x^2) = (w_1(x))^2$.

Correlation estimates via Kendall's tau. A method for estimating correlation that is particularly easy to carry out is based on Kendall's rank correlation coefficient; this method will turn out to be related to a method in Chapter 7 that is used for estimating the parameters of certain copulas. The theoretical version of Kendall's rank correlation (also known as Kendall's tau) for two rvs X_1 and X_2 is denoted by $\rho_\tau(X_1, X_2)$ and is defined formally in Section 7.2.3; it is shown in Proposition 7.43 that if $(X_1, X_2) \sim E_2(\boldsymbol{\mu}, \Sigma, \psi)$, then

$$\rho_\tau(X_1, X_2) = \frac{2}{\pi}\arcsin(\rho), \tag{6.48}$$

where $\rho = \sigma_{12}/(\sigma_{11}\sigma_{22})^{1/2}$ is the *pseudo-correlation coefficient* of the elliptical distribution, which is always defined (even when correlation coefficients are undefined because variances are infinite). This relationship can be inverted to provide a method for estimating ρ from data; we simply replace the left-hand side of (6.48) by the standard textbook estimator of Kendall's tau, which is given in (7.52), to get an estimating equation that is solved for $\hat{\rho}$. This method estimates correlation by exploiting the geometry of an elliptical distribution and does not require us to estimate variances and covariances.

The method can be used to estimate a correlation matrix of a higher-dimensional elliptical distribution by applying the technique to each bivariate margin. This does, however, result in a matrix of pairwise correlation estimates that is not necessarily positive definite; this problem does not always arise, and if it does, a matrix

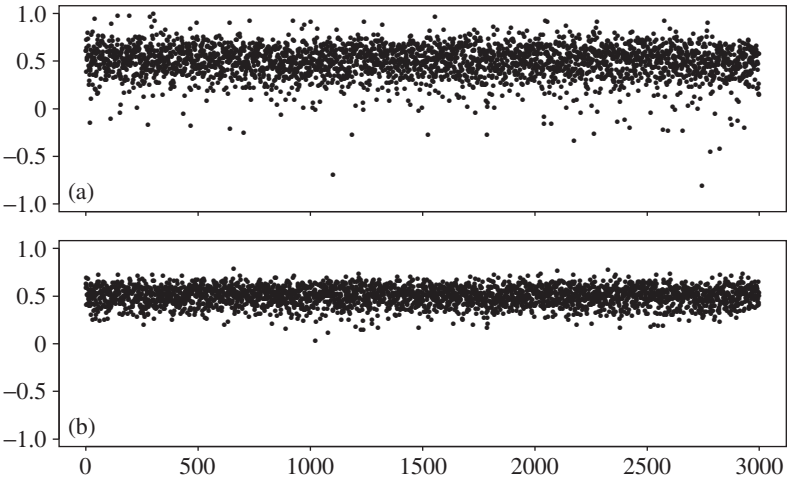

Figure 6.4. For 3000 independent samples of size 90 from a bivariate t distribution with three degrees of freedom and linear correlation 0.5: (a) the standard (Pearson) estimator of correlation; (b) the Kendall's tau transform estimator. See Example 6.30 for commentary.

adjustment method can be used, such as the eigenvalue method of Rousseeuw and Molenberghs (1993), which is given in Algorithm 7.57.

Note that to turn an estimate of a bivariate correlation matrix into a robust estimate of a dispersion matrix we could estimate the ratio of standard deviations $\lambda = (\sigma_{22}/\sigma_{11})^{1/2}$, e.g. by using a ratio of *trimmed* sample standard deviations; in other words, we leave out an equal number of outliers from each of the univariate data sets $X_{1,i}, \ldots, X_{n,i}$ for $i = 1, 2$ and calculate the sample standard deviations with the remaining observations. This would give us the estimate

$$\hat{\Sigma} = \begin{pmatrix} 1 & \hat{\lambda}\hat{\rho} \\ \hat{\lambda}\hat{\rho} & \hat{\lambda}^2 \end{pmatrix}. \tag{6.49}$$

Example 6.30 (efficient correlation estimation for heavy-tailed data). Suppose we calculate correlations of asset or risk-factor returns based on 90 days (somewhat more than three trading months) of data; it would seem that this ought to be enough data to allow us to accurately estimate the "true" underlying correlation under an assumption that we have identically distributed data for that period.

Figure 6.4 displays the results of a simulation experiment where we have generated 3000 bivariate samples of iid data from a t distribution with three degrees of freedom and correlation $\rho = 0.5$; this is a heavy-tailed elliptical distribution. The distribution of the values of the standard correlation coefficient (also known as the Pearson correlation coefficient) is not particularly closely concentrated around the true value and produces some very poor estimates for a number of samples. On the other hand, the Kendall's tau transform method produces estimates that are generally much closer to the true value, and thus provides a more efficient way of estimating ρ.

Notes and Comments

A comprehensive reference for spherical and elliptical distributions is Fang, Kotz and Ng (1990); we have based our brief presentation of the theory on this account. Other references for the theory are Kelker (1970), Cambanis, Huang and Simons (1981) and Bingham and Kiesel (2002), the latter in the context of financial modelling. The original reference for Theorem 6.21 is Schoenberg (1938). Frahm (2004) suggests a generalization of the elliptical class to allow asymmetric models while preserving many of the attractive properties of elliptical distributions. For a more historical discussion (going back to Archimedes) and some surprising properties of the uniform distribution on the unit d-sphere, see Letac (2004).

There is a vast literature on alternative estimators of dispersion and correlation matrices, particularly with regard to better robustness properties. Textbooks with relevant sections include Hampel et al. (1986), Marazzi (1993), Wilcox (1997) and Huber and Ronchetti (2009); the last of those books is recommended more generally for applications of robust statistics in econometrics and finance.

We have concentrated on M-estimation of dispersion matrices, since this is related to the maximum likelihood estimation of alternative elliptical models. M-estimators have a relatively long history and are known to have good local robustness properties (insensitivity to small data perturbations); they do, however, have relatively low breakdown points in high dimensions, so their performance can be poor when data are more contaminated. A small selection of papers on M-estimation is Maronna (1976), Devlin, Gnanadesikan and Kettenring (1975, 1981) and Tyler (1983, 1987); see also Frahm (2004), in which an interesting alternative derivation of a Tyler estimator is given. The method based on Kendall's tau was suggested in Lindskog, McNeil and Schmock (2003).

6.4 Dimension-Reduction Techniques

The techniques of dimension reduction, such as factor modelling and principal components, are central to multivariate statistical analysis and are widely used in econometric model building. In the high-dimensional world of financial risk management they are essential tools.

6.4.1 Factor Models

By using a factor model we attempt to explain the randomness in the components of a d-dimensional vector $\boldsymbol{X}$ in terms of a smaller set of *common factors*. If the components of $\boldsymbol{X}$ represent, for example, equity returns, it is clear that a large part of their variation can be explained in terms of the variation of a smaller set of market index returns. Formally, we define a factor model as follows.

Definition 6.31 (linear factor model). The random vector $\boldsymbol{X}$ is said to follow a p-factor model if it can be decomposed as

$$\boldsymbol{X} = \boldsymbol{a} + B\boldsymbol{F} + \boldsymbol{\varepsilon}, \tag{6.50}$$

where

(i) $\boldsymbol{F} = (F_1, \ldots, F_p)'$ is a random vector of *common factors* with $p < d$ and a covariance matrix that is positive definite,

(ii) $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_d)'$ is a random vector of *idiosyncratic error terms*, which are uncorrelated and have mean 0,

(iii) $B \in \mathbb{R}^{d\times p}$ is a matrix of constant *factor loadings* and $\boldsymbol{a} \in \mathbb{R}^d$ is a vector of constants, and

(iv) $\operatorname{cov}(\boldsymbol{F}, \boldsymbol{\varepsilon}) = E((\boldsymbol{F} - E(\boldsymbol{F}))\boldsymbol{\varepsilon}') = 0$.

The assumptions that the errors are uncorrelated with each other (ii) and also with the common factors (iv) are important parts of this definition. We do not in general require independence, only uncorrelatedness. However, if the vector $\boldsymbol{X}$ is multivariate normally distributed and follows the factor model in (6.50), then it is possible to find a version of the factor model where $\boldsymbol{F}$ and $\boldsymbol{\varepsilon}$ are Gaussian and the errors can be assumed to be mutually independent and independent of the common factors. We elaborate on this assertion in Example 6.32 below.

It follows from the basic assumptions that factor models imply a special structure for the covariance matrix Σ of $\boldsymbol{X}$. If we denote the covariance matrix of $\boldsymbol{F}$ by Ω and that of $\boldsymbol{\varepsilon}$ by the diagonal matrix Υ, it follows that

$$\Sigma = \operatorname{cov}(\boldsymbol{X}) = B\Omega B' + \Upsilon. \tag{6.51}$$

If the factor model holds, the common factors can always be transformed so that they have mean 0 and are orthogonal. By setting $\boldsymbol{F}^* = \Omega^{-1/2}(\boldsymbol{F} - E(\boldsymbol{F}))$ and $B^* = B\Omega^{1/2}$, we have a representation of the factor model of the form $\boldsymbol{X} = \boldsymbol{\mu} + B^*\boldsymbol{F}^* + \boldsymbol{\varepsilon}$, where $\boldsymbol{\mu} = E(\boldsymbol{X})$, as usual, and $\Sigma = B^*(B^*)' + \Upsilon$.

Conversely, it can be shown that whenever a random vector $\boldsymbol{X}$ has a covariance matrix that satisfies

$$\Sigma = BB' + \Upsilon \tag{6.52}$$

for some $B \in \mathbb{R}^{d\times p}$ with $\operatorname{rank}(B) = p < d$ and diagonal matrix Υ, then $\boldsymbol{X}$ has a factor-model representation for some p-dimensional factor vector $\boldsymbol{F}$ and d-dimensional error vector $\boldsymbol{\varepsilon}$.

Example 6.32 (equicorrelation model). Suppose $\boldsymbol{X}$ is a random vector with standardized margins (zero mean and unit variance) and an *equicorrelation matrix*; in other words, the correlation between each pair of components is equal to $\rho > 0$. This means that the covariance matrix Σ can be written as $\Sigma = \rho J_d + (1-\rho)I_d$, where J_d is the d-dimensional square matrix of ones and I_d is the identity matrix, so that Σ is obviously of the form (6.52) for the d-vector $B = \sqrt{\rho}\mathbf{1}$.

To find a factor decomposition of $\boldsymbol{X}$, take *any* zero-mean, unit-variance rv Y that is *independent* of $\boldsymbol{X}$ and define a single common factor F and errors $\boldsymbol{\varepsilon}$ by

$$F = \frac{\sqrt{\rho}}{1+\rho(d-1)}\sum_{j=1}^{d} X_j + \sqrt{\frac{1-\rho}{1+\rho(d-1)}}Y, \qquad \varepsilon_j = X_j - \sqrt{\rho}F,$$

where we note that in this construction F also has mean 0 and variance 1. We therefore have the factor decomposition $\boldsymbol{X} = B\boldsymbol{F} + \boldsymbol{\varepsilon}$, and it may be verified by calculation that $\text{cov}(F, \varepsilon_j) = 0$ for all j and $\text{cov}(\varepsilon_j, \varepsilon_k) = 0$ when $j \neq k$, so that the requirements of Definition 6.31 are satisfied. A random vector with an equicorrelation matrix can be thought of as following a factor model with a single common factor.

Since we can take any Y, the factors and errors in this decomposition are non-unique. Consider the case where the vector $\boldsymbol{X}$ is Gaussian; it is most convenient to take Y to also be Gaussian, since in that case the common factor is normally distributed, the error vector is multivariate normally distributed, Y is independent of ε_j, for all j, and ε_j and ε_k are independent for $j \neq k$. Since $\text{var}(\varepsilon_j) = 1 - \rho$, it is most convenient to write the factor model implied by the equicorrelation model as

$$X_j = \sqrt{\rho}F + \sqrt{1-\rho}Z_j, \quad j = 1, \ldots, d, \tag{6.53}$$

where $F, Z_1, \ldots, Z_d$ are mutually independent standard Gaussian rvs. This model will be used in Section 11.1.5 in the context of modelling homogeneous credit portfolios. For the more general construction on which this example is based, see Mardia, Kent and Bibby (1979, Exercise 9.2.2).

6.4.2 Statistical Estimation Strategies

Now assume that we have data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \mathbb{R}^d$ representing risk-factor changes at times $t = 1, \ldots, n$. Each vector observation $\boldsymbol{X}_t$ is assumed to be a realization from a factor model of the form (6.50) so that we have

$$\boldsymbol{X}_t = \boldsymbol{a} + B\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, n, \tag{6.54}$$

for common-factor vectors $\boldsymbol{F}_t = (F_{t,1}, \ldots, F_{t,p})'$, error vectors $\boldsymbol{\varepsilon}_t$, a vector of constants $\boldsymbol{a} \in \mathbb{R}^d$, and loading matrix $B \in \mathbb{R}^{d \times p}$. There are occasionally situations where we might wish to model $\boldsymbol{a}$ and B as time dependent, but mostly they are assumed to be fixed over time.

The model (6.54) is clearly an idealization. Data will seldom be perfectly explained by a factor model; the aim is to find an approximating factor model that captures the main sources of variability in the data. Three general types of factor model are commonly used in financial risk applications; they are known as *macroeconomic*, *fundamental* and *statistical* factor models.

Macroeconomic factor models. In these models we assume that appropriate factors $\boldsymbol{F}_t$ are also observable and we collect time-series data $\boldsymbol{F}_1, \ldots, \boldsymbol{F}_n \in \mathbb{R}^p$. The name comes from the fact that, in many applications of these models in economics and finance, the observed factors are macroeconomic variables, such as changes in GDP, inflation and interest rates.

A simple example of a macroeconomic model in finance is Sharpe's single-index model, where $F_1, \ldots, F_n$ are observations of the return on a market index and $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are individual equity returns that are explained in terms of the market return. Fitting of the model (estimation of B and $\boldsymbol{a}$) is accomplished by time-series regression techniques; it is described in Section 6.4.3.

Fundamental factor models. In contrast to the macroeconomic factor models, here we assume that the loading matrix B is known but that the underlying factors $\boldsymbol{F}_t$ are unobserved. Factor values $\boldsymbol{F}_1, \ldots, \boldsymbol{F}_n$ have to be estimated from the data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ using cross-sectional regression at each time point.

The name comes from applications in modelling equity returns where the stocks are classified according to their "fundamentals", such as country, industry sector and size (small cap, large cap, etc.). These are generally categorical variables and it is assumed that there are underlying, unobserved factors associated with each level of the categorical variable, e.g. a factor for each country or each industry sector.

If each risk-factor change $X_{t,i}$ can be identified with a unique set of values for the fundamentals, e.g. a unique country or industry, then the matrix B is a matrix consisting of zeros and ones. If $X_{t,i}$ is attributed to different values of the fundamental variable, then B might contain factor weights summing to 1; for example, 60% of a stock return for a multinational company might be attributed to an unobserved US factor and 40% to an unobserved UK factor. There may also be situations in fundamental factor modelling where time-dependent loading matrices B_t are used.

Statistical factor models. In these models we observe neither the factors $\boldsymbol{F}_t$ nor the loadings B. Instead, we use statistical techniques to estimate both from the data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. This can be a very powerful approach to explaining the variability in data, but we note that the factors we obtain, while being explanatory in a statistical sense, may not have any obvious interpretation.

There are two general methods for finding factors. The first method, which is quite common in finance, is to use *principal component analysis* to construct factors; we discuss this technique in detail in Section 6.4.5. The second method, *classical statistical factor analysis*, is less commonly used in finance (see Notes and Comments).

Factor models and systematic risk. In the context of risk management, the goal of all approaches to factor modelling is either to identify or to estimate appropriate factor data $\boldsymbol{F}_1, \ldots, \boldsymbol{F}_n$. If this is achieved, we can then concentrate on modelling the distribution or dynamics of the factors, which is a lower-dimensional problem than modelling $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$.

The factors describe the *systematic risk* and are of primary importance. The unobserved errors $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ describe the *idiosyncratic risk* and are of secondary importance. In situations where we have many risk factors, the risk embodied in the errors is partly mitigated by a diversification effect, whereas the risk embodied in the common factors remains. The following simple example gives an idea why this is the case.

Example 6.33. We continue our analysis of the one-factor model in Example 6.32. Suppose that the random vector $\boldsymbol{X}$ in that example represents the return on d different companies so that the rv $Z_{(d)} = (1/d)\sum_{j=1}^{d} X_j$ can be thought of as the portfolio return for an equal investment in each of the companies. We calculate that

$$Z_{(d)} = \frac{1}{d}\mathbf{1}'BF + \frac{1}{d}\mathbf{1}'\boldsymbol{\varepsilon} = \sqrt{\rho}F + \frac{1}{d}\sum_{j=1}^{d}\varepsilon_j.$$

The risk in the first term is not affected by increasing the size of the portfolio d, whereas the risk in the second term can be reduced. Suppose we measure risk by simply calculating variances; we get

$$\operatorname{var}(Z_{(d)}) = \rho + \frac{1-\rho}{d} \to \rho, \quad d \to \infty,$$

showing that the systematic factor is the main contributor to the risk in a large-portfolio situation.

6.4.3 Estimating Macroeconomic Factor Models

Two equivalent approaches may be used to estimate the model parameters in a macroeconomic factor model of the form (6.54). In the first approach we perform d univariate regression analyses, one for each component of the individual return series. In the second approach we estimate all parameters in a single multivariate regression.

Univariate regression. Writing $X_{t,j}$ for the observation at time t of instrument j, we consider the univariate regression model

$$X_{t,j} = a_j + \boldsymbol{b}_j' \boldsymbol{F}_t + \varepsilon_{t,j}, \quad t = 1, \ldots, n.$$

This is known as a time-series regression, since the responses $X_{1,j}, \ldots, X_{n,j}$ form a univariate time series and the factors $\boldsymbol{F}_1, \ldots, \boldsymbol{F}_n$ form a possibly multivariate time series. Without going into technical details we simply remark that the parameters a_j and $\boldsymbol{b}_j$ are estimated using the standard ordinary least-squares (OLS) method found in all textbooks on linear regression. To justify the use of the method and to derive statistical properties of the method it is usually assumed that, conditional on the factors, the errors $\varepsilon_{1,j}, \ldots, \varepsilon_{n,j}$ are identically distributed and serially uncorrelated. In other words, they form a white noise process as defined in Chapter 4.

The estimate $\hat{a}_j$ obviously estimates the jth component of $\boldsymbol{a}$, while $\hat{\boldsymbol{b}}_j$ is an estimate of the jth row of the matrix B. By performing a regression for each of the univariate time series $X_{1,j}, \ldots, X_{n,j}$ for $j = 1, \ldots, d$, we complete the estimation of the parameters $\boldsymbol{a}$ and B.

Multivariate regression. To set the problem up as a multivariate linear-regression problem, we construct a number of large matrices:

$$X = \underbrace{\begin{pmatrix} \boldsymbol{X}_1' \\ \vdots \\ \boldsymbol{X}_n' \end{pmatrix}}_{n \times d}, \qquad F = \underbrace{\begin{pmatrix} 1 & \boldsymbol{F}_1' \\ \vdots & \vdots \\ 1 & \boldsymbol{F}_n' \end{pmatrix}}_{n \times (p+1)}, \qquad B_2 = \underbrace{\begin{pmatrix} \boldsymbol{a}' \\ B' \end{pmatrix}}_{(p+1) \times d}, \qquad E = \underbrace{\begin{pmatrix} \boldsymbol{\varepsilon}_1' \\ \vdots \\ \boldsymbol{\varepsilon}_n' \end{pmatrix}}_{n \times d}.$$

Each row of the data X corresponds to a vector observation at a fixed time point t, and each column corresponds to a univariate time series for one of the individual returns. The model (6.54) can then be expressed by the matrix equation

$$X = FB_2 + E, \tag{6.55}$$

where B_2 is the matrix of regression parameters to be estimated.

If we assume that the unobserved error vectors $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ comprising the rows of E are identically distributed and serially uncorrelated, conditional on $\boldsymbol{F}_1, \ldots, \boldsymbol{F}_n$, then the equation (6.55) defines a standard multivariate linear regression (see, for example, Mardia, Kent and Bibby (1979) for the standard assumptions). An estimate of B_2 is obtained by multivariate OLS according to the formula

$$\hat{B}_2 = (F'F)^{-1}F'X. \tag{6.56}$$

The factor model is now essentially calibrated, since we have estimates for $\boldsymbol{a}$ and B. The model can now be critically examined with respect to the original conditions of Definition 6.31. Do the error vectors $\boldsymbol{\varepsilon}_t$ come from a distribution with diagonal covariance matrix, and are they uncorrelated with the factors?

To learn something about the errors we can form the model residual matrix $\hat{E} = X - F\hat{B}_2$. Each row of this matrix contains an inferred value of an error vector $\hat{\boldsymbol{\varepsilon}}_t$ at a fixed point in time. Examination of the sample correlation matrix of these inferred error vectors will hopefully show that there is little remaining correlation in the errors (or at least much less than in the original data vectors $\boldsymbol{X}_t$). If this is the case, then the diagonal elements of the sample covariance matrix of the $\hat{\boldsymbol{\varepsilon}}_t$ could be taken as an estimator $\hat{\Upsilon}$ for Υ. It is sometimes of interest to form the covariance matrix implied by the factor model and compare this with the original sample covariance matrix S of the data. The implied covariance matrix is

$$\hat{\Sigma}^{(\mathrm{F})} = \hat{B}\hat{\Omega}\hat{B}' + \hat{\Upsilon}, \quad \text{where } \hat{\Omega} = \frac{1}{n-1}\sum_{t=1}^{n}(\boldsymbol{F}_t - \bar{\boldsymbol{F}})(\boldsymbol{F}_t - \bar{\boldsymbol{F}})'.$$

We would hope that $\hat{\Sigma}^{(\mathrm{F})}$ captures much of the structure of S and that the correlation matrix $R^{(\mathrm{F})} := \wp(\hat{\Sigma}^{(\mathrm{F})})$ captures much of the structure of the sample correlation matrix $R = \wp(S)$.

Example 6.34 (single-index model for Dow Jones 30 returns). As a simple example of the regression approach to fitting factor models we have fitted a single factor model to a set of ten Dow Jones 30 daily stock-return series from 1992 to 1998. Note that these are different returns to those analysed in previous sections of this chapter. They have been chosen to be of two types: technology-related titles such as Hewlett-Packard, Intel, Microsoft and IBM; and food- and consumer-related titles such as Philip Morris, Coca-Cola, Eastman Kodak, McDonald's, Wal-Mart and Disney. The factor chosen is the corresponding return on the Dow Jones 30 index itself.

The estimate of B implied by formula (6.56) is shown in the first line of Table 6.5. The highest values of B correspond to so-called *high-beta* stocks; since a one-factor model implies the relationship $E(X_j) = a_j + B_j E(F)$, these stocks potentially offer high expected returns relative to the market (but are often riskier titles); in this case, the four technology-related stocks have the highest beta values. In the second row, values of r^2, the so-called coefficient of determination, are given for each of the univariate regression models. This number measures the strength of the regression relationship between X_j and F and can be interpreted as the proportion of the variation of the stock return that is explained by variation in the market return;

Table 6.5. The first line gives estimates of B for a multivariate regression model fitted to ten Dow Jones 30 stocks where the observed common factor is the return on the Dow Jones 30 index itself. The second row gives r^2 values for a univariate regression model for each individual time series. The next ten lines of the table give the sample correlation matrix of the data R, while the middle ten lines give the correlation matrix implied by the factor model. The final ten lines show the estimated correlation matrix of the residuals from the regression model, with entries less than 0.1 in absolute value being omitted. See Example 6.34 for full details.

	MO	KO	EK	HWP	INTC	MSFT	IBM	MCD	WMT	DIS
$\hat{B}$	0.87	1.01	0.77	1.12	1.12	1.11	1.07	0.86	1.02	1.03
r^2	0.17	0.33	0.14	0.18	0.17	0.21	0.22	0.23	0.24	0.26
MO	1.00	0.27	0.14	0.17	0.16	0.25	0.18	0.22	0.16	0.22
KO	0.27	1.00	0.17	0.22	0.21	0.25	0.18	0.36	0.33	0.32
EK	0.14	0.17	1.00	0.17	0.17	0.18	0.15	0.14	0.17	0.16
HWP	0.17	0.22	0.17	1.00	0.42	0.38	0.36	0.20	0.22	0.23
INTC	0.16	0.21	0.17	0.42	1.00	0.53	0.36	0.19	0.22	0.21
MSFT	0.25	0.25	0.18	0.38	0.53	1.00	0.33	0.22	0.28	0.26
IBM	0.18	0.18	0.15	0.36	0.36	0.33	1.00	0.20	0.20	0.20
MCD	0.22	0.36	0.14	0.20	0.19	0.22	0.20	1.00	0.26	0.26
WMT	0.16	0.33	0.17	0.22	0.22	0.28	0.20	0.26	1.00	0.28
DIS	0.22	0.32	0.16	0.23	0.21	0.26	0.20	0.26	0.28	1.00
MO	1.00	0.24	0.16	0.18	0.17	0.19	0.20	0.20	0.20	0.21
KO	0.24	1.00	0.22	0.24	0.23	0.26	0.27	0.28	0.28	0.29
EK	0.16	0.22	1.00	0.16	0.15	0.17	0.18	0.18	0.18	0.19
HWP	0.18	0.24	0.16	1.00	0.17	0.19	0.20	0.20	0.21	0.22
INTC	0.17	0.23	0.15	0.17	1.00	0.19	0.19	0.19	0.20	0.21
MSFT	0.19	0.26	0.17	0.19	0.19	1.00	0.22	0.22	0.22	0.23
IBM	0.20	0.27	0.18	0.20	0.19	0.22	1.00	0.23	0.23	0.24
MCD	0.20	0.28	0.18	0.20	0.19	0.22	0.23	1.00	0.23	0.24
WMT	0.20	0.28	0.18	0.21	0.20	0.22	0.23	0.23	1.00	0.25
DIS	0.21	0.29	0.19	0.22	0.21	0.23	0.24	0.24	0.25	1.00
MO	1.00									
KO		1.00					−0.12	0.12		
EK			1.00							
HWP				1.00	0.30	0.24	0.20			
INTC				0.30	1.00	0.43	0.20			
MSFT				0.24	0.43	1.00	0.14			
IBM		−0.12		0.20	0.20	0.14	1.00			
MCD		0.12						1.00		
WMT										
DIS										1.00

the highest r^2 corresponds to Coca-Cola (33%), and in general it seems that about 20% of individual stock-return variation is explained by market-return variation.

The next ten lines of the table give the sample correlation matrix of the data R, while the middle ten lines give the correlation matrix implied by the factor model

(corresponding to $\hat{\Sigma}^{(\mathrm{F})}$). The latter matrix picks up much, but not all, of the structure of the former matrix. The final ten lines show the estimated correlation matrix of the residuals from the regression model, but only those elements that exceed 0.1 in absolute value. The residuals are indeed much less correlated than the original data, but a few larger entries indicate imperfections in the factor-model representation of the data, particularly for the technology stocks. The index return for the broader market is clearly an important common factor, but further systematic effects that are not captured by the index appear to be present in these data.

6.4.4 Estimating Fundamental Factor Models

To estimate a fundamental factor model we consider, at each time point t, a cross-sectional regression model of the form

$$\boldsymbol{X}_t = B\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \tag{6.57}$$

where $\boldsymbol{X}_t \in \mathbb{R}^d$ are the risk-factor change data, $B \in \mathbb{R}^{d\times p}$ is a known matrix of factor loadings (which may be time dependent in some applications), $\boldsymbol{F}_t \in \mathbb{R}^p$ are the factors to be estimated, and $\boldsymbol{\varepsilon}_t$ are errors with diagonal covariance matrix Υ. There is no need for an intercept $\boldsymbol{a}$ in the estimation of a fundamental factor model, as this can be absorbed into the factor estimates.

To obtain precision in the estimation of $\boldsymbol{F}_t$, the dimension d of the risk-factor vector needs to be large with respect to the number of factors p to be estimated. Note also that the components of the error vector $\boldsymbol{\varepsilon}_t$ cannot generally be assumed to have equal variance, so (6.57) is a regression problem with so-called heteroscedastic errors.

We recall that, in typical applications in equity return modelling, the factors are frequently identified with country, industry-sector and company-size effects. The rows of the matrix B can consist of zeros and ones, if $X_{t,i}$ is associated with a single country or industry sector, or weights, if $X_{t,i}$ is attributed to more than one country or industry sector. This kind of interpretation for the factors is also quite common in the factor models used for modelling portfolio credit risk, as we discuss in Section 11.5.1.

Unbiased estimators of the factors $\boldsymbol{F}_t$ may be obtained by forming the OLS estimates

$$\hat{\boldsymbol{F}}_t^{\mathrm{OLS}} = (B'B)^{-1}B'\boldsymbol{X}_t,$$

and these are the best linear unbiased estimates in the case where the errors are homoscedastic, so that $\Upsilon = \upsilon^2 I_d$ for some scalar υ. However, in general, the OLS estimates are not efficient and it is possible to obtain linear unbiased estimates with a smaller covariance matrix using the method of generalized least squares (GLS). If Υ were a known matrix, the GLS estimates would be given by

$$\hat{\boldsymbol{F}}_t^{\mathrm{GLS}} = (B'\Upsilon^{-1}B)^{-1}B'\Upsilon^{-1}\boldsymbol{X}_t. \tag{6.58}$$

In practice, we replace Υ in (6.58) with an estimate $\hat{\Upsilon}$ obtained as follows. Under an assumption that the model (6.57) holds at every time point $t = 1, \ldots, n$, we first

carry out OLS estimation at each time t and form the model residual vectors

$$\hat{\boldsymbol{\varepsilon}}_t = \boldsymbol{X}_t - B\hat{\boldsymbol{F}}_t^{\mathrm{OLS}}, \quad t = 1, \ldots, n.$$

We then form the sample covariance matrix of the residuals $\hat{\boldsymbol{\varepsilon}}_1, \ldots, \hat{\boldsymbol{\varepsilon}}_n$. This matrix should be approximately diagonal, if the factor model assumption holds. We can set off-diagonal elements equal to zero to form an estimate of Υ.

We give an example of the estimation of a fundamental factor model in the context of modelling the yield curve in Section 9.1.4.

6.4.5 Principal Component Analysis

The aim of principal component analysis (PCA) is to reduce the dimensionality of highly correlated data by finding a small number of uncorrelated linear combinations that account for most of the variance of the original data. PCdimensional reductionA is not itself a model, but rather a data-rotation technique. However, it can be used as a way of constructing factors for use in factor modelling, and this is the main application we consider in this section.

The key mathematical result behind the technique is the *spectral decomposition theorem* of linear algebra, which says that any symmetric matrix $A \in \mathbb{R}^{d\times d}$ can be written as

$$A = \Gamma \Lambda \Gamma', \tag{6.59}$$

where

(i) $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ is the diagonal matrix of *eigenvalues of A* that, without loss of generality, are ordered so that $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_d$, and

(ii) Γ is an orthogonal matrix satisfying $\Gamma\Gamma' = \Gamma'\Gamma = I_d$ whose columns are standardized *eigenvectors of A* (i.e. eigenvectors with length 1).

Theoretical principal components. Obviously we can apply this decomposition to any covariance matrix Σ, and in this case the positive semidefiniteness of Σ ensures that $\lambda_j \geqslant 0$ for all j. Suppose the random vector $\boldsymbol{X}$ has mean vector $\boldsymbol{\mu}$ and covariance matrix Σ and we make the decomposition $\Sigma = \Gamma\Lambda\Gamma'$ as in (6.59). The principal components transform of $\boldsymbol{X}$ is then defined to be

$$\boldsymbol{Y} = \Gamma'(\boldsymbol{X} - \boldsymbol{\mu}), \tag{6.60}$$

and it can be thought of as a rotation and a recentring of $\boldsymbol{X}$. The jth component of the rotated vector $\boldsymbol{Y}$ is known as the *jth principal component* of $\boldsymbol{X}$ and is given by

$$Y_j = \boldsymbol{\gamma}_j'(\boldsymbol{X} - \boldsymbol{\mu}), \tag{6.61}$$

where $\boldsymbol{\gamma}_j$ is the eigenvector of Σ corresponding to the jth ordered eigenvalue; this vector is also known as the jth vector of *loadings*.

Simple calculations show that

$$E(\boldsymbol{Y}) = \boldsymbol{0} \quad \text{and} \quad \mathrm{cov}(\boldsymbol{Y}) = \Gamma'\Sigma\Gamma = \Gamma'\Gamma\Lambda\Gamma'\Gamma = \Lambda,$$

so that the principal components of $\boldsymbol{Y}$ are uncorrelated and have variances $\mathrm{var}(Y_j) = \lambda_j$, $\forall j$. The components are thus ordered by variance, from largest to

smallest. Moreover, the first principal component can be shown to be the standardized linear combination of $\boldsymbol{X}$ that has maximal variance among all such combinations; in other words,

$$\operatorname{var}(\boldsymbol{\gamma}_1'\boldsymbol{X}) = \max\{\operatorname{var}(\boldsymbol{a}'\boldsymbol{X}) : \boldsymbol{a}'\boldsymbol{a} = 1\}.$$

For $j = 2, \ldots, d$, the jth principal component can be shown to be the standardized linear combination of $\boldsymbol{X}$ with maximal variance among all such linear combinations that are *orthogonal* to (and hence uncorrelated with) the first $j - 1$ linear combinations. The final dth principal component has minimum variance among standardized linear combinations of $\boldsymbol{X}$.

To measure the ability of the first few principal components to explain the variance of $\boldsymbol{X}$, we observe that

$$\sum_{j=1}^{d} \operatorname{var}(Y_j) = \sum_{j=1}^{d} \lambda_j = \operatorname{trace}(\Sigma) = \sum_{j=1}^{d} \operatorname{var}(X_j).$$

If we interpret $\operatorname{trace}(\Sigma) = \sum_{j=1}^{d} \operatorname{var}(X_j)$ as a measure of the total variance of $\boldsymbol{X}$, then, for $k \leqslant d$, the ratio $\sum_{j=1}^{k} \lambda_j / \sum_{j=1}^{d} \lambda_j$ represents the amount of this variance that is explained by the first k principal components.

Principal components as factors. We note that, by inverting the principal components transform (6.60), we obtain

$$\boldsymbol{X} = \boldsymbol{\mu} + \Gamma\boldsymbol{Y} = \boldsymbol{\mu} + \Gamma_1\boldsymbol{Y}_1 + \Gamma_2\boldsymbol{Y}_2,$$

where we have partitioned $\boldsymbol{Y}$ into vectors $\boldsymbol{Y}_1 \in \mathbb{R}^k$ and $\boldsymbol{Y}_2 \in \mathbb{R}^{d-k}$, such that $\boldsymbol{Y}_1$ contains the first k principal components, and we have partitioned Γ into matrices $\Gamma_1 \in \mathbb{R}^{d\times k}$ and $\Gamma_2 \in \mathbb{R}^{d\times(d-k)}$ correspondingly. Let us assume that the first k principal components explain a large part of the total variance and we decide to focus our attention on them and ignore the further principal components in $\boldsymbol{Y}_2$. If we set $\boldsymbol{\varepsilon} = \Gamma_2\boldsymbol{Y}_2$, we obtain

$$\boldsymbol{X} = \boldsymbol{\mu} + \Gamma_1\boldsymbol{Y}_1 + \boldsymbol{\varepsilon}, \tag{6.62}$$

which is reminiscent of the basic factor model (6.50) with the vector $\boldsymbol{Y}_1$ playing the role of the factors and the matrix Γ_1 playing the role of the factor loading matrix. Although the components of the error vector $\boldsymbol{\varepsilon}$ will tend to have small variances, the assumptions of the factor model are generally violated in (6.62) since $\boldsymbol{\varepsilon}$ need not have a diagonal covariance matrix and need not be uncorrelated with $\boldsymbol{Y}_1$. Nevertheless, principal components are often interpreted as factors and used to develop approximate factor models. We now describe the estimation process that is followed when data are available.

Sample principal components. Assume that we have a time series of multivariate data observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ with identical distribution, unknown mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, with the spectral decomposition $\Sigma = \Gamma\Lambda\Gamma'$ as before.

To construct sample principal components we need to estimate the unknown parameters. We estimate $\boldsymbol{\mu}$ by $\bar{\boldsymbol{X}}$, the sample mean vector, and we estimate Σ by the sample covariance matrix

$$S_x = \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{X}_t - \bar{\boldsymbol{X}})(\boldsymbol{X}_t - \bar{\boldsymbol{X}})'.$$

We apply the spectral decomposition (6.59) to the symmetric, positive-semidefinite matrix S_x to get

$$S_x = GLG', \tag{6.63}$$

where G is the eigenvector matrix, $L = \text{diag}(l_1, \ldots, l_d)$ is the diagonal matrix consisting of ordered eigenvalues, and we switch to roman letters to emphasize that these are now calculated from an empirical covariance matrix. The matrix G provides an estimate of Γ and L provides an estimate of Λ.

By analogy with (6.60) we define vectors of sample principal components

$$\boldsymbol{Y}_t = G'(\boldsymbol{X}_t - \bar{\boldsymbol{X}}), \quad t = 1, \ldots, n. \tag{6.64}$$

The jth component of $\boldsymbol{Y}_t$ is known as the jth sample principal component at time t and is given by

$$Y_{t,j} = \boldsymbol{g}_j'(\boldsymbol{X}_t - \bar{\boldsymbol{X}}),$$

where $\boldsymbol{g}_j$ is the jth column of G, that is, the eigenvector of S_x corresponding to the jth largest eigenvalue.

The rotated vectors $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ have the property that their sample covariance matrix is L, as is easily verified:

$$\begin{aligned} S_y &= \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{Y}_t - \bar{\boldsymbol{Y}})(\boldsymbol{Y}_t - \bar{\boldsymbol{Y}})' = \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{Y}_t\boldsymbol{Y}_t' \\ &= \frac{1}{n}\sum_{t=1}^{n}G'(\boldsymbol{X}_t - \bar{\boldsymbol{X}})(\boldsymbol{X}_t - \bar{\boldsymbol{X}})'G = G'S_xG = L. \end{aligned}$$

The rotated vectors therefore show no correlation between components, and the components are ordered by their sample variances, from largest to smallest.

Remark 6.35. In a situation where the different components of the data vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ have very different sample variances (particularly if they are measured on very different scales), it is to be expected that the component (or components) with largest variance will dominate the first loading vector $\boldsymbol{g}_1$ and dominate the first principal component. In these situations the data are often transformed to have identical variances, which effectively means that principal component analysis is applied to the sample correlation matrix R_x. Note also that we could derive sample principal components from a robust estimate of the correlation matrix or a multivariate dispersion matrix.

We can now use the sample eigenvector matrix G and the sample principal components $\boldsymbol{Y}_t$ to calibrate an approximate factor model of the form (6.62). We assume that our data are realizations from the model

$$\boldsymbol{X}_t = \bar{\boldsymbol{X}} + G_1 \boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, n, \tag{6.65}$$

where G_1 consists of the first k columns of G and $\boldsymbol{F}_t = (Y_{t,1}, \ldots, Y_{t,k})'$, $t = 1, \ldots, n$. The choice of k is based on a subjective choice of the number of sample principal components that are required to explain a substantial part of the total sample variance (see Example 6.36).

Equation (6.65) bears a resemblance to the factor model (6.54) except that, in practice, the errors $\boldsymbol{\varepsilon}_t$ do not generally have a diagonal covariance matrix and are not generally uncorrelated with $\boldsymbol{F}_t$. Nevertheless, the method is a popular approach to constructing time series of statistically explanatory factors from multivariate time series of risk-factor changes.

Example 6.36 (PCA-based factor model for Dow Jones 30 returns). We consider the data in Example 6.34 again. Principal component analysis is applied to the sample covariance matrix of the return data and the results are summarized in Figures 6.5 and 6.6. In the former we see a bar plot of the sample variances of the first eight principal components l_j; the cumulative proportion of the total variance explained by the components is given above each bar; the first two components explain almost 50% of the variation. In the latter figure the first two loading vectors $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ are summarized.

The first vector of loadings is positively weighted for all stocks and can be thought of as describing a kind of index portfolio; of course, the weights in the loading vector do not sum to 1, but they can be scaled to do so and this gives a so-called principal-component-mimicking portfolio. The second vector has positive weights for the consumer titles and negative weights for the technology titles; as a portfolio it can be thought of as prescribing a programme of short selling of technology to buy consumer titles. These first two sample principal components loading vectors are used to define factors.

In Table 6.6 the transpose of the matrix $\hat{B}$ (containing the loadings estimates in the factor model) is shown; the rows are merely the first two loading vectors from the principal component analysis. In the third row, values of r^2, the so-called coefficient of determination, are given for each of the univariate regression models, and these indicate that more of the variation in the data is explained by the two PCA-based factors than was explained by the observed factor in Example 6.34; it seems that the model is best able to explain Intel returns.

The next ten lines give the correlation matrix implied by the factor model (corresponding to $\hat{\Sigma}^{(\mathrm{F})}$). Compared with the true sample correlation matrix in Example 6.34 this seems to pick up more of the structure than did the correlation matrix implied by the observed factor model.

The final ten lines show the estimated correlation matrix of the residuals from the regression model, but only those elements that exceed 0.1 in absolute value. The residuals are again less correlated than the original data, but there are quite a number

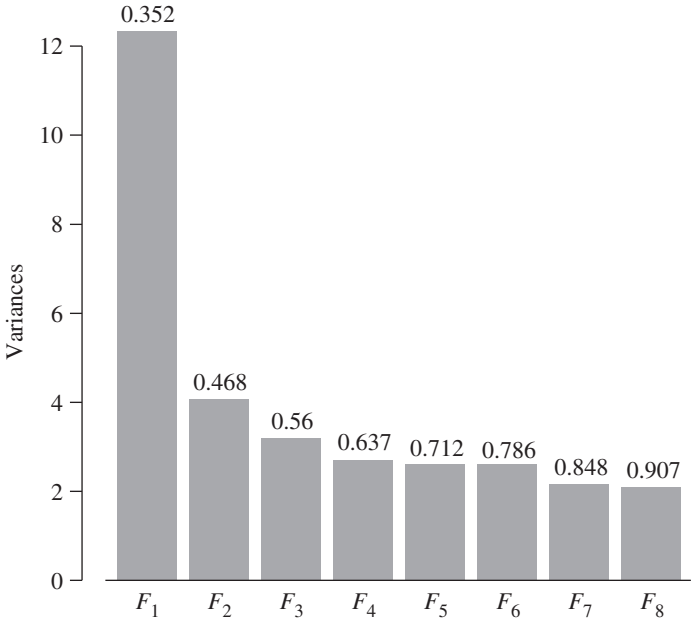

Figure 6.5. Bar plot of the sample variances l_j of the first eight principal components; the cumulative proportion of the total variance explained by the components is given above each bar ($\sum_{j=1}^{k} l_j / \sum_{j=1}^{10} l_j, k = 1, \ldots, 8$).

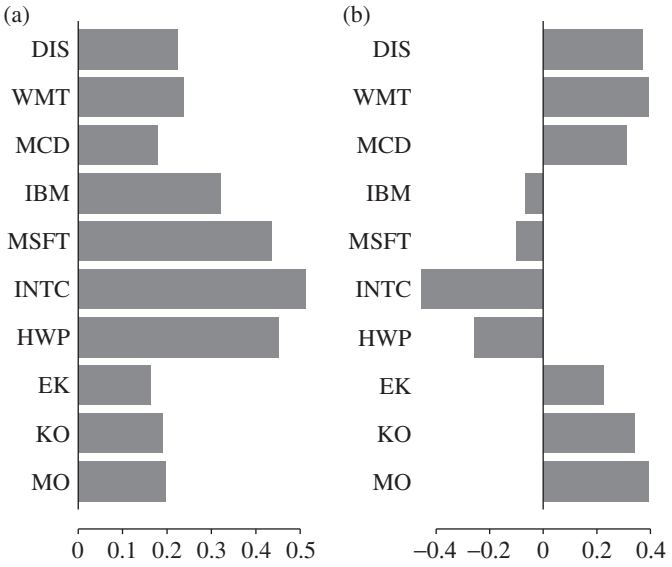

Figure 6.6. Bar plot summarizing the loading vectors $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ defining the first two principal components: (a) factor 1 loadings; (b) factor 2 loadings.

Table 6.6. The first two lines give estimates of the transpose of B for a factor model fitted to ten Dow Jones 30 stocks, where the factors are constructed from the first two sample principal components. The third row gives r^2 values for the univariate regression model for each individual time series. The next ten lines give the correlation matrix implied by the factor model. The final ten lines show the estimated correlation matrix of the residuals from the regression model, with entries less than 0.1 in absolute value omitted. See Example 6.36 for full details.

	MO	KO	EK	HWP	INTC	MSFT	IBM	MCD	WMT	DIS
$\hat{B}'$	0.20	0.19	0.16	0.45	0.51	0.44	0.32	0.18	0.24	0.22
	0.39	0.34	0.23	−0.26	−0.45	−0.10	−0.07	0.31	0.39	0.37
r^2	0.35	0.42	0.18	0.55	0.75	0.56	0.35	0.34	0.42	0.41
MO	1.00	0.39	0.25	0.17	0.13	0.25	0.20	0.35	0.38	0.38
KO	0.39	1.00	0.28	0.21	0.17	0.29	0.23	0.38	0.42	0.42
EK	0.25	0.28	1.00	0.18	0.15	0.22	0.18	0.25	0.28	0.27
HWP	0.17	0.21	0.18	1.00	0.64	0.55	0.43	0.20	0.23	0.23
INTC	0.13	0.17	0.15	0.64	1.00	0.61	0.48	0.16	0.19	0.18
MSFT	0.25	0.29	0.22	0.55	0.61	1.00	0.44	0.27	0.31	0.30
IBM	0.20	0.23	0.18	0.43	0.48	0.44	1.00	0.21	0.25	0.24
MCD	0.35	0.38	0.25	0.20	0.16	0.27	0.21	1.00	0.38	0.37
WMT	0.38	0.42	0.28	0.23	0.19	0.31	0.25	0.38	1.00	0.41
DIS	0.38	0.42	0.27	0.23	0.18	0.30	0.24	0.37	0.41	1.00
MO	1.00	−0.19	−0.15					−0.19	−0.37	−0.26
KO	−0.19	1.00	−0.15		0.11				−0.16	−0.17
EK	−0.15	−0.15	1.00					−0.15	−0.16	−0.16
HWP				1.00	−0.63	−0.37	−0.14			
INTC		0.11		−0.63	1.00	−0.24	−0.31			
MSFT				−0.37	−0.24	1.00	−0.22			
IBM				−0.14	−0.31	−0.22	1.00			
MCD	−0.19		−0.15					1.00	−0.19	−0.19
WMT	−0.37	−0.16	−0.16					−0.19	1.00	−0.23
DIS	−0.26	−0.17	−0.16					−0.19	−0.23	1.00

of larger entries, indicating imperfections in the factor-model representation of the data. In particular, we have introduced a number of larger negative correlations into the residuals; in practice, we seldom expect to find a factor model in which the residuals have a covariance matrix that appears perfectly diagonal.

Notes and Comments

For a more detailed discussion of factor models see the paper by Connor (1995), which provides a comparison of the three types of model, and the book by Campbell, Lo and MacKinlay (1997). An excellent practical introduction to these models with examples in S-Plus is Zivot and Wang (2003). Other accounts of factor models and PCA in finance are found in Alexander (2001) and Tsay (2002).

Much of our discussion of factor models, multivariate regression and principal components is based on Mardia, Kent and Bibby (1979). Statistical approaches to factor models are also treated in Seber (1984) and Johnson and Wichern (2002); these include classical statistical factor analysis, which we have omitted from our account.

7
Copulas and Dependence

In this chapter we use the concept of a copula to look more closely at the issue of modelling a random vector of dependent financial risk factors. In Section 7.1 we define copulas, give a number of examples and establish their basic properties.

Dependence concepts and dependence measures are considered in Section 7.2, beginning with the notion of perfect positive dependence or comonotonicity. This is a very important concept in risk management because it formalizes the idea of undiversifiable risks and therefore has important implications for determining risk-based capital. Dependence measures provide a scalar-valued summary of the strength of dependence between risks and there are many different measures; we consider linear correlation and two further classes of measures—rank correlations and coefficients of tail dependence—that can be directly related to copulas.

Linear correlation is a standard measure for describing the dependence between financial assets but it has a number of limitations, particularly when we leave the multivariate normal and elliptical distributions of Chapter 6 behind. Rank correlations are mainly used to calibrate copulas to data, while tail dependence is an important theoretical concept, since it addresses the phenomenon of joint extreme values in several risk factors, which is one of the major concerns in financial risk management (see also Section 3.2).

In Section 7.3 we look in more detail at the copulas of normal mixture distributions; these are the copulas that are used implicitly when normal mixture distributions are fitted to multivariate risk-factor change data, as in Chapter 6. In Section 7.4 we consider Archimedean copulas, which are widely used as dependence models in low-dimensional applications and which have also found an important niche in portfolio credit risk modelling, as will be seen in Chapters 11 and 12. The chapter ends with a section on fitting copulas to data.

7.1 Copulas

In a sense, every joint distribution function for a random vector of risk factors implicitly contains both a description of the marginal behaviour of individual risk factors and a description of their *dependence structure*; the copula approach provides a way of isolating the description of the dependence structure. We view copulas as an extremely useful concept and see several advantages in introducing and studying them.

First, copulas help in the understanding of dependence at a deeper level. They allow us to see the potential pitfalls of approaches to dependence that focus only on correlation and show us how to define a number of useful alternative dependence measures. Copulas express dependence on a *quantile scale*, which is useful for describing the dependence of extreme outcomes and is natural in a risk-management context, where VaR has led us to think of risk in terms of quantiles of loss distributions.

Moreover, copulas facilitate a *bottom-up approach to multivariate model building*. This is particularly useful in risk management, where we very often have a much better idea about the marginal behaviour of individual risk factors than we do about their dependence structure. An example is furnished by credit risk, where the individual default risk of an obligor, while in itself difficult to estimate, is at least something we can get a better handle on than the dependence among default risks for several obligors. The copula approach allows us to combine our more developed marginal models with a variety of possible dependence models and to investigate the sensitivity of risk to the dependence specification. Since the copulas we present are easily simulated, they lend themselves particularly well to Monte Carlo studies of risk. Of course, while the flexibility of the copula approach allows us, in theory, to build an unlimited number of models with given marginal distributions, we should stress that it is important to have a good understanding of the behaviour of different copulas and their appropriateness for particular kinds of modelling application.

7.1.1 Basic Properties

Definition 7.1 (copula). A d-dimensional copula is a distribution function on $[0, 1]^d$ with standard uniform marginal distributions.

We reserve the notation $C(\boldsymbol{u}) = C(u_1, \ldots, u_d)$ for the multivariate dfs that are copulas. Hence C is a mapping of the form $C\colon [0, 1]^d \to [0, 1]$, i.e. a mapping of the unit hypercube into the unit interval. The following three properties must hold.

(1) $C(u_1, \ldots, u_d) = 0$ if $u_i = 0$ for any i.

(2) $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all $i \in \{1, \ldots, d\}$, $u_i \in [0, 1]$.

(3) For all $(a_1, \ldots, a_d), (b_1, \ldots, b_d) \in [0, 1]^d$ with $a_i \leqslant b_i$ we have

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C(u_{1i_1}, \ldots, u_{di_d}) \geqslant 0, \tag{7.1}$$

where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j \in \{1, \ldots, d\}$.

Note that the second property corresponds to the requirement that marginal distributions are uniform. The so-called rectangle inequality in (7.1) ensures that if the random vector $(U_1, \ldots, U_d)'$ has df C, then $P(a_1 \leqslant U_1 \leqslant b_1, \ldots, a_d \leqslant U_d \leqslant b_d)$ is non-negative. These three properties characterize a copula; if a function C fulfills them, then it is a copula. Note also that, for $2 \leqslant k < d$, the k-dimensional margins of a d-dimensional copula are themselves copulas.

Some preliminaries. In working with copulas we must be familiar with the operations of *probability* and *quantile transformation*, as well as the properties of generalized inverses, which are summarized in Section A.1.2. The following elementary proposition is found in many probability texts.

Proposition 7.2. *Let F be a distribution function and let $F^{\leftarrow}$ denote its generalized inverse, i.e. the function $F^{\leftarrow}(u) = \inf\{x : F(x) \geqslant u\}$.*

(1) Quantile transform. *If $U \sim U(0,1)$ has a standard uniform distribution, then $P(F^{\leftarrow}(U) \leqslant x) = F(x)$.*

(2) Probability transform. *If X has df F, where F is a continuous univariate df, then $F(X) \sim U(0,1)$.*

Proof. Let $x \in \mathbb{R}$ and $u \in (0,1)$. For the first part use the fact that

$$F(x) \geqslant u \iff F^{\leftarrow}(u) \leqslant x$$

(see Proposition A.3 (iv) in Section A.1.2), from which it follows that

$$P(F^{\leftarrow}(U) \leqslant x) = P(U \leqslant F(x)) = F(x).$$

For the second part we infer that

$$\begin{aligned} P(F(X) \leqslant u) &= P(F^{\leftarrow} \circ F(X) \leqslant F^{\leftarrow}(u)) \\ &= P(X \leqslant F^{\leftarrow}(u)) = F \circ F^{\leftarrow}(u) \\ &= u, \end{aligned}$$

where the first inequality follows from the fact that $F^{\leftarrow}$ is strictly increasing (Proposition A.3 (ii)), the second follows from Proposition A.4, and the final equality is Proposition A.3 (viii). □

Proposition 7.2 (1) is the key to stochastic simulation. If we can generate a uniform variate U and compute the inverse of a df F, then we can sample from that df. Both parts of the proposition taken together imply that we can transform risks with a particular continuous df to have any other continuous distribution. For example, if X has a standard normal distribution, then $\Phi(X)$ is uniform by Proposition 7.2 (2), and, since the quantile function of a standard exponential df G is $G^{\leftarrow}(u) = -\ln(1-u)$, the transformed variable $Y := -\ln(1-\Phi(X))$ has a standard exponential distribution by Proposition 7.2 (1).

Sklar's Theorem. The importance of copulas in the study of multivariate distribution functions is summarized by the following elegant theorem, which shows, firstly, that all multivariate dfs contain copulas and, secondly, that copulas may be used in conjunction with univariate dfs to construct multivariate dfs.

Theorem 7.3 (Sklar 1959). *Let F be a joint distribution function with margins $F_1, \ldots, F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that, for all $x_1, \ldots, x_d$ in $\bar{\mathbb{R}} = [-\infty, \infty]$,*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{7.2}$$

If the margins are continuous, then C is unique; otherwise C is uniquely determined on $\operatorname{Ran} F_1 \times \operatorname{Ran} F_2 \times \cdots \times \operatorname{Ran} F_d$, *where* $\operatorname{Ran} F_i = F_i(\bar{\mathbb{R}})$ *denotes the range of* F_i. *Conversely, if C is a copula and $F_1, \ldots, F_d$ are univariate distribution functions, then the function F defined in (7.2) is a joint distribution function with margins $F_1, \ldots, F_d$.*

Proof. We prove the existence and uniqueness of a copula in the case when $F_1, \ldots, F_d$ are continuous and the converse statement in its general form. Remark 7.4 explains how the general result may be proved with the more complicated distributional transform, which is given in Appendix A.1.3.

Let $\boldsymbol{X}$ be a random vector with df F and continuous margins $F_1, \ldots, F_d$, and, for $i = 1, \ldots, d$, set $U_i = F_i(X_i)$. By Proposition 7.2 (2), $U_i \sim U(0, 1)$, and, by Proposition A.4 in the appendix, $F_i^{\leftarrow}(U_i) = X_i$, almost surely. Let C denote the distribution function of $(U_1, \ldots, U_d)$, which is a copula by Definition 7.1. For any $x_1, \ldots, x_d$ in $\bar{\mathbb{R}} = [-\infty, \infty]$ we infer, using Proposition A.3 (iv), that

$$\begin{aligned} F(x_1, \ldots, x_d) &= P(X_1 \leqslant x_1, \ldots, X_d \leqslant x_d) \\ &= P(F_1^{\leftarrow}(U_1) \leqslant x_1, \ldots, F_d^{\leftarrow}(U_d) \leqslant x_d) \\ &= P(U_1 \leqslant F_1(x_1), \ldots, U_d \leqslant F(x_d)) \\ &= C(F_1(x_1), \ldots, F_d(x_d)), \end{aligned} \tag{7.3}$$

and thus we obtain the identity (7.2).

If we evaluate (7.2) at the arguments $x_i = F_i^{\leftarrow}(u_i)$, $0 \leqslant u_i \leqslant 1$, $i = 1, \ldots, d$, and use Proposition A.3 (viii), we obtain

$$C(u_1, \ldots, u_d) = F(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d)), \tag{7.4}$$

which gives an explicit representation of C in terms of F and its margins, and thus shows uniqueness.

For the converse statement assume that C is a copula and that $F_1, \ldots, F_d$ are arbitrary univariate dfs. We construct a random vector with df (7.2) by taking $\boldsymbol{U}$ to be *any* random vector with df C and setting $\boldsymbol{X} := (F_1^{\leftarrow}(U_1), \ldots, F_d^{\leftarrow}(U_d))$. We can then follow exactly the same sequence of equations commencing with (7.3) to establish that the df of $\boldsymbol{X}$ satisfies (7.2). □

Remark 7.4. The general form of Sklar's Theorem can be proved by using the distributional transform in Appendix A.1.3 instead of the probability transform. For a random vector $\boldsymbol{X}$ with arbitrary df F and margins $F_1, \ldots, F_d$ we can set $U_i = \tilde{F}_i(X_i, V_i)$, where $\tilde{F}_i$ is the modified distribution function of X_i defined in (A.6) and $V_1, \ldots, V_d$ are uniform rvs that are independent of $X_1, \ldots, X_d$. Proposition A.6 shows that $U_i \sim U(0, 1)$ and $F_i^{\leftarrow}(U_i) = X_i$, almost surely, so an otherwise-identical proof may be used. The non-uniqueness of the copula is related to the fact that there are different ways of choosing the V_i variables; they need not themselves be independent and could in fact be identical for all i.

Formulas (7.2) and (7.4) are fundamental in dealing with copulas. The former shows how joint distributions F are formed by *coupling together* marginal distributions with copulas C; the latter shows how copulas are *extracted* from multivariate dfs with continuous margins. Moreover, (7.4) shows how copulas express dependence on a quantile scale, since the value $C(u_1, \ldots, u_d)$ is the joint probability that X_1 lies below its u_1-quantile, X_2 lies below its u_2-quantile, and so on. Sklar's Theorem also suggests that, in the case of continuous margins, it is natural to define the notion of the copula of a distribution.

Definition 7.5 (copula of F). If the random vector $\boldsymbol{X}$ has joint df F with continuous marginal distributions $F_1, \ldots, F_d$, then the copula of F (or $\boldsymbol{X}$) is the df C of $(F_1(X_1), \ldots, F_d(X_d))$.

Discrete distributions. The copula concept is slightly less natural for multivariate discrete distributions. This is because there is more than one copula that can be used to join the margins to form the joint df, as the following example shows.

Example 7.6 (copulas of bivariate Bernoulli). Let (X_1, X_2) have a bivariate Bernoulli distribution satisfying

$$P(X_1 = 0, X_2 = 0) = \tfrac{1}{8}, \quad P(X_1 = 1, X_2 = 1) = \tfrac{3}{8},$$
$$P(X_1 = 0, X_2 = 1) = \tfrac{2}{8}, \quad P(X_1 = 1, X_2 = 0) = \tfrac{2}{8}.$$

Clearly, $P(X_1 = 0) = P(X_2 = 0) = \frac{3}{8}$ and the marginal distributions F_1 and F_2 of X_1 and X_2 are the same. From Sklar's Theorem we know that

$$P(X_1 \leqslant x_1, X_2 \leqslant x_2) = C(P(X_1 \leqslant x_1), P(X_2 \leqslant x_2))$$

for all x_1, x_2 and some copula C. Since $\operatorname{Ran} F_1 = \operatorname{Ran} F_2 = \{0, \frac{3}{8}, 1\}$, clearly the only constraint on C is that $C(\frac{3}{8}, \frac{3}{8}) = \frac{1}{8}$. Any copula fulfilling this constraint is a copula of (X_1, X_2), and there are infinitely many such copulas.

Invariance. A useful property of the copula of a distribution is its invariance under *strictly increasing* transformations of the marginals. In view of Sklar's Theorem and this invariance property, we interpret the copula of a distribution as a very natural way of representing the dependence structure of that distribution, certainly in the case of continuous margins.

Proposition 7.7. *Let $(X_1, \ldots, X_d)$ be a random vector with continuous margins and copula C and let $T_1, \ldots, T_d$ be strictly increasing functions. Then $(T_1(X_1), \ldots, T_d(X_d))$ also has copula C.*

Proof. We first note that $(T_1(X_1), \ldots, T_d(X_d))$ is also a random vector with continuous margins and that it will have the same distribution regardless of whether each T_i is left continuous or right continuous at its (countably many) points of discontinuity. By starting with the expression (7.4) for the unique copula C of $(X_1, \ldots, X_d)$, using the fact that $\{X_i \leqslant x\} = \{T_i(X_i) \leqslant T_i(x)\}$ for a strictly increasing transformation

T_i and then applying Proposition A.5 for left-continuous transformations, we obtain

$$\begin{aligned} C(u_1, \ldots, u_d) &= P(X_1 \leqslant F_1^{\leftarrow}(u_1), \ldots, X_d \leqslant F_d^{\leftarrow}(u_d)) \\ &= P(T_1(X_1) \leqslant T_1 \circ F_1^{\leftarrow}(u_1), \ldots, T_d(X_d) \leqslant T_d \circ F_d^{\leftarrow}(u_d)) \\ &= P(T_1(X_1) \leqslant F_{T_1(X_1)}^{\leftarrow}(u_1), \ldots, T_d(X_d) \leqslant F_{T_d(X_d)}^{\leftarrow}(u_d)), \end{aligned}$$

which shows that C is also the unique copula of $(T_1(X_1), \ldots, T_d(X_d))$. □

Fréchet bounds. We close this section by establishing the important *Fréchet bounds* for copulas, which turn out to have important dependence interpretations that are discussed further in Sections 7.1.2 and 7.2.1.

Theorem 7.8. *For every copula $C(u_1, \ldots, u_d)$ we have the bounds*

$$\max\left(\sum_{i=1}^{d} u_i + 1 - d, 0\right) \leqslant C(\boldsymbol{u}) \leqslant \min(u_1, \ldots, u_d). \tag{7.5}$$

Proof. The second inequality follows from the fact that, for all i,

$$\bigcap_{1 \leqslant j \leqslant d} \{U_j \leqslant u_j\} \subset \{U_i \leqslant u_i\}.$$

For the first inequality observe that

$$\begin{aligned} C(\boldsymbol{u}) = P\left(\bigcap_{1 \leqslant i \leqslant d} \{U_i \leqslant u_i\}\right) &= 1 - P\left(\bigcup_{1 \leqslant i \leqslant d} \{U_i > u_i\}\right) \\ &\geqslant 1 - \sum_{i=1}^{d} P(U_i > u_i) = 1 - d + \sum_{i=1}^{d} u_i. \end{aligned}$$

□

The lower and upper bounds will be given the notation $W(u_1, \ldots, u_d)$ and $M(u_1, \ldots, u_d)$, respectively.

Remark 7.9. Although we give Fréchet bounds for a copula, Fréchet bounds may be given for any multivariate df. For a multivariate df F with margins $F_1, \ldots, F_d$ we establish by similar reasoning that

$$\max\left(\sum_{i=1}^{d} F_i(x_i) + 1 - d, 0\right) \leqslant F(\boldsymbol{x}) \leqslant \min(F_1(x_1), \ldots, F_d(x_d)), \tag{7.6}$$

so we have bounds for F in terms of its own marginal distributions.

7.1.2 Examples of Copulas

We provide a number of examples of copulas in this section and these are subdivided into three categories: *fundamental* copulas represent a number of important special dependence structures; *implicit* copulas are extracted from well-known multivariate distributions using Sklar's Theorem, but do not necessarily possess simple closed-form expressions; *explicit* copulas have simple closed-form expressions and follow

mathematical constructions known to yield copulas. Note, however, that implicit and explicit are not mutually exclusive categories, and some copulas may have both implicit and explicit representations, as shown later in Example 7.14.

Fundamental copulas. The *independence copula* is

$$\Pi(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i. \tag{7.7}$$

It is clear from Sklar's Theorem, and equation (7.2) in particular, that rvs with continuous distributions are independent if and only if their dependence structure is given by (7.7).

The *comonotonicity copula* is the Fréchet upper bound copula from (7.5):

$$M(u_1, \ldots, u_d) = \min(u_1, \ldots, u_d). \tag{7.8}$$

Observe that this special copula is the joint df of the random vector $(U, \ldots, U)$, where $U \sim U(0, 1)$. Suppose that the rvs $X_1, \ldots, X_d$ have continuous dfs and are *perfectly positively dependent* in the sense that they are almost surely strictly increasing functions of each other, so that $X_i = T_i(X_1)$ almost surely for $i = 2, \ldots, d$. By Proposition 7.7, the copula of $(X_1, \ldots, X_d)$ and the copula of $(X_1, \ldots, X_1)$ are the same. But the copula of $(X_1, \ldots, X_1)$ is just the df of $(U, \ldots, U)$, where $U = F_1(X_1)$, i.e. the copula (7.8).

The *countermonotonicity copula* is the two-dimensional Fréchet lower bound copula from (7.5) given by

$$W(u_1, u_2) = \max(u_1 + u_2 - 1, 0). \tag{7.9}$$

This copula is the joint df of the random vector $(U, 1 - U)$, where $U \sim U(0, 1)$. If X_1 and X_2 have continuous dfs and are *perfectly negatively dependent* in the sense that X_2 is almost surely a strictly decreasing function of X_1, then (7.9) is their copula.

We discuss both perfect positive and perfect negative dependence in more detail in Section 7.2.1, where we see that an extension of the countermonotonicity concept to dimensions higher than two is not possible.

Perspective pictures and contour plots for the three fundamental copulas are given in Figure 7.1. The Fréchet bounds (7.5) imply that all bivariate copulas lie between the surfaces in (a) and (c).

Implicit copulas. If $\boldsymbol{Y} \sim N_d(\boldsymbol{\mu}, \Sigma)$ is a multivariate normal random vector, then its copula is a so-called *Gauss copula* (or Gaussian copula). Since the operation of standardizing the margins amounts to applying a series of strictly increasing transformations, Proposition 7.7 implies that the copula of $\boldsymbol{Y}$ is exactly the same as the copula of $\boldsymbol{X} \sim N_d(\boldsymbol{0}, P)$, where $P = \wp(\Sigma)$ is the correlation matrix of $\boldsymbol{Y}$. By Definition 7.5 this copula is given by

$$\begin{aligned} C_P^{\mathrm{Ga}}(\boldsymbol{u}) &= P(\Phi(X_1) \leqslant u_1, \ldots, \Phi(X_d) \leqslant u_d) \\ &= \boldsymbol{\Phi}_P(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)), \end{aligned} \tag{7.10}$$

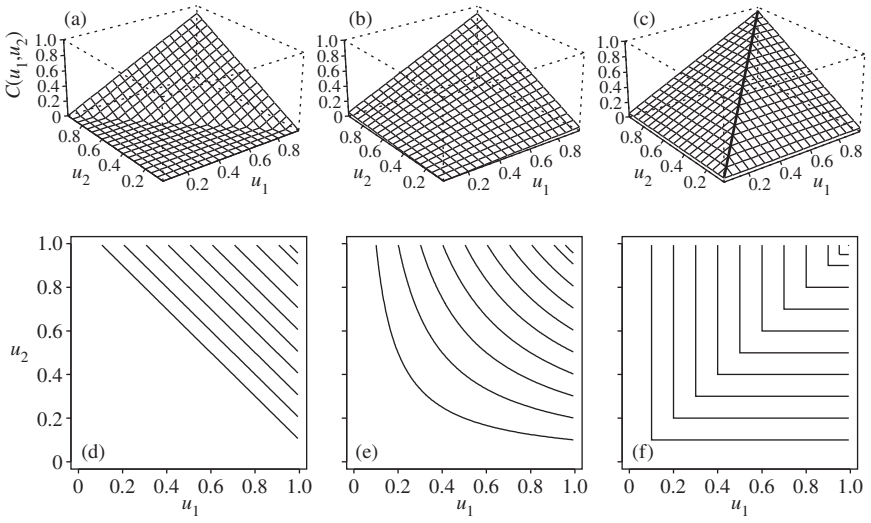

Figure 7.1. (a)–(c) Perspective plots and (d)–(f) contour plots of the three fundamental copulas: (a), (d) countermonotonicity, (b), (e) independence and (c), (f) comonotonicity. Note that these are plots of distribution functions.

where Φ denotes the standard univariate normal df and $\boldsymbol{\Phi}_P$ denotes the joint df of $\boldsymbol{X}$. The notation C_P^{Ga} emphasizes that the copula is parametrized by the $\frac{1}{2}d(d-1)$ parameters of the correlation matrix; in two dimensions we write C_ρ^{Ga}, where $\rho = \rho(X_1, X_2)$.

The Gauss copula does not have a simple closed form, but can be expressed as an integral over the density of $\boldsymbol{X}$; in two dimensions for $|\rho| < 1$ we have, using (7.10), that

$$C_\rho^{\mathrm{Ga}}(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left(\frac{-(s_1^2 - 2\rho s_1 s_2 + s_2^2)}{2(1-\rho^2)}\right) \mathrm{d}s_1\, \mathrm{d}s_2.$$

Note that both the independence and comonotonicity copulas are special cases of the Gauss copula. If $P = I_d$, we obtain the independence copula (7.7); if $P = J_d$, the $d \times d$ matrix consisting entirely of ones, then we obtain the comonotonicity copula (7.8). Also, for $d = 2$ and $\rho = -1$ the Gauss copula is equal to the countermonotonicity copula (7.9). Thus in two dimensions the Gauss copula can be thought of as a dependence structure that interpolates between perfect positive and negative dependence, where the parameter ρ represents the strength of dependence.

Perspective plots and contour lines of the bivariate Gauss copula with $\rho = 0.7$ are shown in Figure 7.2 (a),(c); these may be compared with the contour lines of the independence and perfect dependence copulas in Figure 7.1. Note that these pictures show contour lines of distribution functions and not densities; a picture of the Gauss copula density is given in Figure 7.5.

In the same way that we can extract a copula from the multivariate normal distribution, we can extract an implicit copula from any other distribution with continuous

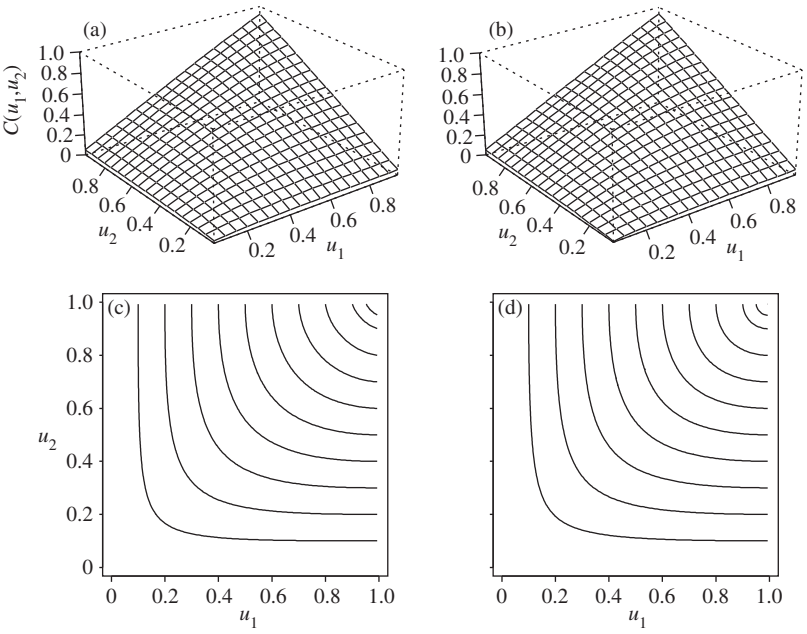

Figure 7.2. (a), (b) Perspective plots and (c), (d) contour plots of the Gaussian and Gumbel copulas, with parameters $\rho = 0.7$ and $\theta = 2$, respectively. Note that these are plots of distribution functions; a picture of the Gauss copula density is given in Figure 7.5.

marginal dfs. For example, the d-dimensional *t copula* takes the form

$$C^t_{\nu,P}(\boldsymbol{u}) = \boldsymbol{t}_{\nu,P}(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)), \tag{7.11}$$

where t_ν is the df of a standard univariate t distribution with ν degrees of freedom, $\boldsymbol{t}_{\nu,P}$ is the joint df of the vector $\boldsymbol{X} \sim t_d(\nu, \mathbf{0}, P)$, and P is a correlation matrix. As in the case of the Gauss copula, if $P = J_d$ then we obtain the comonotonicity copula (7.8). However, in contrast to the Gauss copula, if $P = I_d$ we do not obtain the independence copula (assuming $\nu < \infty$) since uncorrelated multivariate t-distributed rvs are not independent (see Lemma 6.5).

Explicit copulas. While the Gauss and t copulas are copulas implied by well-known multivariate dfs and do not themselves have simple closed forms, we can write down a number of copulas that do have simple closed forms. An example is the bivariate *Gumbel copula*:

$$C^{\mathrm{Gu}}_\theta(u_1, u_2) = \exp(-((-\ln u_1)^\theta + (-\ln u_2)^\theta)^{1/\theta}), \quad 1 \leqslant \theta < \infty. \tag{7.12}$$

If $\theta = 1$ we obtain the independence copula as a special case, and the limit of C^{Gu}_θ as $\theta \to \infty$ is the two-dimensional comonotonicity copula. Thus the Gumbel copula interpolates between independence and perfect dependence and the parameter θ represents the strength of dependence. Perspective plot and contour lines for the Gumbel copula with parameter $\theta = 2$ are shown in Figure 7.2 (b),(d). They appear to be very similar to the picture for the Gauss copula, but Example 7.13 will show that the Gaussian and Gumbel dependence structures are quite different.

A further example is the bivariate *Clayton copula*:

$$C_\theta^{\mathrm{Cl}}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, \quad 0 \leqslant \theta < \infty. \tag{7.13}$$

The case $\theta = 0$ should be interpreted as the limit of (7.13) as $\theta \to 0$, which is the independence copula; as $\theta \to \infty$ we approach the two-dimensional comonotonicity copula.

The Gumbel and Clayton copulas belong to the *Archimedean* copula family and we provide more discussion of this family in Sections 7.4 and 15.2.

7.1.3 Meta Distributions

The converse statement of Sklar's Theorem provides a very powerful technique for constructing multivariate distributions with arbitrary margins and copulas; we know that if we start with a copula C and margins $F_1, \ldots, F_d$, then $F(\boldsymbol{x}) := C(F_1(x_1), \ldots, F_d(x_d))$ defines a multivariate df with margins $F_1, \ldots, F_d$.

Consider, for example, building a distribution with the Gauss copula C_P^{Ga} but arbitrary margins; such a model is sometimes called a *meta-Gaussian* distribution. We extend the meta terminology to other distributions, so, for example, a *meta-t_ν* distribution has the copula $C_{\nu,P}^t$ and arbitrary margins, and a *meta-Clayton* distribution has the Clayton copula and arbitrary margins.

7.1.4 Simulation of Copulas and Meta Distributions

It should be apparent from the way the implicit copulas in Section 7.1.2 were extracted from well-known distributions that it is particularly easy to sample from these copulas, provided we can sample from the distribution from which they are extracted. The steps are summarized in the following algorithm.

Algorithm 7.10 (simulation of implicit copulas).

(1) Generate $\boldsymbol{X} \sim F$, where F is a df with continuous margins $F_1, \ldots, F_d$.

(2) Return $\boldsymbol{U} = (F_1(X_1), \ldots, F_d(X_d))'$. The random vector $\boldsymbol{U}$ has df C, where C is the unique copula of F.

Particular examples are given in the following algorithms.

Algorithm 7.11 (simulation of Gauss copula).

(1) Generate $\boldsymbol{Z} \sim N_d(\mathbf{0}, P)$ using Algorithm 6.2.

(2) Return $\boldsymbol{U} = (\Phi(Z_1), \ldots, \Phi(Z_d))'$, where Φ is the standard normal df. The random vector $\boldsymbol{U}$ has df C_P^{Ga}.

Algorithm 7.12 (simulation of t copula).

(1) Generate $\boldsymbol{X} \sim t_d(\nu, \mathbf{0}, P)$ using Algorithm 6.10.

(2) Return $\boldsymbol{U} = (t_\nu(X_1), \ldots, t_\nu(X_d))'$, where t_ν denotes the df of a standard univariate t distribution. The random vector $\boldsymbol{U}$ has df $C_{\nu,P}^t$.

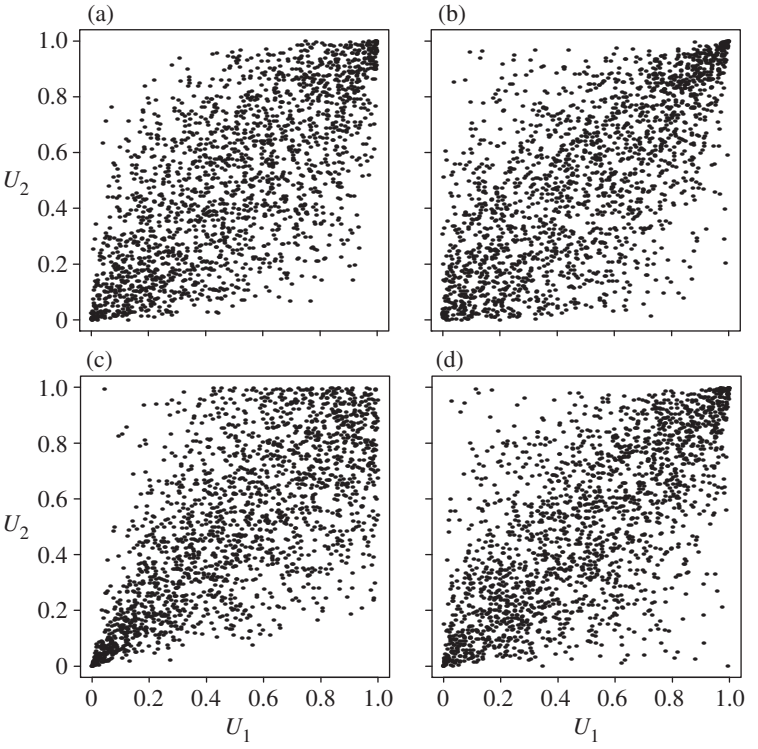

Figure 7.3. Two thousand simulated points from the (a) Gaussian, (b) Gumbel, (c) Clayton and (d) t copulas. See Example 7.13 for parameter choices and interpretation.

The Clayton and Gumbel copulas present slightly more challenging simulation problems and we give algorithms in Section 7.4 after looking at the structure of these copulas in more detail. These algorithms will, however, be used in Example 7.13 below.

Assume that the problem of generating realizations $\boldsymbol{U}$ from a particular copula has been solved. The converse of Sklar's Theorem shows us how we can sample from meta distributions that combine this copula with an arbitrary choice of marginal distribution. If $\boldsymbol{U}$ has df C, then we use quantile transformation to obtain $\boldsymbol{X} := (F_1^{\leftarrow}(U_1), \ldots, F_d^{\leftarrow}(U_d))'$, which is a random vector with margins $F_1, \ldots, F_d$ and multivariate df $C(F_1(x_1), \ldots, F_d(x_d))$. This technique is extremely useful in Monte Carlo studies of risk and will be discussed further in the context of Example 7.58.

Example 7.13 (various copulas compared). In Figure 7.3 we show 2000 simulated points from four copulas: the Gauss copula (7.10) with parameter $\rho = 0.7$; the Gumbel copula (7.12) with parameter $\theta = 2$; the Clayton copula (7.13) with parameter $\theta = 2.2$; and the t copula (7.11) with parameters $\nu = 4$ and $\rho = 0.71$.

In Figure 7.4 we transform these points componentwise using the quantile function of the standard normal distribution to get realizations from four different meta distributions with standard normal margins. The Gaussian picture shows data generated from a standard bivariate normal distribution with correlation 70%. The

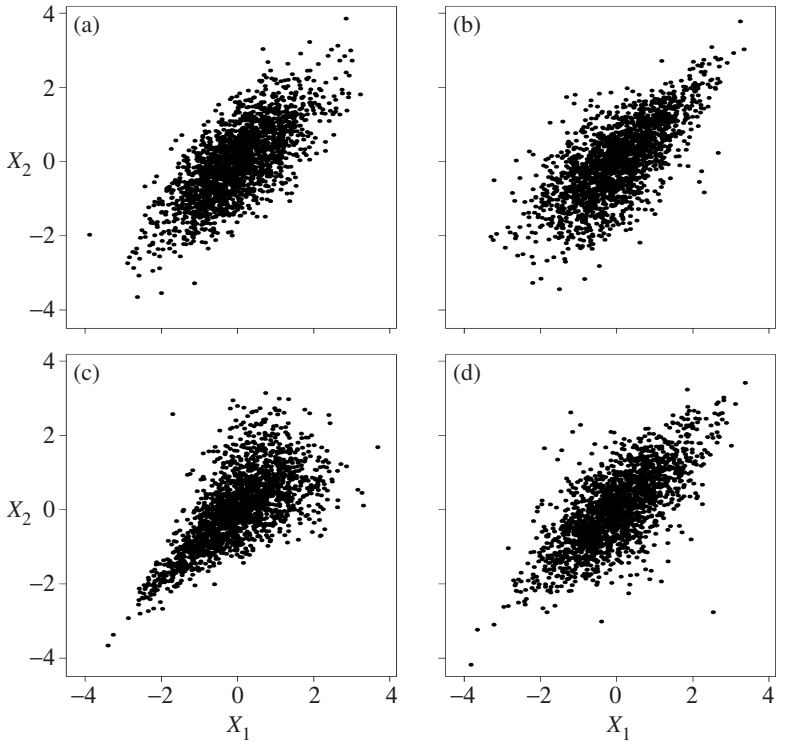

Figure 7.4. Two thousand simulated points from four distributions with standard normal margins, constructed using the copula data from Figure 7.3 ((a) Gaussian, (b) Gumbel, (c) Clayton and (d) t). The Gaussian picture shows points from a standard bivariate normal with correlation 70%; other pictures show distributions with non-Gauss copulas constructed to have a linear correlation of roughly 70%. See Example 7.13 for parameter choices and interpretation.

other pictures show data generated from unusual distributions that have been created using the converse of Sklar's Theorem; the parameters of the copulas have been chosen so that all of these distributions have a linear correlation that is roughly 70%.

Considering the Gumbel picture, these are bivariate data with a meta-Gumbel distribution with df $C_\theta^{\mathrm{Gu}}(\Phi(x_1), \Phi(x_2))$, where $\theta = 2$. The Gumbel copula causes this distribution to have *upper tail dependence*, a concept defined formally in Section 7.2.4. Roughly speaking, there is much more of a tendency for X_2 to be extreme when X_1 is extreme, and vice versa, a phenomenon that would obviously be worrying when X_1 and X_2 are interpreted as potential financial losses. The Clayton copula turns out to have *lower tail dependence*, and the t copula to have both lower and upper tail dependence; in contrast, the Gauss copula does not have tail dependence and this can also be glimpsed in Figure 7.2. In the upper-right-hand corner the contours of the Gauss copula are more like those of the independence copula of Figure 7.1 than the perfect dependence copula.

Note that the qualitative differences between the distributions are explained by the copula alone; we can construct similar pictures where the marginal distributions are exponential or Student t, or any other univariate distribution.

7.1.5 Further Properties of Copulas

Survival copulas. A version of Sklar's identity (7.2) also applies to multivariate survival functions of distributions. Let $\boldsymbol{X}$ be a random vector with multivariate survival function $\bar{F}$, marginal dfs $F_1, \ldots, F_d$ and marginal survival functions $\bar{F}_1, \ldots, \bar{F}_d$, i.e. $\bar{F}_i = 1 - F_i$. We have the identity

$$\bar{F}(x_1, \ldots, x_d) = \hat{C}(\bar{F}_1(x_1), \ldots, \bar{F}_d(x_d)) \tag{7.14}$$

for a copula $\hat{C}$, which is known as a survival copula. In the case where $F_1, \ldots, F_d$ are continuous this identity is easily established by noting that

$$\begin{aligned}\bar{F}(x_1, \ldots, x_d) &= P(X_1 > x_1, \ldots, X_d > x_d) \\ &= P(1 - F_1(X_1) \leqslant \bar{F}_1(x_1), \ldots, 1 - F_d(X_d) \leqslant \bar{F}_d(x_d)),\end{aligned}$$

so (7.14) follows by writing $\hat{C}$ for the distribution function of $\mathbf{1} - \boldsymbol{U}$, where $\boldsymbol{U} := (F_1(X_1), \ldots, F_d(X_d))'$ and $\mathbf{1}$ is the vector of ones in $\mathbb{R}^d$. In general, the term *survival copula of a copula* C will be used to denote the df of $\mathbf{1} - \boldsymbol{U}$ when $\boldsymbol{U}$ has df C.

In the case where $F_1, \ldots, F_d$ are continuous and strictly increasing we can give a representation for $\hat{C}$ in (7.14) by setting $x_i = \bar{F}_i^{-1}(u_i)$ for $i = 1, \ldots, d$ to obtain

$$\hat{C}(u_1, \ldots, u_d) = \bar{F}(\bar{F}_1^{-1}(u_1), \ldots, \bar{F}_d^{-1}(u_d)). \tag{7.15}$$

The next example illustrates the derivation of a survival copula from a bivariate survival function using (7.15).

Example 7.14 (survival copula of a bivariate Pareto distribution). A well-known generalization of the important univariate Pareto distribution is the bivariate Pareto distribution with survivor function given by

$$\bar{F}(x_1, x_2) = \left(\frac{x_1 + \kappa_1}{\kappa_1} + \frac{x_2 + \kappa_2}{\kappa_2} - 1\right)^{-\alpha}, \quad x_1, x_2 \geqslant 0,\ \alpha, \kappa_1, \kappa_2 > 0.$$

It is easily confirmed that the marginal survivor functions are given by $\bar{F}_i(x) = (\kappa_i/(\kappa_i + x))^\alpha$, $i = 1, 2$, and we then infer using (7.15) that the survival copula is given by $\hat{C}(u_1, u_2) = (u_1^{-1/\alpha} + u_2^{-1/\alpha} - 1)^{-\alpha}$. Comparison with (7.13) reveals that this is the Clayton copula with parameter $\theta = 1/\alpha$.

The useful concept of *radial symmetry* can be expressed in terms of copulas and survival copulas.

Definition 7.15 (radial symmetry). A random vector $\boldsymbol{X}$ (or its df) is radially symmetric about a point $\boldsymbol{a}$ if $\boldsymbol{X} - \boldsymbol{a} \overset{\mathrm{d}}{=} \boldsymbol{a} - \boldsymbol{X}$.

An elliptical random vector $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ is obviously radially symmetric about $\boldsymbol{\mu}$. If $\boldsymbol{U}$ has df C, where C is a copula, then the only possible centre of symmetry is $(0.5, \ldots, 0.5)$, so C is radially symmetric if

$$(U_1 - 0.5, \ldots, U_d - 0.5) \overset{\mathrm{d}}{=} (0.5 - U_1, \ldots, 0.5 - U_d) \iff \boldsymbol{U} \overset{\mathrm{d}}{=} \mathbf{1} - \boldsymbol{U}.$$

Thus if a copula C is radially symmetric and $\hat{C}$ is its survival copula, we have $\hat{C} = C$. It is easily seen that the copulas of elliptical distributions are radially symmetric but the Gumbel and Clayton copulas are not.

Survival copulas should not be confused with the *survival functions* of copulas, which are not themselves copulas. Since copulas are simply multivariate dfs, they have survival or tail functions, which we denote by $\bar{C}$. If $\boldsymbol{U}$ has df C and the survival copula of C is $\hat{C}$, then

$$\begin{aligned}\bar{C}(u_1, \ldots, u_d) &= P(U_1 > u_1, \ldots, U_d > u_d) \\ &= P(1 - U_1 \leqslant 1 - u_1, \ldots, 1 - U_d \leqslant 1 - u_d) \\ &= \hat{C}(1 - u_1, \ldots, 1 - u_d).\end{aligned}$$

A useful relationship between a copula and its survival copula in the bivariate case is that

$$\hat{C}(1 - u_1, 1 - u_2) = 1 - u_1 - u_2 + C(u_1, u_2). \tag{7.16}$$

Conditional distributions of copulas. It is often of interest to look at conditional distributions of copulas. We concentrate on two dimensions and suppose that (U_1, U_2) has df C. Since a copula is an increasing continuous function in each argument,

$$\begin{aligned}C_{U_2|U_1}(u_2 \mid u_1) = P(U_2 \leqslant u_2 \mid U_1 = u_1) &= \lim_{\delta \to 0} \frac{C(u_1 + \delta, u_2) - C(u_1, u_2)}{\delta} \\ &= \frac{\partial}{\partial u_1} C(u_1, u_2),\end{aligned} \tag{7.17}$$

where this partial derivative exists almost everywhere (see Nelsen (2006) for precise details). The conditional distribution is a distribution on the interval $[0, 1]$ that is only a uniform distribution in the case where C is the independence copula. A risk-management interpretation of the conditional distribution is the following. Suppose continuous risks (X_1, X_2) have the (unique) copula C. Then $1 - C_{U_2|U_1}(q \mid p)$ is the probability that X_2 exceeds its qth quantile given that X_1 attains its pth quantile.

Copula densities. Copulas do not always have joint densities; the comonotonicity and countermonotonicity copulas are examples of copulas that are not absolutely continuous. However, the parametric copulas that we have met so far do have densities given by

$$c(u_1, \ldots, u_d) = \frac{\partial C(u_1, \ldots, u_d)}{\partial u_1 \cdots \partial u_d}, \tag{7.18}$$

and we are sometimes required to calculate them, e.g. if we wish to fit copulas to data by maximum likelihood.

It is useful to note that, for the implicit copula of an absolutely continuous joint df F with strictly increasing, continuous marginal dfs $F_1, \ldots, F_d$, we may differentiate $C(u_1, \ldots, u_d) = F(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d))$ to see that the copula density is given by

$$c(u_1, \ldots, u_d) = \frac{f(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \cdots f_d(F_d^{-1}(u_d))}, \tag{7.19}$$

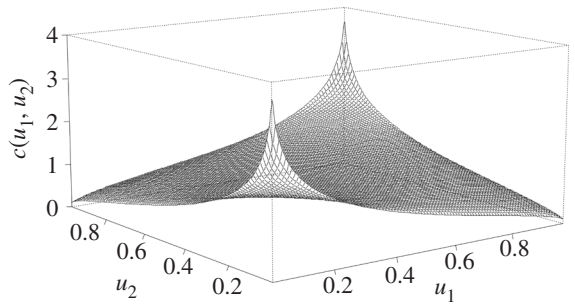

Figure 7.5. Perspective plot of the density of the bivariate Gauss copula with parameter $\rho = 0.3$.

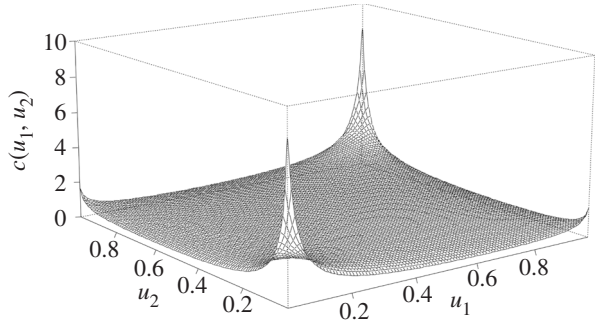

Figure 7.6. Perspective plot of the density of the bivariate t copula with parameters $\nu = 4$ and $\rho = 0.3$.

where f is the joint density of F, $f_1, \ldots, f_d$ are the marginal densities, and $F_1^{-1}, \ldots, F_d^{-1}$ are the ordinary inverses of the marginal dfs.

Using this technique we can calculate the densities of the Gaussian and t copulas as shown in Figures 7.5 and 7.6, respectively. Observe that the t copula assigns much more probability mass to the corners of the unit square; this may be explained by the tail dependence of the t copula, as discussed in Section 7.2.4.

Exchangeability.

Definition 7.16 (exchangeability). A random vector $\boldsymbol{X}$ is exchangeable if

$$(X_1, \ldots, X_d) \stackrel{\mathrm{d}}{=} (X_{\Pi(1)}, \ldots, X_{\Pi(d)})$$

for any permutation $(\Pi(1), \ldots, \Pi(d))$ of $(1, \ldots, d)$.

We will refer to a copula as an *exchangeable* copula if it is the df of an exchangeable random vector of uniform variates $\boldsymbol{U}$. Clearly, for such a copula we must have

$$C(u_1, \ldots, u_d) = C(u_{\Pi(1)}, \ldots, u_{\Pi(d)}) \tag{7.20}$$

for all possible permutations of the arguments of C. Such copulas will prove useful in modelling the default dependence for homogeneous groups of companies in the context of credit risk.

Examples of exchangeable copulas include both the Gumbel and Clayton copulas as well as the Gaussian and t copulas, C_P^{Ga} and $C_{\nu,P}^t$, in the case where P is an *equicorrelation matrix*, i.e. a matrix of the form $P = \rho J_d + (1-\rho) I_d$, where J_d is the square matrix consisting entirely of ones and $\rho \geqslant -1/(d-1)$.

It follows from (7.20) and (7.17) that if the df of the vector (U_1, U_2) is an exchangeable bivariate copula, then

$$P(U_2 \leqslant u_2 \mid U_1 = u_1) = P(U_1 \leqslant u_2 \mid U_2 = u_1), \tag{7.21}$$

which implies quite strong symmetry. If a random vector (X_1, X_2) has such a copula, then the probability that X_2 exceeds its u_2-quantile given that X_1 attains its u_1-quantile is exactly the same as the probability that X_1 exceeds its u_2-quantile given that X_2 attains its u_1-quantile. Not all bivariate copulas must satisfy (7.21). For an example of a non-exchangeable bivariate copula see Section 15.2.2 and Figure 15.4.

Notes and Comments

Sklar's Theorem is first found in Sklar (1959); see also Schweizer and Sklar (1983) or Nelsen (2006) for a proof of the result. The elegant proof using the distributional transform, as mentioned in Remark 7.4, is due to Rüschendorf (2009). A systematic development of the theory of copulas, particularly bivariate ones, with many examples is found in Nelsen (2006). Pitfalls related to discontinuity of marginal distributions are presented in Marshall (1996), and a primer on copulas for discrete count data is given in Genest and Nešlehová (2007). For extensive lists of parametric copula families see Hutchinson and Lai (1990), Joe (1997) and Nelsen (2006).

A reference on copula methods in finance is Cherubini, Luciano and Vecchiato (2004). Embrechts (2009) contains some references to the discussion concerning the pros and cons of copula modelling in insurance and finance.

7.2 Dependence Concepts and Measures

In this section we first provide formal definitions of the concepts of perfect positive and negative dependence (comonotonicity and countermonotonicity) and we present some of the properties of perfectly dependent random vectors.

We then focus on three kinds of dependence measure: the usual Pearson linear correlation, rank correlation, and the coefficients of tail dependence. All of these dependence measures yield a *scalar measurement* of the "strength of the dependence" for a pair of rvs (X_1, X_2), although the nature and properties of the measure are different in each case.

The rank correlations and tail-dependence coefficients are *copula-based* dependence measures. In contrast to ordinary correlation, these measures are functions of the copula only and can thus be used in the parametrization of copulas, as discussed in Section 7.5.

7.2.1 Perfect Dependence

We define the concepts of perfect positive and perfect negative dependence using the fundamental copulas of Section 7.1.2. Alternative names for these concepts are comonotonicity and countermonotonicity, respectively.

Comonotonicity. This concept may be defined in a number of equivalent ways. We give a copula-based definition and then derive alternative representations that show that comonotonic random variables can be thought of as undiversifiable random variables.

Definition 7.17 (comonotonicity). The rvs $X_1, \dots, X_d$ are said to be comonotonic if they admit as copula the Fréchet upper bound $M(u_1, \dots, u_d) = \min(u_1, \dots, u_d)$.

The following result shows that comonotonic rvs are monotonically increasing functions of a single rv. In other words, there is a single source of risk and the comonotonic variables move deterministically in lockstep with that risk.

Proposition 7.18. *$X_1, \dots, X_d$ are comonotonic if and only if*

$$(X_1, \dots, X_d) \stackrel{\mathrm{d}}{=} (v_1(Z), \dots, v_d(Z)) \tag{7.22}$$

for some rv Z and increasing functions $v_1, \dots, v_d$.

Proof. Assume that $X_1, \dots, X_d$ are comonotonic according to Definition 7.2.1. Let U be any uniform rv and write $F, F_1, \dots, F_d$ for the joint df and marginal dfs of $X_1, \dots, X_d$, respectively. From (7.2) we have

$$\begin{aligned} F(x_1, \dots, x_d) &= \min(F_1(x_1), \dots, F_d(x_d)) \\ &= P(U \leqslant \min(F_1(x_1), \dots, F_d(x_d))) \\ &= P(U \leqslant F_1(x_1), \dots, U \leqslant F_d(x_d)) \\ &= P(F_1^{\leftarrow}(U) \leqslant x_1, \dots, F_d^{\leftarrow}(U) \leqslant x_d) \end{aligned}$$

for any $U \sim U(0, 1)$, where we use Proposition A.3 (iv) in the last equality. It follows that

$$(X_1, \dots, X_d) \stackrel{\mathrm{d}}{=} (F_1^{\leftarrow}(U), \dots, F_d^{\leftarrow}(U)), \tag{7.23}$$

which is of the form (7.22). Conversely, if (7.22) holds, then

$$F(x_1, \dots, x_d) = P(v_1(Z) \leqslant x_1, \dots, v_d(Z) \leqslant x_d) = P(Z \in A_1, \dots, Z \in A_d),$$

where each A_i is an interval of the form $(-\infty, k_i]$ or $(-\infty, k_i)$, so one interval A_i is a subset of all other intervals. Therefore,

$$\begin{aligned} F(x_1, \dots, x_d) &= \min(P(Z \in A_1), \dots, P(Z \in A_d)) \\ &= \min(F_1(x_1), \dots, F_d(x_d)), \end{aligned}$$

which proves comonotonicity. □

In the case of rvs with continuous marginal distributions we have a simpler and stronger result.

Corollary 7.19. *Let $X_1, \ldots, X_d$ be rvs with continuous dfs. They are comonotonic if and only if for every pair (i, j) we have $X_j = T_{ji}(X_i)$ almost surely for some increasing transformation T_{ji}.*

Proof. The result follows from the proof of Proposition 7.18 by noting that the rv U may be taken to be $F_i(X_i)$ for any i. Without loss of generality set $d = 2$ and $i = 1$ and use (7.23) and Proposition A.4 to obtain

$$(X_1, X_2) \stackrel{\mathrm{d}}{=} (F_1^{\leftarrow} \circ F_1(X_1), F_2^{\leftarrow} \circ F_1(X_1)) \stackrel{\mathrm{d}}{=} (X_1, F_2^{\leftarrow} \circ F_1(X_1)).$$

□

Comonotone additivity of quantiles. A very important result for comonotonic rvs is the additivity of the quantile function as shown in the following proposition. In addition to the VaR risk measure, the property of so-called comonotone additivity will be shown to apply to a class of risk measures known as distortion risk measures in Section 8.2.1; this class includes expected shortfall.

Proposition 7.20. *Let $0 < \alpha < 1$ and $X_1, \ldots, X_d$ be comonotonic rvs with dfs $F_1, \ldots, F_d$. Then*

$$F^{\leftarrow}_{X_1+\cdots+X_d}(\alpha) = F_1^{\leftarrow}(\alpha) + \cdots + F_d^{\leftarrow}(\alpha). \tag{7.24}$$

Proof. For ease of notation take $d = 2$. From Proposition 7.18 we have that $(X_1, X_2) \stackrel{\mathrm{d}}{=} (F_1^{\leftarrow}(U), F_2^{\leftarrow}(U))$ for some $U \sim U(0, 1)$. It follows that

$$F^{\leftarrow}_{X_1+X_2}(\alpha) = F^{\leftarrow}_{T(U)}(\alpha),$$

where T is the increasing left-continuous function given by $T(x) = F_1^{\leftarrow}(x) + F_2^{\leftarrow}(x)$. The result follows by applying Proposition A.5 to get

$$F^{\leftarrow}_{T(U)}(\alpha) = T(F_U^{\leftarrow}(\alpha)) = T(\alpha) = F_1^{\leftarrow}(\alpha) + F_2^{\leftarrow}(\alpha).$$

□

Countermonotonicity. In an analogous manner to the way we have defined comonotonicity, we define countermonotonicity as a copula concept, albeit restricted to the case $d = 2$.

Definition 7.21 (countermonotonicity). The rvs X_1 and X_2 are countermonotonic if they have as copula the Fréchet lower bound $W(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$.

Proposition 7.22. *X_1 and X_2 are countermonotonic if and only if*

$$(X_1, X_2) \stackrel{\mathrm{d}}{=} (v_1(Z), v_2(Z))$$

for some rv Z with v_1 increasing and v_2 decreasing, or vice versa.

Proof. The proof is similar to that of Proposition 7.18 and is given in Embrechts, McNeil and Straumann (2002). □

Remark 7.23. In the case where X_1 and X_2 are continuous we have the simpler result that countermonotonicity is equivalent to $X_2 = T(X_1)$ almost surely for some decreasing function T.

The concept of countermonotonicity does not generalize to higher dimensions. The Fréchet lower bound $W(u_1, \ldots, u_d)$ is not itself a copula for $d > 2$ since it is not a proper distribution function and does not satisfy (7.1), as the following example taken from Nelsen (2006, Exercise 2.36) shows.

Example 7.24 (the Fréchet lower bound is not a copula for $d > 2$). Consider the d-cube $[\frac{1}{2}, 1]^d \subset [0, 1]^d$. If the Fréchet lower bound for copulas were a df on $[0, 1]^d$, then (7.1) would imply that the probability mass $P(d)$ of this cube would be given by

$$\begin{aligned} P(d) &= \max(1 + \cdots + 1 - d + 1, 0) - d \max(\tfrac{1}{2} + 1 + \cdots + 1 - d + 1, 0) \\ &\quad + \binom{d}{2} \max(\tfrac{1}{2} + \tfrac{1}{2} + \cdots + 1 - d + 1, 0) - \cdots \\ &\quad + \max(\tfrac{1}{2} + \cdots + \tfrac{1}{2} - d + 1, 0) \\ &= 1 - \tfrac{1}{2}d. \end{aligned}$$

Hence the Fréchet lower bound cannot be a copula for $d > 2$.

Some additional insight into the impossibility of countermonotonicity for dimensions higher than two is given by the following simple example.

Example 7.25. Let X_1 be a positive-valued rv and take $X_2 = 1/X_1$ and $X_3 = \mathrm{e}^{-X_1}$. Clearly, (X_1, X_2) and (X_1, X_3) are countermonotonic random vectors. However, (X_2, X_3) is comonotonic and the copula of the vector (X_1, X_2, X_3) is the df of the vector $(U, 1 - U, 1 - U)$, which may be calculated to be

$$C(u_1, u_2, u_3) = \max(\min(u_2, u_3) + u_1 - 1, 0).$$

7.2.2 Linear Correlation

Correlation plays a central role in financial theory, but it is important to realize that the concept is only really a natural one in the context of multivariate normal or, more generally, elliptical models. As we have seen, elliptical distributions are fully described by a mean vector, a covariance matrix and a characteristic generator function. Since means and variances are features of marginal distributions, the copulas of elliptical distributions can be thought of as depending only on the correlation matrix and characteristic generator; the correlation matrix thus has a natural parametric role in these models, which it does not have in more general multivariate models. Our discussion of correlation will focus on the shortcomings of correlation and the subtle pitfalls that the naive user of correlation may encounter when moving away from elliptical models. The concept of copulas will help us to illustrate these pitfalls.

The correlation $\rho(X_1, X_2)$ between rvs X_1 and X_2 was defined in (6.3). It is a measure of *linear* dependence and takes values in $[-1, 1]$. If X_1 and X_2 are independent, then $\rho(X_1, X_2) = 0$, but it is important to be clear that the converse is false:

the uncorrelatedness of X_1 and X_2 does not in general imply their independence. Examples are provided by the class of uncorrelated normal mixture distributions (see Lemma 6.5) and the class of spherical distributions (with the single exception of the multivariate normal). For an even simpler example, we can take $X_1 = Z \sim N(0, 1)$ and $X_2 = Z^2$; these are clearly dependent rvs but have zero correlation.

If $|\rho(X_1, X_2)| = 1$, then this is equivalent to saying that X_2 and X_1 are *perfectly linearly dependent*, meaning that $X_2 = \alpha + \beta X_1$ almost surely for some $\alpha \in \mathbb{R}$ and $\beta \neq 0$, with $\beta > 0$ for positive linear dependence and $\beta < 0$ for negative linear dependence. Moreover, for $\beta_1, \beta_2 > 0$,

$$\rho(\alpha_1 + \beta_1 X_1, \alpha_2 + \beta_2 X_2) = \rho(X_1, X_2),$$

so correlation is invariant under strictly increasing *linear* transformations. However, correlation is *not* invariant under nonlinear strictly increasing transformations $T : \mathbb{R} \to \mathbb{R}$. For two real-valued rvs we have, in general, $\rho(T(X_1), T(X_2)) \neq \rho(X_1, X_2)$.

Another obvious, but important, remark is that correlation is only defined when the variances of X_1 and X_2 are finite. This restriction to finite-variance models is not ideal for a dependence measure and can cause problems when we work with heavy-tailed distributions. For example, actuaries who model losses in different business lines with infinite-variance distributions may not describe the dependence of their risks using correlation. We will encounter similar examples in Section 13.1.4 on operational risk.

Correlation fallacies. We now discuss further pitfalls in the use of correlation, which we present in the form of fallacies. We believe these fallacies are worth highlighting because they illustrate the dangers of attempting to construct multivariate risk models starting from marginal distributions and estimates of the correlations between risks. The statements we make are true if we restrict our attention to elliptically distributed risk factors, but they are false in general. For background to these fallacies, alternative examples and a discussion of the relevance to multivariate Monte Carlo simulation, see Embrechts, McNeil and Straumann (2002).

Fallacy 1. The marginal distributions and pairwise correlations of a random vector determine its joint distribution.

It should already be clear to readers of this chapter that this is not true. Figure 7.4 shows the key to constructing counterexamples. Suppose the rvs X_1 and X_2 have continuous marginal distributions F_1 and F_2 and joint df $C(F_1(x_1), F_2(x_2))$ for some copula C, and suppose their linear correlation is $\rho(X_1, X_2) = \rho$. It will generally be possible to find an alternative copula $C_2 \neq C$ and to construct a random vector (Y_1, Y_2) with df $C_2(F_1(x_1), F_2(x_2))$ such that $\rho(Y_1, Y_2) = \rho$. The following example illustrates this idea in a case where $\rho = 0$.

Example 7.26. Consider two rvs representing profits and losses on two portfolios. Suppose we are given the information that both risks have standard normal distributions and that their correlation is 0. We construct two random vectors that are consistent with this information.

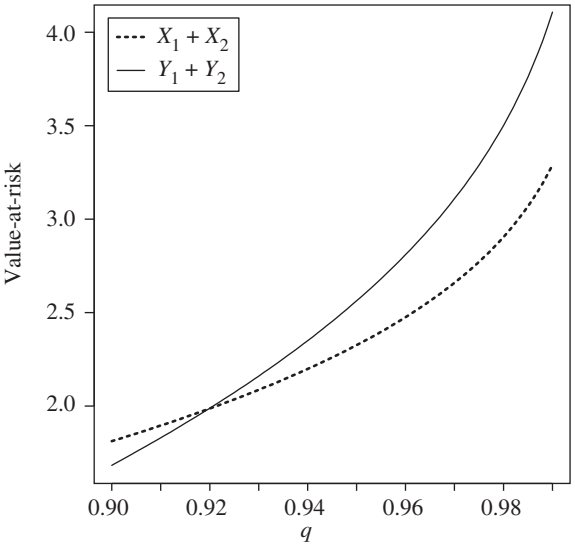

Figure 7.7. VaR for the risks $X_1 + X_2$ and $Y_1 + Y_2$ as described in Example 7.26. Both these pairs have standard normal margins and a correlation of zero; X_1 and X_2 are independent, whereas Y_1 and Y_2 are dependent.

Model 1 is the standard bivariate normal $\boldsymbol{X} \sim N_2(\mathbf{0}, I_2)$. Model 2 is constructed by taking V to be an independent discrete rv such that $P(V = 1) = P(V = -1) = 0.5$ and setting $(Y_1, Y_2) = (X_1, VX_1)$ with X_1 as in Model 1. This model obviously also has normal margins and correlation zero; its copula is given by

$$C(u_1, u_2) = 0.5 \max(u_1 + u_2 - 1, 0) + 0.5 \min(u_1, u_2),$$

which is a mixture of the two-dimensional comonotonicity and countermonotonicity copulas. This could be roughly interpreted as representing two equiprobable states of the world: in one state financial outcomes in the two portfolios are comonotonic and we are certain to make money in both or lose money in both; in the other state they are countermonotonic and we will make money in one and lose money in the other.

We can calculate analytically the distribution of the total losses $X_1 + X_2$ and $Y_1 + Y_2$; the latter sum does not itself have a univariate normal distribution. For $k \geqslant 0$ we get that

$$P(X_1 + X_2 > k) = \bar{\Phi}(k/\sqrt{2}), \qquad P(Y_1 + Y_2 > k) = \tfrac{1}{2}\bar{\Phi}(\tfrac{1}{2}k),$$

from which it follows that, for $\alpha > 0.75$,

$$F^{\leftarrow}_{X_1+X_2}(\alpha) = \sqrt{2}\Phi^{-1}(\alpha), \qquad F^{\leftarrow}_{Y_1+Y_2}(\alpha) = 2\Phi^{-1}(2\alpha - 1).$$

In Figure 7.7 we see that the quantile of $Y_1 + Y_2$ dominates that of $X_1 + X_2$ for probability levels above 93%. This example also illustrates that the VaR of a sum of risks is clearly not determined by marginal distributions and pairwise correlations.

The correlation of two risks does not only depend on their copula—if it did, then Proposition 7.7 would imply that correlation would be invariant under strictly increasing transformations, which is not the case. Correlation is also inextricably linked to the marginal distributions of the risks, and this imposes certain constraints on the values that correlation can take. This is the subject of the second fallacy.

Fallacy 2. For given univariate distributions F_1 and F_2 and any correlation value ρ in $[-1, 1]$ it is always possible to construct a joint distribution F with margins F_1 and F_2 and correlation ρ.

Again, this statement is true if F_1 and F_2 are the margins of an elliptical distribution, but is in general false. The so-called *attainable* correlations can form a strict subset of the interval $[-1, 1]$, as is shown in the next theorem. In the proof of the theorem we require the formula of Höffding, which is given in the next lemma.

Lemma 7.27. *If (X_1, X_2) has joint df F and marginal dfs F_1 and F_2, then the covariance of X_1 and X_2, when finite, is given by*

$$\operatorname{cov}(X_1, X_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (F(x_1, x_2) - F_1(x_1)F_2(x_2))\,\mathrm{d}x_1\,\mathrm{d}x_2. \tag{7.25}$$

Proof. Let (X_1, X_2) have df F and let $(\tilde{X}_1, \tilde{X}_2)$ be an *independent copy* (i.e. a second pair with df F independent of (X_1, X_2)). We have

$$2\operatorname{cov}(X_1, X_2) = E((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2)).$$

We now use a useful identity that says that, for any $a \in \mathbb{R}$ and $b \in \mathbb{R}$, we can always write $(a - b) = \int_{-\infty}^{\infty}(I_{\{b \leqslant x\}} - I_{\{a \leqslant x\}})\,\mathrm{d}x$, and we apply this to the random pairs $(X_1 - \tilde{X}_1)$ and $(X_2 - \tilde{X}_2)$. We obtain

$$\begin{aligned}
&2\operatorname{cov}(X_1, X_2)\\
&\quad = E\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(I_{\{\tilde{X}_1 \leqslant x_1\}} - I_{\{X_1 \leqslant x_1\}})(I_{\{\tilde{X}_2 \leqslant x_2\}} - I_{\{X_2 \leqslant x_2\}})\,\mathrm{d}x_1\,\mathrm{d}x_2\right)\\
&\quad = 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(P(X_1 \leqslant x_1, X_2 \leqslant x_2) - P(X_1 \leqslant x_1)P(X_2 \leqslant x_2))\,\mathrm{d}x_1\,\mathrm{d}x_2.
\end{aligned}$$

□

Theorem 7.28 (attainable correlations). *Let (X_1, X_2) be a random vector with finite-variance marginal dfs F_1 and F_2 and an unspecified joint df; assume also that $\operatorname{var}(X_1) > 0$ and $\operatorname{var}(X_2) > 0$. The following statements hold.*

(1) *The attainable correlations form a closed interval $[\rho_{\min}, \rho_{\max}]$ with $\rho_{\min} < 0 < \rho_{\max}$.*

(2) *The minimum correlation $\rho = \rho_{\min}$ is attained if and only if X_1 and X_2 are countermonotonic. The maximum correlation $\rho = \rho_{\max}$ is attained if and only if X_1 and X_2 are comonotonic.*

(3) *$\rho_{\min} = -1$ if and only if X_1 and $-X_2$ are of the same type (see Section A.1.1), and $\rho_{\max} = 1$ if and only if X_1 and X_2 are of the same type.*

Proof. We begin with (2) and use the identity (7.25). We also recall the two-dimensional Fréchet bounds for a general df in (7.6):

$$\max(F_1(x_1) + F_2(x_2) - 1, 0) \leqslant F(x_1, x_2) \leqslant \min(F_1(x_1), F_2(x_2)).$$

Clearly, when F_1 and F_2 are fixed, the integrand in (7.25) is maximized pointwise when X_1 and X_2 have the Fréchet upper bound copula $C(u_1, u_2) = \min(u_1, u_2)$, i.e. when they are comonotonic. Similarly, the integrand is minimized when X_1 and X_2 are countermonotonic.

To complete the proof of (1), note that clearly $\rho_{\max} \geqslant 0$. However, $\rho_{\max} = 0$ can be ruled out since this would imply that $\min(F_1(x_1), F_2(x_2)) = F_1(x_1)F_2(x_2)$ for all x_1, x_2. This can only occur if F_1 or F_2 is a degenerate distribution consisting of point mass at a single point, but this is excluded by the assumption that variances are non-zero. By a similar argument we have that $\rho_{\min} < 0$. If $W(F_1, F_2)$ and $M(F_1, F_2)$ denote the Fréchet lower and upper bounds, respectively, then the mixture $\lambda W(F_1, F_2) + (1-\lambda)M(F_1, F_2)$, $0 \leqslant \lambda \leqslant 1$, has correlation $\lambda\rho_{\min} + (1-\lambda)\rho_{\max}$. Thus for any $\rho \in [\rho_{\min}, \rho_{\max}]$ we can set $\lambda = (\rho_{\max} - \rho)/(\rho_{\max} - \rho_{\min})$ to construct a joint df that attains the correlation value ρ.

Part (3) is clear since $\rho_{\min} = -1$ or $\rho_{\max} = 1$ if and only if there is an almost sure linear relationship between X_1 and X_2. □

Example 7.29 (attainable correlations for lognormal rvs). An example where the maximal and minimal correlations can be easily calculated occurs when $\ln X_1 \sim N(0, 1)$ and $\ln X_2 \sim N(0, \sigma^2)$. For $\sigma \neq 1$ the lognormally distributed rvs X_1 and X_2 are not of the same type (although $\ln X_1$ and $\ln X_2$ are) so that, by part (3) of Theorem 7.28, we have $\rho_{\max} < 1$. The rvs X_1 and $-X_2$ are also not of the same type, so $\rho_{\min} > -1$.

To calculate the actual boundaries of the attainable interval let $Z \sim N(0, 1)$ and observe that if X_1 and X_2 are comonotonic, then $(X_1, X_2) \overset{\mathrm{d}}{=} (\mathrm{e}^Z, \mathrm{e}^{\sigma Z})$. Clearly, $\rho_{\max} = \rho(\mathrm{e}^Z, \mathrm{e}^{\sigma Z})$ and, by a similar argument, $\rho_{\min} = \rho(\mathrm{e}^Z, \mathrm{e}^{-\sigma Z})$. The analytical calculation now follows easily and yields

$$\rho_{\min} = \frac{\mathrm{e}^{-\sigma} - 1}{\sqrt{(\mathrm{e} - 1)(\mathrm{e}^{\sigma^2} - 1)}}, \qquad \rho_{\max} = \frac{\mathrm{e}^{\sigma} - 1}{\sqrt{(\mathrm{e} - 1)(\mathrm{e}^{\sigma^2} - 1)}}.$$

See Figure 7.8 for an illustration of the attainable correlation interval for different values of σ and note how the boundaries of the interval both tend rapidly to zero as σ is increased. This shows, for example, that we can have situations where comonotonic rvs have very small correlation values. Since comonotonicity is the strongest form of positive dependence, this provides a correction to the widely held view that small correlations always imply weak dependence.

Fallacy 3. For rvs $X_1 \sim F_1$ and $X_2 \sim F_2$ and for given α, the quantile of the sum $F^{\leftarrow}_{X_1+X_2}(\alpha)$ is maximized when the joint distribution F has maximal correlation.

While once again this is true if (X_1, X_2) are jointly elliptical, the statement is not true in general and any example of the superadditivity of the quantile function (or VaR risk measure) yields a counterexample.

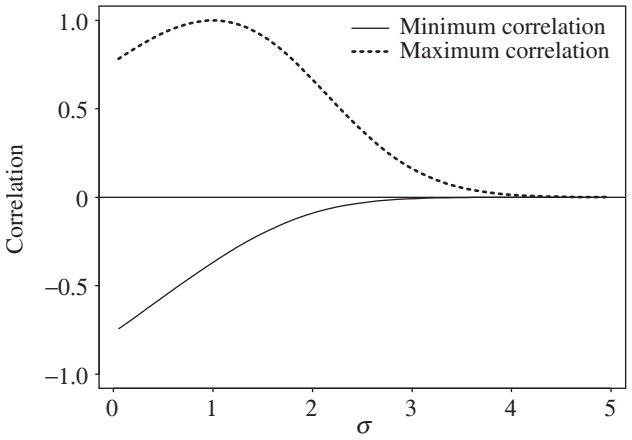

Figure 7.8. Maximum and minimum attainable correlations for lognormal rvs X_1 and X_2, where $\ln X_1$ is standard normal and $\ln X_2$ is normal with mean 0 and variance σ^2.

In a superadditive VaR example, we have, for some value of α, that

$$F^{\leftarrow}_{X_1+X_2}(\alpha) > F^{\leftarrow}_{X_1}(\alpha) + F^{\leftarrow}_{X_2}(\alpha), \tag{7.26}$$

but, by Proposition 7.20, the right-hand side of this inequality is equal to $F^{\leftarrow}_{Y_1+Y_2}(\alpha)$ for a pair of comonotonic rvs (Y_1, Y_2) with $Y_1 \overset{\mathrm{d}}{=} X_1$ and $Y_2 \overset{\mathrm{d}}{=} X_2$. Moreover, by part (2) of Theorem 7.28, (Y_1, Y_2) will attain the maximal correlation $\rho_{\max}$ and $\rho_{\max} > \rho(X_1, X_2)$. A simple example where this occurs is as follows.

Example 7.30. Let $X_1 \sim \mathrm{Exp}(1)$ and $X_2 \sim \mathrm{Exp}(1)$ be two independent standard exponential rvs. Let $Y_1 = Y_2 = X_1$ and take $\alpha = 0.7$. Since $X_1 + X_2 \sim \mathrm{Ga}(2, 1)$ (see Appendix A.2.4) it is easily checked that

$$F^{\leftarrow}_{X_1+X_2}(\alpha) > F^{\leftarrow}_{X_1}(\alpha) + F^{\leftarrow}_{X_2}(\alpha) = F^{\leftarrow}_{Y_1+Y_2}(\alpha)$$

but $\rho(X_1, X_2) = 0$ and $\rho(Y_1, Y_2) = 1$. This example is also discussed in Section 8.3.3.

In Section 8.4.4 we will look at the problem of discovering how "bad" the quantile of the sum of the two risks in (7.26) can be when the marginal distributions are known.

A common message can be extracted from the fallacies of this section: namely that the concept of correlation is meaningless unless applied in the context of a well-defined joint model. Any interpretation of correlation values in the absence of such a model should be avoided.

7.2.3 Rank Correlation

Rank correlations are simple scalar measures of dependence that depend only on the copula of a bivariate distribution and not on the marginal distributions, unlike linear correlation, which depends on both. The standard empirical estimators of rank correlation may be calculated by looking at the *ranks* of the data alone, hence the

name. In other words, we only need to know the ordering of the sample for each variable of interest and not the actual numerical values.

The main practical reason for looking at rank correlations is that they can be used to calibrate copulas to empirical data. At a theoretical level, being direct functionals of the copula, rank correlations have more appealing properties than linear correlations, as is discussed below. There are two main varieties of rank correlation, Kendall's and Spearman's, and both can be understood as a measure of concordance for bivariate random vectors. Two points in $\mathbb{R}^2$, denoted by (x_1, x_2) and $(\tilde{x}_1, \tilde{x}_2)$, are said to be *concordant* if $(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) > 0$ and to be *discordant* if $(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) < 0$.

Kendall's tau. Consider a random vector (X_1, X_2) and an independent copy $(\tilde{X}_1, \tilde{X}_2)$ (i.e. a second vector with the same distribution, but independent of the first). If X_2 tends to increase with X_1, then we expect the probability of concordance to be high relative to the probability of discordance; if X_2 tends to decrease with increasing X_1, then we expect the opposite. This motivates Kendall's rank correlation, which is simply the probability of concordance minus the probability of discordance for these pairs:

$$\rho_\tau(X_1, X_2) = P((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0) - P((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) < 0). \quad (7.27)$$

It is easily seen that there is a more compact way of writing this as an expectation, which also leads to an obvious estimator in Section 7.5.1.

Definition 7.31. For rvs X_1 and X_2 Kendall's tau is given by

$$\rho_\tau(X_1, X_2) = E(\text{sign}((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2))),$$

where $(\tilde{X}_1, \tilde{X}_2)$ is an independent copy of (X_1, X_2) and $\text{sign}(x) = I_{\{x>0\}} - I_{\{x<0\}}$.

In higher dimensions the Kendall's tau matrix of a random vector $\boldsymbol{X}$ may be written as $\rho_\tau(\boldsymbol{X}) = \text{cov}(\boldsymbol{Y})$, where $\boldsymbol{Y} = \text{sign}(\boldsymbol{X} - \tilde{\boldsymbol{X}})$ and $\tilde{\boldsymbol{X}}$ is an independent copy of $\boldsymbol{X}$; note that $\boldsymbol{Y}$ is obtained by the componentwise application of the sign function, so that $\boldsymbol{Y} = (Y_1, \ldots, Y_d)'$, where $Y_i = \text{sign}(X_i - \tilde{X}_i)$ for $i = 1, \ldots, d$. Since $\rho_\tau(\boldsymbol{X})$ can be expressed as the covariance matrix of $\boldsymbol{Y}$, it is obviously positive semidefinite.

We now show that, for random variables with continuous dfs, Kendall's tau depends only on the unique copula C of (X_1, X_2) and we give an explicit formula for computing ρ_τ from C.

Proposition 7.32. *Suppose X_1 and X_2 have continuous marginal distributions and unique copula C. Then*

$$\rho_\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2)\, \mathrm{d}C(u_1, u_2) - 1. \quad (7.28)$$

Proof. Starting from (7.27) and writing F_1 and F_2 for the marginals dfs, we have

$$\begin{aligned} \rho_\tau(X_1, X_2) &= 2P((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0) - 1 \\ &= 4P(X_1 < \tilde{X}_1, X_2 < \tilde{X}_2) - 1 \quad (7.29) \\ &= 4P(F_1(X_1) < F_1(\tilde{X}_1), F_2(X_2) < F_2(\tilde{X}_2)) - 1, \end{aligned}$$

where the second equality follows from the interchangeability of the pairs (X_1, X_2) and $(\tilde{X}_1, \tilde{X}_2)$ and the third equality from the continuity of F_1 and F_2. Introducing the uniform random variables $U_i = F_i(X_i)$ and $\tilde{U}_i = F_i(\tilde{X}_i)$, $i = 1, 2$, and noting that the df of the pairs (U_1, U_2) and $(\tilde{U}_1, \tilde{U}_2)$ is C, we obtain

$$\begin{aligned}\rho_\tau(X_1, X_2) &= 4E(P(U_1 < \tilde{U}_1, U_2 < \tilde{U}_2 \mid \tilde{U}_1, \tilde{U}_2)) - 1 \\ &= 4\int_0^1 \int_0^1 P(U_1 < u_1, U_2 < u_2)\, \mathrm{d}C(u_1, u_2) - 1,\end{aligned}$$

from which (7.28) follows. □

Spearman's rho. This measure can also be defined in terms of the concordance and discordance of two bivariate random vectors, but this time we consider the independent pairs (X_1, X_2) and $(\tilde{X}_1, \bar{X}_2)$ and assume that they have identical marginal distributions but that the second pair is a pair of *independent* random variables.

Definition 7.33. For rvs X_1 and X_2 Spearman's rho is given by

$$\rho_S(X_1, X_2) = 3(P((X_1 - \tilde{X}_1)(X_2 - \bar{X}_2) > 0) - P((X_1 - \tilde{X}_1)(X_2 - \bar{X}_2) < 0)), \tag{7.30}$$

where $\tilde{X}_1$ and $\bar{X}_2$ are random variables satisfying $\tilde{X}_1 \overset{\mathrm{d}}{=} X_1$ and $\bar{X}_2 \overset{\mathrm{d}}{=} X_2$ and where (X_1, X_2), $\tilde{X}_1$ and $\bar{X}_2$ are all independent.

It is not immediately apparent that this definition gives a sensible correlation measure, i.e. a number in the interval $[-1, 1]$. The following proposition gives an alternative representation, which makes this clear for continuous random variables.

Proposition 7.34. *If X_1 and X_2 have continuous marginal distributions F_1 and F_2, then $\rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2))$.*

Proof. If the random vectors (X_1, X_2) and $(\tilde{X}_1, \bar{X}_2)$ have continuous marginal distributions, we may write

$$\begin{aligned}\rho_S(X_1, X_2) &= 6P((X_1 - \tilde{X}_1)(X_2 - \bar{X}_2) > 0) - 3 \qquad (7.31)\\ &= 6P((F_1(X_1) - F_1(\tilde{X}_1))(F_2(X_2) - F_2(\bar{X}_2)) > 0) - 3 \\ &= 6P((U_1 - \tilde{U}_1)(U_2 - \bar{U}_2) > 0) - 3,\end{aligned}$$

where $(U_1, U_2) := (F_1(X_1), F_2(X_2))$, $\tilde{U}_1 := F_1(\tilde{X}_1)$ and $\bar{U}_2 = F_2(\bar{X}_2)$ and U_1, U_2, $\tilde{U}_1$ and $\bar{U}_2$ all have standard uniform distributions. Conditioning on U_1 and U_2 we obtain

$$\begin{aligned}\rho_S(X_1, X_2) &= 6E(P((U_1 - \tilde{U}_1)(U_2 - \bar{U}_2) > 0 \mid U_1, U_2)) - 3 \\ &= 6E(P(\tilde{U}_1 < U_1, \bar{U}_2 < U_2 \mid U_1, U_2) \\ &\qquad + P(\tilde{U}_1 > U_1, \bar{U}_2 > U_2 \mid U_1, U_2)) - 3 \\ &= 6E(U_1U_2 + (1 - U_1)(1 - U_2)) - 3 \\ &= 12E(U_1U_2) - 6E(U_1) - 6E(U_2) + 3 \\ &= 12\,\mathrm{cov}(U_1, U_2), \qquad (7.32)\end{aligned}$$

where we have used the fact that $E(U_1) = E(U_2) = \frac{1}{2}$. The conclusion that $\rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2))$ follows from noting that $\text{var}(U_1) = \text{var}(U_2) = \frac{1}{12}$. □

In other words, Spearman's rho, for continuous random variables, is simply the linear correlation of their unique copula. The Spearman's rho matrix for the general multivariate random vector $\boldsymbol{X}$ with continuous margins is given by $\rho_S(\boldsymbol{X}) = \rho(F_1(X_1), \ldots, F_d(X_d))$ and must be positive semidefinite. In the bivariate case, the formula for Spearman's rho in terms of the copula C of (X_1, X_2) follows from a simple application of Höffding's formula (7.25) to formula (7.32).

Corollary 7.35. *Suppose X_1 and X_2 have continuous marginal distributions and unique copula C. Then*

$$\rho_S(X_1, X_2) = 12 \int_0^1 \int_0^1 (C(u_1, u_2) - u_1 u_2)\, \mathrm{d}u_1\, \mathrm{d}u_2. \tag{7.33}$$

Properties of rank correlation. Kendall's tau and Spearman's rho have many properties in common. They are both symmetric dependence measures taking values in the interval $[-1, 1]$. They give the value zero for independent rvs, although a rank correlation of 0 does not necessarily imply independence. It can be shown that they take the value 1 if and only if X_1 and X_2 are comonotonic (see Embrechts, McNeil and Straumann 2002) and the value -1 if and only if they are countermonotonic (which contrasts with the behaviour of linear correlation observed in Theorem 7.28). They are invariant under strictly increasing transformations of X_1 and X_2.

To what extent do the fallacies of linear correlation identified in Section 7.2.2 carry over to rank correlation? Clearly, Fallacy 1 remains relevant: marginal distributions and pairwise rank correlations do not fully determine the joint distribution of a vector of risks. Fallacy 3 also still applies when we switch from linear to rank correlations; although two comonotonic risks will have the maximum rank correlation value of one, this does not imply that the quantile of their sum is maximized over the class of all joint models with the same marginal distributions.

However, Fallacy 2 no longer applies when we consider rank correlations: for any choice of continuous marginal distributions it is possible to specify a bivariate distribution that has any desired rank correlation value in $[-1, 1]$. One way of doing this is to take a convex combination of the form

$$F(x_1, x_2) = \lambda W(F_1(x_1), F_2(x_2)) + (1 - \lambda) M(F_1(x_1), F_2(x_2)),$$

where W and M are the countermonotonicity and comonotonicity copulas, respectively. A random pair (X_1, X_2) with this df has rank correlation

$$\rho_\tau(X_1, X_2) = \rho_S(X_1, X_2) = 1 - 2\lambda,$$

which yields any desired value in $[-1, 1]$ for an appropriate choice of λ in $[0, 1]$. But this is only one of many possible constructions; a model with the Gauss copula of the form $F(x_1, x_2) = C^{\text{Ga}}_\rho(F_1(x_1), F_2(x_2))$ can also be parametrized by an appropriate choice of $\rho \in [-1, 1]$ to have any rank correlation in $[-1, 1]$. In Section 7.3.2 we will calculate Spearman's rho and Kendall's tau values for the Gauss copula and other copulas of normal variance mixture distributions.

7.2.4 Coefficients of Tail Dependence

Like the rank correlations, the coefficients of tail dependence are measures of pairwise dependence that depend only on the copula of a pair of rvs X_1 and X_2 with continuous marginal dfs. The motivation for looking at these coefficients is that they provide measures of *extremal dependence* or, in other words, measures of the strength of dependence in the tails of a bivariate distribution. The coefficients we describe are defined in terms of limiting conditional probabilities of *quantile exceedances*. We note that there are a number of other definitions of tail-dependence measures in the literature (see Notes and Comments).

In the case of upper tail dependence we look at the probability that X_2 exceeds its q-quantile, given that X_1 exceeds its q-quantile, and then consider the limit as q goes to one. Obviously the roles of X_1 and X_2 are interchangeable. Formally we have the following.

Definition 7.36. Let X_1 and X_2 be rvs with dfs F_1 and F_2. The coefficient of upper tail dependence of X_1 and X_2 is

$$\lambda_u := \lambda_u(X_1, X_2) = \lim_{q \to 1^-} P(X_2 > F_2^{\leftarrow}(q) \mid X_1 > F_1^{\leftarrow}(q)),$$

provided a limit $\lambda_u \in [0, 1]$ exists. If $\lambda_u \in (0, 1]$, then X_1 and X_2 are said to show upper tail dependence or extremal dependence in the upper tail; if $\lambda_u = 0$, they are *asymptotically independent* in the upper tail. Analogously, the coefficient of lower tail dependence is

$$\lambda_l := \lambda_l(X_1, X_2) = \lim_{q \to 0^+} P(X_2 \leqslant F_2^{\leftarrow}(q) \mid X_1 \leqslant F_1^{\leftarrow}(q)),$$

provided a limit $\lambda_l \in [0, 1]$ exists.

If F_1 and F_2 are continuous dfs, then we get simple expressions for λ_l and λ_u in terms of the unique copula C of the bivariate distribution. Using elementary conditional probability and (7.4) we have

$$\begin{aligned} \lambda_l &= \lim_{q \to 0^+} \frac{P(X_2 \leqslant F_2^{\leftarrow}(q), X_1 \leqslant F_1^{\leftarrow}(q))}{P(X_1 \leqslant F_1^{\leftarrow}(q))} \\ &= \lim_{q \to 0^+} \frac{C(q, q)}{q}. \end{aligned} \tag{7.34}$$

For upper tail dependence we use (7.14) to obtain

$$\lambda_u = \lim_{q \to 1^-} \frac{\hat{C}(1-q, 1-q)}{1-q} = \lim_{q \to 0^+} \frac{\hat{C}(q, q)}{q}, \tag{7.35}$$

where $\hat{C}$ is the survival copula of C (see (7.16)). For radially symmetric copulas we must have $\lambda_l = \lambda_u$, since $C = \hat{C}$ for such copulas.

Calculation of these coefficients is straightforward if the copula in question has a simple closed form, as is the case for the Gumbel copula in (7.12) and the Clayton copula in (7.13). In Section 7.3.1 we will use a slightly more involved method to calculate tail-dependence coefficients for copulas of normal variance mixture distributions, such as the Gaussian and t copulas.

Example 7.37 (Gumbel and Clayton copulas). Writing $\hat{C}_\theta^{\mathrm{Gu}}$ for the Gumbel survival copula we first use (7.16) to infer that

$$\lambda_{\mathrm{u}} = \lim_{q \to 1^-} \frac{\hat{C}_\theta^{\mathrm{Gu}}(1-q, 1-q)}{1-q} = 2 - \lim_{q \to 1^-} \frac{C_\theta^{\mathrm{Gu}}(q,q) - 1}{q-1}.$$

We now use L'Hôpital's rule and the fact that $C_\theta^{\mathrm{Gu}}(u,u) = u^{2^{1/\theta}}$ to infer that

$$\lambda_{\mathrm{u}} = 2 - \lim_{q \to 1^-} \frac{\mathrm{d}C_\theta^{\mathrm{Gu}}(q,q)}{\mathrm{d}q} = 2 - 2^{1/\theta}.$$

Provided that $\theta > 1$, the Gumbel copula has upper tail dependence. The strength of this tail dependence tends to 1 as $\theta \to \infty$, which is to be expected since the Gumbel copula tends to the comonotonicity copula as $\theta \to \infty$. Using a similar technique the coefficient of lower tail dependence for the Clayton copula may be shown to be $\lambda_{\mathrm{l}} = 2^{-1/\theta}$ for $\theta > 0$.

The consequences of the lower tail dependence of the Clayton copula and the upper tail dependence of the Gumbel copula can be seen in Figures 7.3 and 7.4, where there is obviously an increased tendency for these copulas to generate joint extreme values in the respective corners. In Section 7.3.1 we will see that the Gauss copula is asymptotically independent in both tails, while the t copula has both upper and lower tail dependence of the same magnitude (due to its radial symmetry).

Notes and Comments

The concept of comonotonicity or perfect positive dependence is discussed by many authors, including Schmeidler (1986) and Yaari (1987). See also Wang and Dhaene (1998), whose proof we use in Proposition 7.18, and the entry in the *Encyclopedia of Actuarial Science* by Vyncke (2004).

The discussion of correlation fallacies is based on Embrechts, McNeil and Straumann (2002), which contains a number of other examples illustrating these pitfalls. Throughout this book we make numerous references to this paper, which also played a major role in popularizing the copula concept mainly, but not solely, in finance, insurance and economics (see, for example, Genest, Gendron and Bourdeau-Brien 2009). The ETH-RiskLab preprint of this paper was available as early as 1998, with a published abridged version appearing as Embrechts, McNeil and Straumann (1999).

For Höffding's formula and its use in proving the bounds on attainable correlations see Höffding (1940), Fréchet (1957) and Shea (1983). Useful references for rank correlations are Kruskal (1958) and Joag-Dev (1984). The relationship between rank correlation and copulas is discussed in Schweizer and Wolff (1981) and Nelsen (2006). The definition of tail dependence that we use stems from Joe (1993, 1997). There are a number of alternative definitions of tail-dependence measures, as discussed, for example, in Coles, Heffernan and Tawn (1999).

Important books that treat dependence concepts and emphasize links to copulas include Joe (1997), Denuit et al. (2005) and Rüschendorf (2013).

7.3 Normal Mixture Copulas

A unique copula is contained in every multivariate distribution with continuous marginal distributions, and a useful class of parametric copulas are those contained in the multivariate normal mixture distributions of Section 6.2. We view these copulas as particularly important in market-risk applications; indeed, in most cases, these copulas are used implicitly, without the user necessarily recognizing the fact. Whenever normal mixture distributions are fitted to multivariate return data or used as innovation distributions in multivariate time-series models, normal mixture copulas are used. They are also found in a number of credit risk models, as we discuss in Section 12.2.

In this section we first focus on normal variance mixture copulas; in Section 7.3.1 we examine their tail-dependence properties; and in Section 7.3.2 we calculate rank correlation coefficients, which are useful for calibrating these copulas to data. Then, in Sections 7.3.3 and 7.3.4, we look at more exotic examples of copulas arising from multivariate normal mixture constructions.

7.3.1 Tail Dependence

Coefficients of tail dependence. Consider a pair of uniform rvs (U_1, U_2) whose distribution $C(u_1, u_2)$ is a normal variance mixture copula. Due to the radial symmetry of C (see Section 7.1.5), it suffices to consider the formula for the lower tail-dependence coefficient in (7.34) to calculate the coefficient of tail dependence λ of C. By applying L'Hôpital's rule and using (7.17) we obtain

$$\lambda = \lim_{q \to 0^+} \frac{\mathrm{d}C(q,q)}{\mathrm{d}q} = \lim_{q \to 0^+} P(U_2 \leqslant q \mid U_1 = q) + \lim_{q \to 0^+} P(U_1 \leqslant q \mid U_2 = q).$$

Since C is *exchangeable*, we have from (7.21) that

$$\lambda = 2 \lim_{q \to 0^+} P(U_2 \leqslant q \mid U_1 = q). \tag{7.36}$$

We now show the interesting contrast between the Gaussian and t copulas that we alluded to in Example 7.13, namely that the t copula has tail dependence whereas the Gauss copula is asymptotically independent in the tail.

Example 7.38 (asymptotic independence of the Gauss copula). To evaluate the tail-dependence coefficient for the Gauss copula C_ρ^{Ga}, let $(X_1, X_2) := (\Phi^{-1}(U_1), \Phi^{-1}(U_2))$, so that (X_1, X_2) has a bivariate normal distribution with standard margins and correlation ρ. It follows from (7.36) that

$$\begin{aligned}\lambda &= 2 \lim_{q \to 0^+} P(\Phi^{-1}(U_2) \leqslant \Phi^{-1}(q) \mid \Phi^{-1}(U_1) = \Phi^{-1}(q)) \\ &= 2 \lim_{x \to -\infty} P(X_2 \leqslant x \mid X_1 = x).\end{aligned}$$

Using the fact that $X_2 \mid X_1 = x \sim N(\rho x, 1 - \rho^2)$, it can be calculated that

$$\lambda = 2 \lim_{x \to -\infty} \Phi(x\sqrt{1-\rho}/\sqrt{1+\rho}) = 0,$$

Table 7.1. Values of λ, the coefficient of upper and lower tail dependence, for the t copula $C^t_{\nu,\rho}$ for various values of ν, the degrees of freedom, and ρ, the correlation. The last row represents the Gauss copula.

	ρ				
ν	-0.5	0	0.5	0.9	1
2	0.06	0.18	0.39	0.72	1
4	0.01	0.08	0.25	0.63	1
10	0.00	0.01	0.08	0.46	1
∞	0	0	0	0	1

provided $\rho < 1$. Hence, the Gauss copula is asymptotically independent in both tails. Regardless of how high a correlation we choose, if we go far enough into the tail, extreme events appear to occur independently in each margin.

Example 7.39 (asymptotic dependence of the *t* copula). To evaluate the tail-dependence coefficient for the t copula $C^t_{\nu,\rho}$, let $(X_1, X_2) := (t_\nu^{-1}(U_1), t_\nu^{-1}(U_2))$, where t_ν denotes the df of a univariate t distribution with ν degrees of freedom. Thus $(X_1, X_2) \sim t_2(\nu, \mathbf{0}, P)$, where P is a correlation matrix with off-diagonal element ρ. By calculating the conditional density from the joint and marginal densities of a bivariate t distribution, it may be verified that, conditional on $X_1 = x$,

$$\left(\frac{\nu+1}{\nu+x^2}\right)^{1/2} \frac{X_2 - \rho x}{\sqrt{1-\rho^2}} \sim t_{\nu+1}. \tag{7.37}$$

Using an argument similar to Example 7.38 we find that

$$\lambda = 2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right). \tag{7.38}$$

Provided that $\rho > -1$, the copula of the bivariate t distribution is asymptotically dependent in both the upper and lower tails.

In Table 7.1 we tabulate the coefficient of tail dependence for various values of ν and ρ. For fixed ρ the strength of the tail dependence increases as ν decreases, and for fixed ν tail dependence increases as ρ increases. Even for zero or negative correlation values there is some tail dependence. This is not too surprising and can be grasped intuitively by recalling from Section 6.2.1 that the t distribution is a normal mixture distribution with a mixing variable W whose distribution is inverse gamma (which is a heavy-tailed distribution): if $|X_1|$ is large, there is a good chance that this is because W is large, increasing the probability of $|X_2|$ being large.

We could use the same method used in the previous examples to calculate tail-dependence coefficients for other copulas of normal variance mixtures. In doing so we would find that most examples, such as copulas of symmetric hyperbolic or NIG distributions, fell into the same category as the Gauss copula and were asymptotically independent in the tails. The essential determinant of whether the copula of a normal variance mixture has tail dependence or not is the tail of the distribution of the mixing

variable W in Definition 6.4. If W has a distribution with a power tail, then we get tail dependence, otherwise we get asymptotic independence. This is a consequence of a general result for elliptical distributions given in Section 16.1.3.

Joint quantile exceedance probabilities. Coefficients of tail dependence are of course asymptotic quantities, and in the remainder of this section we look at joint exceedances of *finite high quantiles* for the Gauss and t copulas in order to learn more about the practical consequences of the differences between the extremal behaviours of these two models.

As motivation we consider Figure 7.9, where 5000 simulated points from four different distributions are displayed. The distributions in (a) and (b) are meta-Gaussian distributions (see Section 7.1.3); they share the same copula C_ρ^{Ga}. The distributions in (c) and (d) are meta-t distributions; they share the same copula $C_{\nu,\rho}^t$. The values of ν and ρ in all parts are 4 and 0.5, respectively. The distributions in (a) and (c) share the same margins, namely standard normal margins. The distributions in (b) and (d) both have Student t margins with four degrees of freedom. The distributions in (a) and (d) are, of course, elliptical, being a standard bivariate normal and a bivariate t distribution with four degrees of freedom; they both have linear correlation $\rho = 0.5$. The other distributions are not elliptical and do not necessarily have linear correlation 50%, since altering the margins alters the linear correlation. All four distributions have identical Kendall's tau values (see Proposition 7.43). The meta-Gaussian distributions have the same Spearman's rho value, as do the meta-t distributions, although the two values are not identical (see Section 7.2.3).

The vertical and horizontal lines mark the true theoretical 0.005 and 0.995 quantiles for all distributions. Note that for the meta-t distributions the number of points that lie below both 0.005 quantiles or exceed both 0.995 quantiles is clearly greater than for the meta-Gaussian distributions, and this can be explained by the tail dependence of the t copula. The true theoretical ratio by which the number of these joint exceedances in the meta-t models should exceed the number in the meta-Gaussian models is 2.79, as may be read from Table 7.2, whose interpretation we now discuss.

In Table 7.2 we have calculated values of $C_\rho^{\mathrm{Ga}}(u,u)/C_{\nu,\rho}^t(u,u)$ for various ρ and ν and $u = 0.05, 0.01, 0.005, 0.001$. The rows marked Gauss contain values of $C_\rho^{\mathrm{Ga}}(u,u)$, which is the probability that two rvs with this copula are below their u-quantiles; we term this event a joint quantile exceedance (thinking of exceedance in the downwards direction). It is obviously identical to the probability that both rvs are larger than their $(1-u)$-quantiles. The remaining rows give the values of the ratio and thus express the amount by which the joint quantile exceedance probabilities must be inflated when we move from models with a Gauss copula to models with a t copula.

In Table 7.3 we extend Table 7.2 to higher dimensions. We now focus only on joint exceedances of the 1% (or 99%) quantile(s). We tabulate values of the ratio $C_P^{\mathrm{Ga}}(u,\ldots,u)/C_{\nu,P}^t(u,\ldots,u)$, where P is an equicorrelation matrix with all correlations equal to ρ. It is noticeable that not only do these values grow as the correlation parameter or number of degrees of freedom falls, but they also grow with the

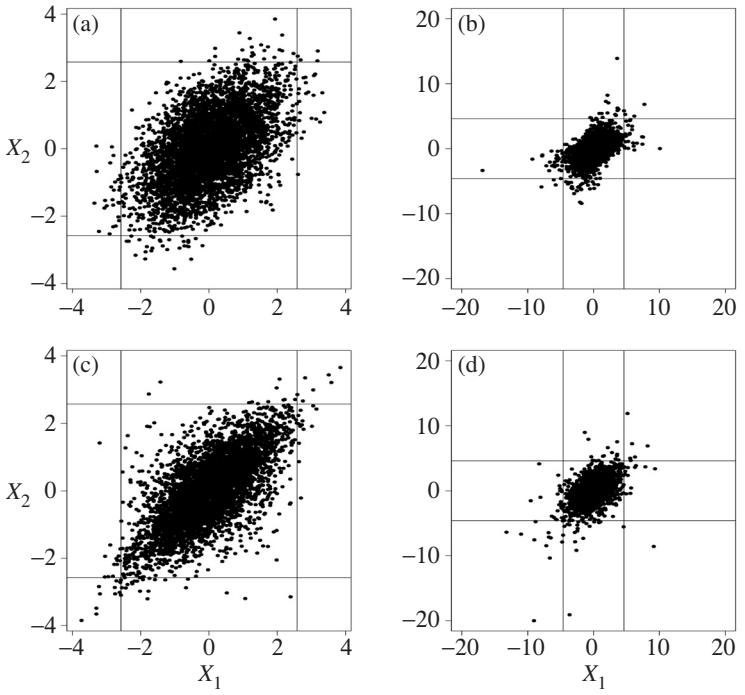

Figure 7.9. Five thousand simulated points from four distributions. (a) Standard bivariate normal with correlation parameter $\rho = 0.5$. (b) Meta-Gaussian distribution with copula C_ρ^{Ga} and Student t margins with four degrees of freedom. (c) Meta-t distribution with copula $C_{4,\rho}^t$ and standard normal margins. (d) Standard bivariate t distribution with four degrees of freedom and correlation parameter $\rho = 0.5$. Horizontal and vertical lines mark the 0.005 and 0.995 quantiles. See Section 7.3.1 for a commentary.

Table 7.2. Joint quantile exceedance probabilities for bivariate Gauss and t copulas with correlation parameter values of 0.5 and 0.7. For Gauss copulas the probability of joint quantile exceedance is given; for the t copulas the factors by which the Gaussian probability must be multiplied are given.

			Quantile			
ρ	Copula	ν	0.05	0.01	0.005	0.001
0.5	Gauss		1.21×10^{-2}	1.29×10^{-3}	4.96×10^{-4}	5.42×10^{-5}
0.5	t	8	1.20	1.65	1.94	3.01
0.5	t	4	1.39	2.22	2.79	4.86
0.5	t	3	1.50	2.55	3.26	5.83
0.7	Gauss		1.95×10^{-2}	2.67×10^{-3}	1.14×10^{-3}	1.60×10^{-4}
0.7	t	8	1.11	1.33	1.46	1.86
0.7	t	4	1.21	1.60	1.82	2.52
0.7	t	3	1.27	1.74	2.01	2.83

Table 7.3. Joint 1% quantile exceedance probabilities for multivariate Gaussian and t equicorrelation copulas with correlation parameter values of 0.5 and 0.7. For Gauss copulas the probability of joint quantile exceedance is given; for the t copulas the factors by which the Gaussian probability must be multiplied are given.

			Dimension d			
ρ	Copula	ν	2	3	4	5
0.5	Gauss		1.29×10^{-3}	3.66×10^{-4}	1.49×10^{-4}	7.48×10^{-5}
0.5	t	8	1.65	2.36	3.09	3.82
0.5	t	4	2.22	3.82	5.66	7.68
0.5	t	3	2.55	4.72	7.35	10.34
0.7	Gauss		2.67×10^{-3}	1.28×10^{-3}	7.77×10^{-4}	5.35×10^{-4}
0.7	t	8	1.33	1.58	1.78	1.95
0.7	t	4	1.60	2.10	2.53	2.91
0.7	t	3	1.74	2.39	2.97	3.45

dimension of the copula. The next example gives an interpretation of one of these numbers.

Example 7.40 (joint quantile exceedances: an interpretation). Consider daily returns on five stocks. Suppose we are unsure about the best multivariate elliptical model for these returns, but we believe that the correlation between any two returns on the same day is 50%. If returns follow a multivariate Gaussian distribution, then the probability that on any day all returns are below the 1% quantiles of their respective distributions is 7.48×10^{-5}. In the long run, such an event will happen once every 13 369 trading days on average: that is, roughly once every 51.4 years (assuming 260 trading days in a year). On the other hand, if returns follow a multivariate t distribution with four degrees of freedom, then such an event will happen 7.68 times more often: that is, roughly once every 6.7 years. In the life of a risk manager, 50-year events and 7-year events have a very different significance.

7.3.2 Rank Correlations

To calculate rank correlations for normal variance mixture copulas we use the following preliminary result for elliptical distributions.

Proposition 7.41. *Let $\boldsymbol{X} \sim E_2(\mathbf{0}, \Sigma, \psi)$ and $\rho = \wp(\Sigma)_{12}$, where $\wp$ denotes the correlation operator in (6.5). Assume $P(\boldsymbol{X} = \mathbf{0}) = 0$. Then*

$$P(X_1 > 0, X_2 > 0) = \tfrac{1}{4} + \frac{\arcsin \rho}{2\pi}.$$

Proof. First we make a standardization of the variables and observe that if $\boldsymbol{Y} \sim E_2(0, P, \psi)$ and $P = \wp(\Sigma)$, then $P(X_1 > 0, X_2 > 0) = P(Y_1 > 0, Y_2 > 0)$. Now introduce a pair of spherical variates $\boldsymbol{Z} \sim S_2(\psi)$; it follows that

$$\begin{aligned}(Y_1, Y_2) &\stackrel{\mathrm{d}}{=} (Z_1, \rho Z_1 + \sqrt{1-\rho^2} Z_2) \\ &\stackrel{\mathrm{d}}{=} R(\cos \Theta, \rho \cos \Theta + \sqrt{1-\rho^2} \sin \Theta),\end{aligned}$$

where R is a positive radial rv and Θ is an independent, uniformly distributed angle on $[-\pi, \pi)$ (see Section 6.3.1 and Theorem 6.21). Let $\phi = \arcsin \rho$ and observe that $\sin\phi = \rho$ and $\cos\phi = \sqrt{1-\rho^2}$. Since $P(R = 0) = P(\boldsymbol{X} = \mathbf{0}) = 0$ we conclude that

$$\begin{aligned} P(X_1 > 0, X_2 > 0) &= P(\cos\Theta > 0, \sin\phi\cos\Theta + \cos\phi\sin\Theta > 0) \\ &= P(\cos\Theta > 0, \sin(\Theta + \phi) > 0). \end{aligned}$$

The angle Θ must jointly satisfy $\Theta \in (-\frac{1}{2}\pi, \frac{1}{2}\pi)$ and $\Theta + \phi \in (0, \pi)$, and it is easily seen that for any value of ϕ this has probability $(\frac{1}{2}\pi + \phi)/(2\pi)$, which gives the result. □

Theorem 7.42 (rank correlations for the Gauss copula). *Let $\boldsymbol{X}$ have a bivariate meta-Gaussian distribution with copula C_ρ^{Ga} and continuous margins. Then the rank correlations are*

$$\rho_\tau(X_1, X_2) = \frac{2}{\pi}\arcsin\rho, \tag{7.39}$$

$$\rho_S(X_1, X_2) = \frac{6}{\pi}\arcsin\tfrac{1}{2}\rho. \tag{7.40}$$

Proof. Since rank correlation is a copula property, we can of course simply assume that $\boldsymbol{X} \sim N_2(\mathbf{0}, P)$, where P is a correlation matrix with off-diagonal element ρ; the calculations are then easy. For Kendall's tau, formula (7.29) implies

$$\rho_\tau(X_1, X_2) = 4P(Y_1 > 0, Y_2 > 0) - 1,$$

where $\boldsymbol{Y} = \tilde{\boldsymbol{X}} - \boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ is an independent copy of $\boldsymbol{X}$. Since, by the convolution property of the multivariate normal distribution in Section 6.1.3, $\boldsymbol{Y} \sim N_2(\mathbf{0}, 2P)$, we have that $\rho(Y_1, Y_2) = \rho$ and formula (7.39) follows from Proposition 7.41.

For Spearman's rho, let $\boldsymbol{Z} = (Z_1, Z_2)'$ be a vector consisting of two independent standard normal random variables and observe that formula (7.31) implies

$$\begin{aligned} \rho_S(X_1, X_2) &= 3(2P((X_1 - Z_1)(X_2 - Z_2) > 0) - 1) \\ &= 3(4P(X_1 - Z_1 > 0, X_2 - Z_2 > 0) - 1) \\ &= 3(4P(Y_1 > 0, Y_2 > 0) - 1), \end{aligned}$$

where $\boldsymbol{Y} = \boldsymbol{X} - \boldsymbol{Z}$. Since $\boldsymbol{Y} \sim N_2(\mathbf{0}, (P + I_2))$, the formula (7.40) follows from Proposition 7.41 and the fact that $\rho(Y_1, Y_2) = \rho/2$. □

These relationships between the rank correlations and ρ are illustrated in Figure 7.10. Note that the right-hand side of (7.40) may be approximated by the value ρ itself. This approximation turns out to be very accurate, as shown in the figure; the error bounds are $|6\arcsin(\rho/2)/\pi - \rho| \leqslant (\pi - 3)|\rho|/\pi \leqslant 0.0181$.

The relationship between Kendall's tau and the correlation parameter of the Gauss copula C_ρ^{Ga} expressed by (7.39) holds more generally for the copulas of all elliptical distributions that exclude point mass at their centre, including, for example, the t copula $C_{\nu,\rho}^t$. This is a consequence of the following result, which was already used to derive an alternative correlation estimator for bivariate distributions in Section 6.3.4.

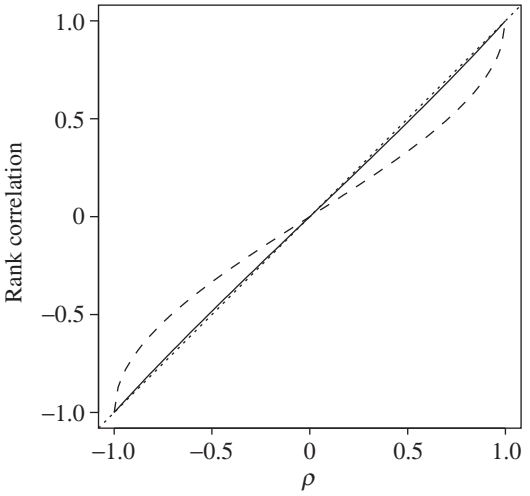

Figure 7.10. The solid line shows the relationship between Spearman's rho and the correlation parameter ρ of the Gauss copula C_ρ^{Ga} for meta-Gaussian rvs with continuous dfs; this is very close to the line $y = x$, which is just visible as a dotted line. The dashed line shows the relationship between Kendall's tau and ρ; this relationship holds for the copulas of other normal variance mixture distributions with correlation parameter ρ, such as the t copula $C_{\nu,\rho}^t$.

Proposition 7.43. *Let $\boldsymbol{X} \sim E_2(\mathbf{0}, P, \psi)$ for a correlation matrix P with off-diagonal element ρ, and assume that $P(\boldsymbol{X} = \mathbf{0}) = 0$. Then the relationship $\rho_\tau(X_1, X_2) = (2/\pi)\arcsin\rho$ holds.*

Proof. The result relies on the convolution property of elliptical distributions in (6.47). Setting $\boldsymbol{Y} = \tilde{\boldsymbol{X}} - \boldsymbol{X}$, where $\tilde{\boldsymbol{X}}$ is an independent copy of $\boldsymbol{X}$, we note that $\boldsymbol{Y} \sim E_2(\mathbf{0}, P, \tilde{\psi})$ for some characteristic generator $\tilde{\psi}$. We need to evaluate $\rho_\tau(X_1, X_2) = 4P(Y_1 > 0, Y_2 > 0) - 1$ as in the proof of Theorem 7.42, but Proposition 7.41 shows that $P(Y_1 > 0, Y_2 > 0)$ takes the same value whenever $\boldsymbol{Y}$ is elliptical. □

The relationship (7.40) between Spearman's rho and the correlation parameter of the Gauss copula does not hold for the copulas of all elliptical distributions. We can, however, derive a formula for the copulas of normal variance mixture distributions based on the following result.

Proposition 7.44. *Let $\boldsymbol{X} \sim M_2(\mathbf{0}, P, \hat{H})$ be distributed according to a normal variance mixture distribution for a correlation matrix P with off-diagonal element ρ, and assume that $P(\boldsymbol{X} = \mathbf{0}) = 0$. Then*

$$\rho_S(X_1, X_2) = \frac{6}{\pi} E\left(\arcsin\left(\rho \frac{W}{\sqrt{(W + \tilde{W})(W + \bar{W})}}\right)\right), \tag{7.41}$$

where W, $\tilde{W}$ and $\bar{W}$ are independent random variables with df H such that the Laplace–Stieltjes transform of H is $\hat{H}$.

Proof. Assume that $\boldsymbol{X} = \sqrt{W}\boldsymbol{Z}$, where $\boldsymbol{Z} \sim N_2(\mathbf{0}, P)$. Let $\tilde{Z}$ and $\bar{Z}$ be standard normal variables and assume that $\boldsymbol{Z}$, $\tilde{Z}$, $\bar{Z}$, W, $\tilde{W}$ and $\bar{W}$ are all independent. Write $\tilde{X} := \sqrt{\tilde{W}}\tilde{Z}$, $\bar{X} := \sqrt{\bar{W}}\bar{Z}$ and

$$Y_1 := X_1 - \tilde{X} = \sqrt{W}Z_1 - \sqrt{\tilde{W}}\tilde{Z},$$
$$Y_2 := X_2 - \bar{X} = \sqrt{W}Z_2 - \sqrt{\bar{W}}\bar{Z}.$$

The result is proved by applying a similar approach to Theorem 7.42 and conditioning on the variables W, $\tilde{W}$ and $\bar{W}$. We note that the conditional distribution of $\boldsymbol{Y} = (Y_1, Y_2)'$ satisfies

$$\boldsymbol{Y} \mid W, \tilde{W}, \bar{W} \sim N_2\left(\mathbf{0}, \begin{pmatrix} W + \tilde{W} & W\rho \\ W\rho & W + \bar{W} \end{pmatrix}\right).$$

Using formula (7.31) we calculate

$$\begin{aligned}\rho_S(X_1, X_2) &= 6P((X_1 - \tilde{X}_1)(X_2 - \bar{X}_2) > 0) - 3 \\ &= 3(2E(P(Y_1Y_2 > 0 \mid W, \tilde{W}, \bar{W})) - 1) \\ &= 3E(4P(Y_1 > 0, Y_2 > 0 \mid W, \tilde{W}, \bar{W}) - 1),\end{aligned}$$

and (7.41) is obtained by applying Proposition 7.41 and using the fact that, given W, $\tilde{W}$ and $\bar{W}$,

$$\rho(Y_1, Y_2) = \rho\frac{W}{\sqrt{(W + \tilde{W})(W + \bar{W})}}.$$

□

The formula (7.41) reduces to the formula (7.40) in the case where $W = \tilde{W} = \bar{W} = k$ for some positive constant k. In general, Spearman's rho for the copulas of normal variance mixtures can be calculated accurately by approximating the integral in (7.41) using Monte Carlo. For example, to calculate Spearman's rho for the t copula $C^t_{\nu,\rho}$ we would generate a set of inverse gamma variates $\{W_j, \tilde{W}_j, \bar{W}_j,\ j = 1, \ldots, m\}$ for m large, such that each variable in the set had an independent $\text{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$ distribution. We would then use

$$\rho_S(X_1, X_2) \approx \frac{6}{m\pi}\sum_{j=1}^{m}\arcsin\left(\rho\frac{W_j}{\sqrt{(W_j + \tilde{W}_j)(W_j + \bar{W}_j)}}\right). \tag{7.42}$$

7.3.3 Skewed Normal Mixture Copulas

A skewed normal mixture copula is the copula of any normal mixture distribution that is not elliptically symmetric. An example is provided by the *skewed t copula*, which is the copula of the distribution whose density is given in (6.31).

A random vector $\boldsymbol{X}$ with a skewed t distribution and ν degrees of freedom is denoted by $\boldsymbol{X} \sim \text{GH}_d(-\frac{1}{2}\nu, \nu, 0, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$ in the notation of Section 6.2.3. Its marginal distributions satisfy $X_i \sim \text{GH}_1(-\frac{1}{2}\nu, \nu, 0, \mu_i, \Sigma_{ii}, \gamma_i)$ (from Proposition 6.13) and its copula depends on ν, $P = \wp(\Sigma)$ and $\boldsymbol{\gamma}$ and will be denoted

by $C^t_{\nu,P,\boldsymbol{\gamma}}$ or, in the bivariate case, $C^t_{\nu,\rho,\gamma_1,\gamma_2}$. Random sampling from the skewed t copula follows the general approach of Algorithm 7.10.

Algorithm 7.45 (simulation of the skewed t copula).

(1) Generate $\boldsymbol{X} \sim \mathrm{GH}_d(-\frac{1}{2}\nu, \nu, 0, \mathbf{0}, P, \boldsymbol{\gamma})$ using Algorithm 6.10.

(2) Return $\boldsymbol{U} = (F_1(X_1), \ldots, F_d(X_d))'$, where F_i is the distribution function of a $\mathrm{GH}_1(-\frac{1}{2}\nu, \nu, 0, 0, 1, \gamma_i)$ distribution. The random vector $\boldsymbol{U}$ has df $C^t_{\nu,P,\boldsymbol{\gamma}}$.

Note that the evaluation of F_i requires the numerical integration of the density of a skewed univariate t distribution.

To appreciate the flexibility of the skewed t copula it suffices to consider the bivariate case for different values of the skewness parameters γ_1 and γ_2. In Figure 7.11 we have plotted simulated points from nine different examples of this copula. Part (e) corresponds to the case when $\gamma_1 = \gamma_2 = 0$ and is thus the ordinary t copula. All other pictures show copulas that are non-radially symmetric (see Section 7.1.5), as is obvious by rotating each picture 180° about the point $(\frac{1}{2}, \frac{1}{2})$; (c), (e) and (g) show exchangeable copulas satisfying (7.20), while the remaining six are non-exchangeable.

Obviously, the main advantage of the skewed t copula over the ordinary t copula is that its asymmetry allows us to have different levels of tail dependence in "opposite corners" of the distribution. In the context of market risk it is often claimed that joint negative returns on stocks show more tail dependence than joint positive returns.

7.3.4 Grouped Normal Mixture Copulas

Technically speaking, a grouped normal mixture copula is not itself the copula of a normal mixture distribution, but rather a way of attaching together a set of normal mixture copulas. We will illustrate the idea by considering the *grouped t copula*. Here, the basic idea is to construct a copula for a random vector $\boldsymbol{X}$ such that certain subvectors of $\boldsymbol{X}$ have t copulas but quite different levels of tail dependence.

We create a distribution using a generalization of the variance-mixing construction $\boldsymbol{X} = \sqrt{W}\boldsymbol{Z}$ in (6.18). Rather than multiplying all components of a correlated Gaussian vector $\boldsymbol{Z}$ with the root of a single inverse-gamma-distributed variate W, as in Example 6.7, we instead multiply different subgroups with different variates W_j, where $W_j \sim \mathrm{Ig}(\frac{1}{2}\nu_j, \frac{1}{2}\nu_j)$ and the W_j are themselves comonotonic (see Section 7.2.1). We therefore create subgroups whose dependence properties are described by t copulas with different ν_j parameters. The groups may even consist of a single member for each ν_j parameter, an idea that has been developed by Luo and Shevchenko (2010) under the name of the generalized t copula.

Like the t copula, the skewed t copula and anything based on a mixture of multivariate normals, a grouped t copula is easy to simulate and thus to use in Monte Carlo risk studies—this has been a major motivation for its development. We formally define the grouped t copula by explaining in more detail how to generate a random vector $\boldsymbol{U}$ with that distribution.

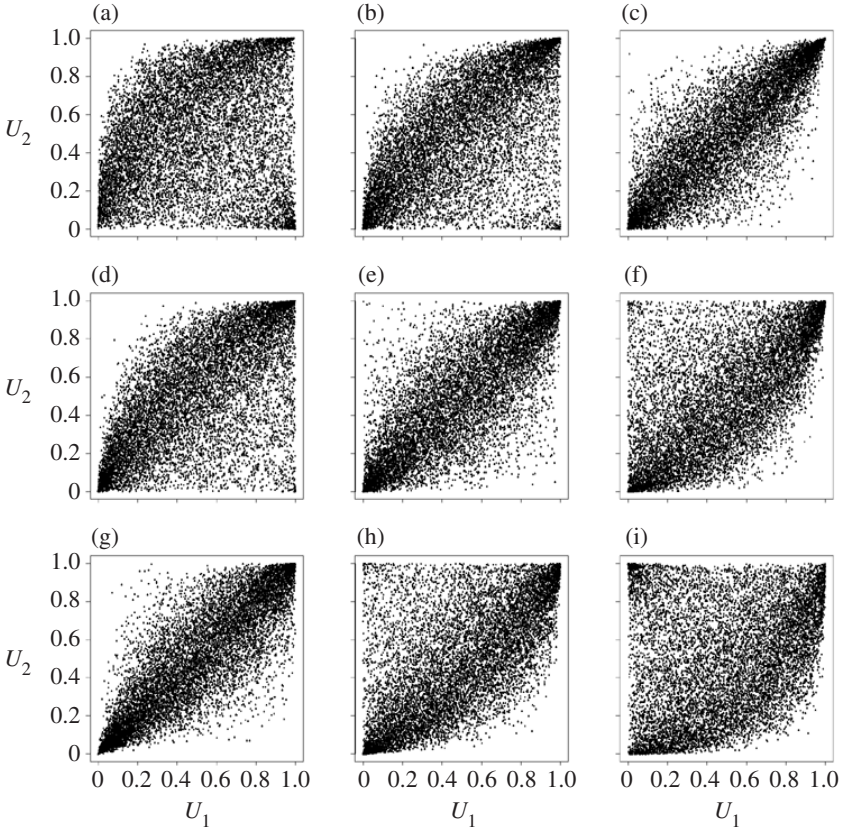

Figure 7.11. Ten thousand simulated points from bivariate skewed t copula $C^t_{\nu,\rho,\gamma_1,\gamma_2}$ for $\nu = 5$, $\rho = 0.8$ and various values of the parameters $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$: (a) $\boldsymbol{\gamma} = (0.8, -0.8)'$; (b) $\boldsymbol{\gamma} = (0.8, 0)'$; (c) $\boldsymbol{\gamma} = (0.8, 0.8)'$; (d) $\boldsymbol{\gamma} = (0, -0.8)'$; (e) $\boldsymbol{\gamma} = (0, 0)'$; (f) $\boldsymbol{\gamma} = (0, 0.8)'$; (g) $\boldsymbol{\gamma} = (-0.8, -0.8)'$; (h) $\boldsymbol{\gamma} = (-0.8, 0)'$; and (i) $\boldsymbol{\gamma} = (-0.8, 0.8)'$.

Algorithm 7.46 (simulation of the grouped t copula).

(1) Generate independently $\boldsymbol{Z} \sim N_d(\mathbf{0}, P)$ and $U \sim U(0, 1)$.

(2) Partition $\{1, \ldots, d\}$ into m subsets of sizes $s_1, \ldots, s_m$, and for $k = 1, \ldots, m$ let ν_k be the degrees-of-freedom parameter associated with group k.

(3) Set $W_k = G^{-1}_{\nu_k}(U)$, where G_ν is the df of the univariate $\mathrm{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$ distribution, so that $W_1, \ldots, W_m$ are comonotonic and inverse-gamma-distributed variates.

(4) Construct vectors $\boldsymbol{X}$ and $\boldsymbol{U}$ by

$$\boldsymbol{X} = (\sqrt{W_1}Z_1, \ldots, \sqrt{W_1}Z_{s_1}, \sqrt{W_2}Z_{s_1+1}, \ldots, \sqrt{W_2}Z_{s_1+s_2}, \ldots, \sqrt{W_m}Z_d)',$$
$$\boldsymbol{U} = (t_{\nu_1}(X_1), \ldots, t_{\nu_1}(X_{s_1}), t_{\nu_2}(X_{s_1+1}), \ldots, t_{\nu_2}(X_{s_1+s_2}), \ldots, t_{\nu_m}(X_d))'.$$

The former has a grouped t distribution and the latter is distributed according to a grouped t copula.

If we have an a priori idea of the desired group structure, we can calibrate the grouped t copula to data using a method based on Kendall's tau rank correlations. The use of this method for the ordinary t copula is described later in Section 7.5.1 and Example 7.56.

Notes and Comments

The coefficient of tail dependence for the t copula was first derived in Embrechts, McNeil and Straumann (2002). A more general result for the copulas of elliptical distributions is given in Hult and Lindskog (2002) and will be discussed in Section 16.1.3. The formula for Kendall's tau for elliptical distributions can be found in Lindskog, McNeil and Schmock (2003) and Fang and Fang (2002).

Proposition 7.44 is due to Andrew D. Smith (personal correspondence), who has also derived the attractive alterative formula

$$\rho_S(X_1, X_2) = (6/\pi)E(\arcsin(\rho \sin(\Theta/2))),$$

where Θ is the (random) angle in a triangle with side lengths $(W^{-1} + \tilde{W}^{-1})$, $(W^{-1} + \bar{W}^{-1})$ and $(\tilde{W}^{-1} + \bar{W}^{-1})$.

The skewed t copula was introduced in Demarta and McNeil (2005), which also describes the grouped t copula. The grouped t copula and a method for its calibration were first proposed in Daul et al. (2003). The special case of the grouped t copula with one member in each group has been investigated by Luo and Shevchenko (2010, 2012), who refer to this as a generalized t copula.

7.4 Archimedean Copulas

The Gumbel copula (7.12) and the Clayton copula (7.13) belong to the family of so-called Archimedean copulas, which has been very extensively studied. This family has proved useful for modelling portfolio credit risk, as will be seen in Example 11.4. In this section we look at the simple structure of these copulas and establish some of the properties that we will need.

7.4.1 Bivariate Archimedean Copulas

As well as the Gumbel and Clayton copulas, two further examples we consider are the *Frank copula*,

$$C_\theta^{\mathrm{Fr}}(u_1, u_2) = -\frac{1}{\theta} \ln\left(1 + \frac{(\mathrm{e}^{-\theta u_1} - 1)(\mathrm{e}^{-\theta u_2} - 1)}{\mathrm{e}^{-\theta} - 1}\right), \quad \theta \in \mathbb{R},$$

and a two-parameter copula that we refer to as a *generalized Clayton copula*,

$$C_{\theta,\delta}^{\mathrm{GC}}(u_1, u_2) = (((u_1^{-\theta} - 1)^\delta + (u_2^{-\theta} - 1)^\delta)^{1/\delta} + 1)^{-1/\theta}, \quad \theta \geqslant 0,\ \delta \geqslant 1.$$

It may be verified that, provided the parameter θ lies in the ranges we have specified in the copula definitions, all four examples that we have met have the form

$$C(u_1, u_2) = \psi(\psi^{-1}(u_1) + \psi^{-1}(u_2)), \tag{7.43}$$

Table 7.4. Table summarizing the generators, permissible parameter values and limiting special cases for four selected Archimedean copulas. The case $\theta = 0$ should be taken to mean the limit $\lim_{\theta\to 0}\psi_\theta(t)$. For the Clayton and Frank copulas this limit is e^{-t}, which is the generator of the independence copula.

Copula	Generator $\psi(t)$	Parameter range	$\psi \in \Psi_\infty$	Lower	Upper
C^{Gu}_θ	$\exp(-t^{1/\theta})$	$\theta \geqslant 1$	Yes	Π	M
C^{Cl}_θ	$\max((1+\theta t)^{-1/\theta}, 0)$	$\theta \geqslant -1$	$\theta \geqslant 0$	W	M
C^{Fr}_θ	$-\frac{1}{\theta}\ln(1+(\mathrm{e}^{-\theta}-1)\mathrm{e}^{-t})$	$\theta \in \mathbb{R}$	$\theta \geqslant 0$	W	M
$C^{\mathrm{GC}}_{\theta,\delta}$	$(1+\theta t^{1/\delta})^{-1/\theta}$	$\theta \geqslant 0, \delta \geqslant 1$	Yes	N/A	N/A

where $\psi : [0,\infty) \to [0,1]$ is a decreasing and continuous function that satisfies the conditions $\psi(0) = 1$ and $\lim_{t\to\infty}\psi(t) = 0$. The function ψ is known as the *generator* of the copula. For example, for the Gumbel copula $\psi(t) = \exp(-t^{1/\theta})$ for $\theta \geqslant 1$, and for the other copulas the generators are given in Table 7.4.

When we introduced the Clayton copula in (7.13) we insisted that its parameter should be non-negative. In the table we also define a Clayton copula for $-1 \leqslant \theta < 0$. To accommodate this case, the generator is written $\psi(t) = \max((1+\theta t)^{-1/\theta}, 0)$. We observe that $\psi(t)$ is strictly decreasing on $[0, -1/\theta)$ but $\psi(t) = 0$ on $[-1/\theta, \infty)$. To define the generator inverse uniquely at zero we set $\psi^{-1}(0) = \inf\{t : \psi(t) = 0\} = -1/\theta$.

In Table 7.4 we also give the lower and upper limits of the families as the parameter θ goes to the boundaries of the parameter space. Both the Frank and Clayton copulas are known as *comprehensive* copulas, since they interpolate between a lower limit of countermonotonicity and an upper limit of comonotonicity. For a more extensive table of Archimedean copulas see Nelsen (2006).

The following important theorem clarifies the conditions under which a function ψ is the generator of a bivariate Archimedean copula and allows us to define an Archimedean copula generator formally.

Theorem 7.47 (bivariate Archimedean copula). *Let $\psi : [0,\infty) \to [0,1]$ be a decreasing, continuous function that satisfies the conditions $\psi(0) = 1$ and $\lim_{t\to\infty}\psi(t) = 0$. Then*

$$C(u_1, u_2) = \psi(\psi^{-1}(u_1) + \psi^{-1}(u_2)) \tag{7.44}$$

is a copula if and only if ψ is convex.

Proof. See Nelsen (2006, Theorem 4.1.4). □

Definition 7.48 (Archimedean copula generator). A decreasing, continuous, convex function $\psi : [0,\infty) \to [0,1]$ satisfying $\psi(0) = 1$ and $\lim_{t\to\infty}\psi(t) = 0$ is known as an Archimedean copula generator.

Table 7.5. Kendall's rank correlations and coefficients of tail dependence for the copulas of Table 7.4. $D_1(\theta)$ is the Debye function $D_1(\theta) = \theta^{-1} \int_0^\theta t/(\mathrm{e}^t - 1)\,\mathrm{d}t$.

Copula	ρ_τ	λ_u	λ_l
C_θ^{Gu}	$1 - 1/\theta$	$2 - 2^{1/\theta}$	0
C_θ^{Cl}	$\theta/(\theta+2)$	0	$\begin{cases} 2^{-1/\theta}, & \theta > 0, \\ 0, & \theta \leqslant 0, \end{cases}$
C_θ^{Fr}	$1 - 4\theta^{-1}(1 - D_1(\theta))$	0	0
$C_{\theta,\delta}^{\mathrm{GC}}$	$\dfrac{(2+\theta)\delta - 2}{(2+\theta)\delta}$	$2 - 2^{1/\delta}$	$2^{-1/(\theta\delta)}$

Note that this definition automatically implies that ψ is strictly decreasing at all values t for which $\psi(t) > 0$, but there may be a flat piece if ψ attains the value zero. This is the only point where there is ambiguity about the inverse ψ^{-1}, and we set $\psi^{-1}(0) = \inf\{t\colon \psi(t) = 0\}$.

Kendall's rank correlations can be calculated for Archimedean copulas directly from the generator inverse using Proposition 7.49 below. The formula obtained can be used to calibrate Archimedean copulas to empirical data using the sample version of Kendall's tau, as we discuss in Section 7.5.

Proposition 7.49. *Let X_1 and X_2 be continuous rvs with unique Archimedean copula C generated by ψ. Then*

$$\rho_\tau(X_1, X_2) = 1 + 4 \int_0^1 \frac{\psi^{-1}(t)}{\mathrm{d}\psi^{-1}(t)/\mathrm{d}t}\,\mathrm{d}t. \tag{7.45}$$

Proof. See Nelsen (2006, Corollary 5.1.4). □

For the closed-form copulas of the Archimedean class, coefficients of tail dependence are easily calculated using methods of the kind used in Example 7.37. Values for Kendall's tau and the coefficients of tail dependence for the copulas of Table 7.4 are given in Table 7.5. It is interesting to note that the generalized Clayton copula $C_{\theta,\delta}^{\mathrm{GC}}$ combines, in a sense, both Gumbel's family and Clayton's family for positive parameter values, and thus succeeds in having tail dependence in both tails.

7.4.2 Multivariate Archimedean Copulas

It seems natural to attempt to construct a higher-dimensional Archimedean copula according to

$$C(u_1, \ldots, u_d) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)), \tag{7.46}$$

where ψ is an Archimedean generator function as in Definition 7.48. However, this construction may fail to define a proper distribution function for arbitrary dimension d. An example where this occurs is obtained if we take the generator $\psi(t) = 1 - t$, which is the Clayton generator for $\theta = -1$. In this case we obtain the Fréchet lower bound for copulas, which is not itself a copula for $d > 2$.

In order to guarantee that we will obtain a proper copula in any dimension we have to impose the property of *complete monotonicity* on ψ. A decreasing function $f(t)$ is completely monotonic on an interval $[a, b]$ if it satisfies

$$(-1)^k \frac{\mathrm{d}^k}{\mathrm{d}t^k} f(t) \geqslant 0, \quad k \in \mathbb{N},\ t \in (a, b). \tag{7.47}$$

Theorem 7.50. *If $\psi : [0, \infty) \to [0, 1]$ is an Archimedean copula generator, then the construction (7.46) gives a copula in any dimension d if and only if ψ is completely monotonic.*

Proof. See Kimberling (1974). □

If an Archimedean copula generator is completely monotonic, we write $\psi \in \Psi_\infty$. A column in Table 7.4 shows the cases where the generators are completely monotonic. For example, the Clayton generator is completely monotonic when $\theta \geqslant 0$ and the d-dimensional Clayton copula takes the form

$$C_\theta^{\mathrm{Cl}}(\boldsymbol{u}) = (u_1^{-\theta} + \cdots + u_d^{-\theta} - d + 1)^{-1/\theta}, \quad \theta \geqslant 0, \tag{7.48}$$

where the limiting case $\theta = 0$ should be interpreted as the d-dimensional independence copula.

The class of completely monotonic Archimedean copula generators is equivalent to the class of Laplace–Stieltjes transforms of dfs G on $[0, \infty)$ such that $G(0) = 0$. Let X be an rv with such a df G. We recall that the Laplace–Stieltjes transform of G (or X) is given by

$$\hat{G}(t) = \int_0^\infty \mathrm{e}^{-tx}\, \mathrm{d}G(x) = E(\mathrm{e}^{-tX}), \quad t \geqslant 0. \tag{7.49}$$

It is not difficult to verify that $\hat{G} : [0, \infty) \to [0, 1]$ is a continuous, strictly decreasing function with the property of complete monotonicity (7.47). Moreover, $\hat{G}(0) = 1$ and the exclusion of distributions with point mass at zero ensures $\lim_{t\to\infty} \hat{G}(t) = 0$. $\hat{G}$ therefore provides a candidate for an Archimedean generator that will generate a copula in any dimension.

This insight has a number of practical implications. On the one hand, we can create a rich variety of Archimedean copulas by considering Laplace–Stieltjes transforms of different distributions on $[0, \infty)$. On the other hand, we can derive a generic method of sampling from Archimedean copulas based on the following result.

Proposition 7.51. *Let G be a df on $[0, \infty)$ satisfying $G(0) = 0$ with Laplace–Stieltjes transform $\hat{G}$ as in (7.49). Let V be an rv with df G and let $Y_1, \ldots, Y_d$ be a sequence of independent, standard exponential rvs that are also independent of V. Then the following hold.*

(i) *The survival copula of the random vector $\boldsymbol{X} := (Y_1/V, \ldots, Y_d/V)$ is an Archimedean copula C with generator $\psi = \hat{G}$.*

(ii) *The random vector $\boldsymbol{U} := (\psi(X_1), \ldots, \psi(X_d))'$ is distributed according to C.*

(iii) *The components of* $\boldsymbol{U}$ *are conditionally independent given* V *with conditional df* $P(U_i \leqslant u \mid V = v) = \exp(-v\psi^{-1}(u))$.

Proof. For part (i) we calculate that, for $\boldsymbol{x} \in \mathbb{R}_+^d$,

$$\begin{aligned} P(X_1 > x_1, \ldots, X_d > x_d) &= \int_0^\infty \prod_{i=1}^d \mathrm{e}^{-x_i v} \, \mathrm{d}G(v) \\ &= \int_0^\infty \mathrm{e}^{-v(x_1+\cdots+x_d)} \, \mathrm{d}G(v) \\ &= \hat{G}(x_1 + \cdots + x_d). \end{aligned} \tag{7.50}$$

Since the marginal survival functions are given by $P(X_i > x) = \hat{G}(x)$ and $\hat{G}$ is continuous and strictly decreasing, the result follows from writing

$$P(X_1 > x_1, \ldots, X_d > x_d) = \hat{G}(\hat{G}^{-1}(P(X_1 > x_1)) + \cdots + \hat{G}^{-1}(P(X_d > x_d))).$$

Part (ii) follows easily from (7.50) since

$$\begin{aligned} P(\boldsymbol{U} \leqslant \boldsymbol{u}) &= P(U_1 < u_1, \ldots, U_d < u_d) \\ &= P(X_1 > \psi^{-1}(u_1), \ldots, X_d > \psi^{-1}(u_d)). \end{aligned}$$

The conditional independence is obvious in part (iii) and we calculate that

$$\begin{aligned} P(U_i \leqslant u \mid V = v) &= P(X_i > \psi^{-1}(u) \mid V = v) \\ &= P(Y_i > v\psi^{-1}(u)) \\ &= \mathrm{e}^{-v\psi^{-1}(u)}. \end{aligned}$$

□

Because of the importance of such copulas, particularly in the field of credit risk, we will call these copulas LT-Archimedean (LT stands for "Laplace transform") and make the following definition.

Definition 7.52 (LT-Archimedean copula). An LT-Archimedean copula is a copula of the form (7.46), where ψ is the Laplace–Stieltjes transform of a df G on $[0, \infty)$ satisfying $G(0) = 0$.

The sampling algorithm is based on parts (i) and (ii) of Proposition 7.51. We give explicit instructions for the Clayton, Gumbel and Frank copulas.

Algorithm 7.53 (simulation of LT-Archimedean copulas).

(1) Generate a variate V with df G such that $\hat{G}$, the Laplace–Stieltjes transform of G, is the generator ψ of the required copula.

(2) Generate independent uniform variates $Z_1, \ldots, Z_d$ and set $Y_i = -\ln(Z_i)$ for $i = 1, \ldots, d$ so that $Y_1, \ldots, Y_d$ are standard exponential.

(3) Return $\boldsymbol{U} = (\psi(Y_1/V), \ldots, \psi(Y_d/V))'$.

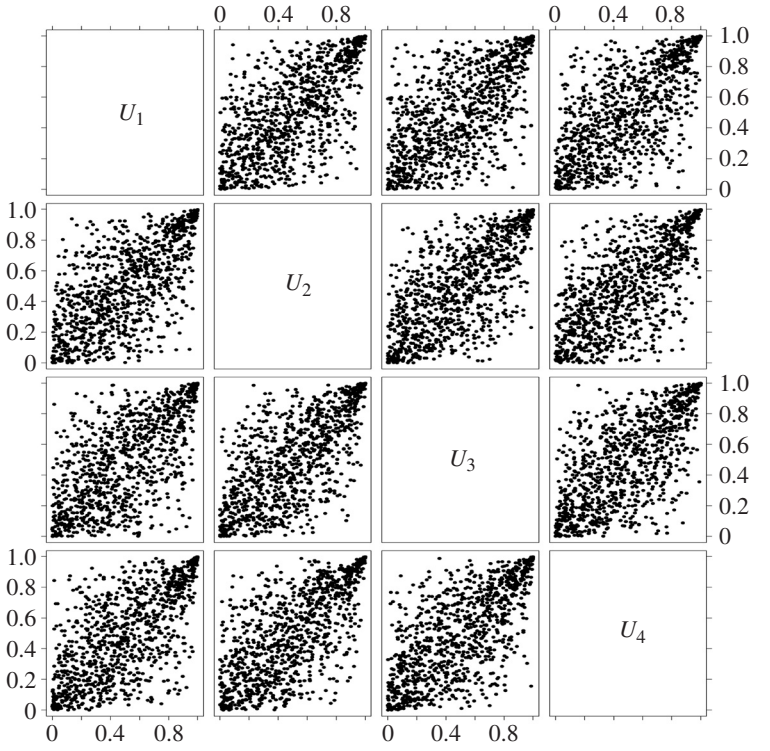

Figure 7.12. Pairwise scatterplots of 1000 simulated points from a four-dimensional exchangeable Gumbel copula with $\theta = 2$. Data are simulated using Algorithm 7.53.

(a) For the special case of the Clayton copula we generate a gamma variate $V \sim \text{Ga}(1/\theta, 1)$ with $\theta > 0$ (see Section A.2.4). The df of V has Laplace transform $\hat{G}(t) = (1+t)^{-1/\theta}$. This differs slightly from the Clayton generator in Table 7.4 but we note that $\hat{G}(\theta t)$ and $\hat{G}(t)$ generate the same Archimedean copula.

(b) For the special case of the Gumbel copula we generate a positive stable variate $V \sim \text{St}(1/\theta, 1, \gamma, 0)$, where $\gamma = (\cos(\pi/(2\theta)))^\theta$ and $\theta > 1$ (see Section A.2.9 for more details and a reference to a simulation algorithm). This df has Laplace transform $\hat{G}(t) = \exp(-t^{1/\theta})$ as desired.

(c) For the special case of the Frank copula we generate a discrete rv V with probability mass function $p(k) = P(V = k) = (1 - \mathrm{e}^{-\theta})^k/(k\theta)$ for $k = 1, 2, \ldots$ and $\theta > 0$. This can be achieved by standard simulation methods for discrete distributions (see Ripley 1987, p. 71).

See Figure 7.12 for an example of data simulated from a four-dimensional Gumbel copula using this algorithm. Note the upper tail dependence in each bivariate margin of this copula.

While Archimedean copulas with completely monotonic generators (Laplace–Stieltjes transforms) can be used in any dimension, if we are interested in

Archimedean copulas in a given dimension d, we can relax the requirement of complete monotonicity and substitute the weaker requirement of *d-monotonicity*. See Section 15.2.1 for more details.

A copula obtained from construction (7.46) is obviously an *exchangeable* copula conforming to (7.20). While exchangeable bivariate Archimedean copulas are widely used in modelling applications, their exchangeable multivariate extensions represent a very specialized form of dependence structure and have more limited applications. An exception to this is in the area of credit risk, although even here more general models with group structures are also needed. It is certainly natural to enquire whether there are extensions to the Archimedean class that are not rigidly exchangeable. We present some non-exchangeable Archimedean copulas in Section 15.2.2.

Notes and Comments

The name *Archimedean* relates to an algebraic property of the copulas that resembles the Archimedean axiom for real numbers (see Nelsen 2006, p. 122). The Clayton copula was introduced in Clayton (1978), although it has also been called the Cook and Johnson copula (see Genest and MacKay 1986) and the Pareto copula (see Hutchinson and Lai 1990). For the Frank copula see Frank (1979); this copula has radial symmetry and is the only such Archimedean copula. A useful reference, particularly for bivariate Archimedean copulas, is Nelsen (2006).

Theorem 7.47 is a result of Schweizer and Sklar (1983) (see also Alsina, Frank and Schweizer 2006). The formula for Kendall's tau in the Archimedean family is due to Genest and MacKay (1986). The link between completely monotonic functions and generators which give Archimedean copulas of the form (7.46) is found in Kimberling (1974). See also Feller (1971) for more on the concept of complete monotonicity. For more on the important connection between Archimedean generators and Laplace transforms, see Joe (1997).

Proposition 7.51 and Algorithm 7.53 are essentially due to Marshall and Olkin (1988). See Frees and Valdez (1997), Schönbucher (2005) and Frey and McNeil (2003) for further discussion of this technique. A text on simulation techniques for copula families is Mai and Scherer (2012).

Other copula families we have not considered include the Marshall–Olkin copulas (Marshall and Olkin 1967a,b) and the extremal copulas in Tiit (1996). There is also a large literature on pair copulas and vine copulas; fundamental references include Bedford and Cooke (2001), Kurowicka and Cooke (2006), Aas et al. (2009) and Czado (2010).

7.5 Fitting Copulas to Data

We assume that we have data vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ with identical distribution function F, describing financial losses or financial risk-factor returns; we write $\boldsymbol{X}_t = (X_{t,1}, \ldots, X_{t,d})'$ for an individual data vector and $\boldsymbol{X} = (X_1, \ldots, X_d)'$ for

a generic random vector with df F. We assume further that this df F has continuous margins $F_1, \ldots, F_d$ and thus, by Sklar's Theorem, a unique representation $F(x) = C(F_1(x_1), \ldots, F_d(x_d))$.

It is often very difficult, particularly in higher dimensions and in situations where we are dealing with skewed loss distributions or heterogeneous risk factors, to find a good multivariate model that describes both marginal behaviour and dependence structure effectively. For multivariate risk-factor return data of a similar kind, such as stock returns or exchange-rate returns, we have discussed useful overall models such as the GH family of Section 6.2.3, but even in these situations there can be value in separating the marginal-modelling and dependence-modelling issues and looking at each in more detail. The copula approach to multivariate models facilitates this approach and allows us to consider, for example, the issue of whether tail dependence appears to be present in our data.

This section is thus devoted to the problem of estimating the parameters $\boldsymbol{\theta}$ of a parametric copula $C_{\boldsymbol{\theta}}$. The main method we consider is maximum likelihood in Section 7.5.3. First we outline a simpler method-of-moments procedure using sample rank correlation estimates. This method has the advantage that marginal distributions do not need to be estimated, and consequently inference about the copula is in a sense "margin free".

7.5.1 Method-of-Moments Using Rank Correlation

Depending on which particular copula we want to fit, it may be easier to use empirical estimates of either Spearman's or Kendall's rank correlation to infer an estimate for the copula parameter. We begin by discussing the standard estimators of both of these rank correlations.

Proposition 7.34 suggests that we could estimate $\rho_S(X_i, X_j)$ by calculating the usual correlation coefficient for the *pseudo-observations*: $\{(F_{i,n}(X_{t,i}), F_{j,n}(X_{t,j})) : t = 1, \ldots, n\}$, where $F_{i,n}$ denotes the standard empirical df for the ith margin. In fact, we estimate $\rho_S(X_i, X_j)$ by calculating the correlation coefficient of the *ranks* of the data, a quantity known as the Spearman's rank correlation coefficient. This coincides with the correlation coefficient of the pseudo-observations when there are no tied observations (that is, observations with $X_{t,i} = X_{s,i}$ or $X_{t,j} = X_{s,j}$ for some $t \neq s$).

The rank of $X_{t,i}$, written $\text{rank}(X_{t,i})$, is simply the position of $X_{t,i}$ in the sample $X_{1,i}, \ldots, X_{n,i}$ when the observations are ordered from smallest to largest. If there are tied observations, we assign them a rank equal to the average rank that the observations would have if the ties were randomly broken; for example, the sample of four observations $\{2, 3, 2, 1\}$ would have ranks $\{2.5, 4, 2.5, 1\}$.

In the case of no ties the Spearman's rank correlation coefficient is given by the formula

$$r_{ij}^S = \frac{12}{n(n^2-1)} \sum_{t=1}^{n} (\text{rank}(X_{t,i}) - \tfrac{1}{2}(n+1))(\text{rank}(X_{t,j}) - \tfrac{1}{2}(n+1)). \quad (7.51)$$

We denote by $R^{\mathrm{S}} = (r_{ij}^{\mathrm{S}})$ the matrix of pairwise Spearman's rank correlation coefficients; since this is the sample correlation matrix of the vectors of ranks, it is clearly a positive-semidefinite matrix.

The standard estimator of Kendall's tau $\rho_\tau(X_i, X_j)$ is Kendall's rank correlation coefficient:

$$r_{ij}^{\tau} = \binom{n}{2}^{-1} \sum_{1 \leqslant t < s \leqslant n} \operatorname{sign}((X_{t,i} - X_{s,i})(X_{t,j} - X_{s,j})). \tag{7.52}$$

This is clearly the empirical analogue of the theoretical Kendall's tau in Definition 7.31. Note that the actual evaluation of this estimator for large n is time-consuming (in comparison with Spearman's rho) because every pair of observations must be considered. Again we can collect pairwise Kendall's rank correlation coefficients in a matrix $R^{\tau} = (r_{ij}^{\tau})$; by observing that this matrix may be written as

$$R^{\tau} = \binom{n}{2}^{-1} \sum_{1 \leqslant t < s \leqslant n} \operatorname{sign}(\boldsymbol{X}_t - \boldsymbol{X}_s) \operatorname{sign}(\boldsymbol{X}_t - \boldsymbol{X}_s)',$$

it is again apparent that this gives a positive-semidefinite matrix.

In a series of examples we show how these sample rank correlations can be used to calibrate (or partially calibrate) various copulas. Obviously we assume that there are a priori grounds for considering the chosen copula to be an appropriate model, such as symmetry or the lack of it and the presence or absence of tail dependence. The general method will always be similar: we look for a theoretical relationship between one of the rank correlations and the parameters of the copula and substitute empirical values of the rank correlation into this relationship to get estimates of some or all of the copula parameters.

Example 7.54 (bivariate Archimedean copulas with a single parameter). Suppose our assumed model is of the form $F(x_1, x_2) = C_\theta(F_1(x_1), F_2(x_2))$, where θ is a single parameter to be estimated. For many such copulas a simple functional relationship exists between either Kendall's tau and θ or Spearman's rho and θ. For specific examples consider the Gumbel, Clayton and Frank copulas of Section 7.4; in these cases we have simple relationships of the form $\rho_\tau(X_1, X_2) = f(\theta)$, as shown in Table 7.5. This suggests that we can calibrate these copulas by first calculating a sample value r^{τ} for Kendall's tau and then solving the equation $r^{\tau} = f(\hat{\theta})$ for $\hat{\theta}$, assuming that $\hat{\theta}$ is a valid value in the parameter space of the copula. For example, Gumbel's copula is calibrated by taking $\hat{\theta} = (1 - r^{\tau})^{-1}$, provided that $r^{\tau} \geqslant 0$. Clayton's copula interpolates between perfect negative and perfect positive dependence and can be calibrated to any sample Kendall's tau value in $(-1, 1)$. For the calibration of higher-dimensional Archimedean copulas using rank correlations, see Hofert, Mächler and McNeil (2012).

Example 7.55 (calibrating Gauss copulas using Spearman's rho). Suppose we assume a meta-Gaussian model for $\boldsymbol{X}$ with copula C_P^{Ga}, and we wish to estimate the correlation matrix P. It follows from Theorem 7.42 that

$$\rho_{\mathrm{S}}(X_i, X_j) = (6/\pi) \arcsin \tfrac{1}{2}\rho_{ij} \approx \rho_{ij},$$

where the final approximation is very accurate (see Figure 7.10). This suggests we estimate P by the matrix of pairwise Spearman's rank correlation coefficients R^{S}.

The method of Example 7.55 could be used to estimate P in a t copula model $C^t_{\nu,P}(F_1(x_1), \ldots, F_d(x_d))$, although the calibration would not be as accurate as in the Gaussian case. Empirical investigations of the relationship (7.41) based on the Monte Carlo approximation (7.42) suggest that the error $|\rho_{\mathrm{S}}(X_i, X_j) - \rho_{ij}|$, while still modest, is larger than in the Gaussian case and increases with decreasing degrees of freedom ν. Instead we propose a method based on Kendall's tau in the next example, which is based on Proposition 7.43 and could be applied to all elliptical copulas.

Example 7.56 (calibrating *t* copulas using Kendall's tau). Suppose we assume a meta-t model for $\boldsymbol{X}$ with copula $C^t_{\nu,P}$ and we wish to estimate the correlation matrix P. It follows from Proposition 7.43 that

$$\rho_\tau(X_i, X_j) = (2/\pi) \arcsin \rho_{ij},$$

so that a possible estimator of P is the matrix R^* with components given by $r^*_{ij} = \sin(\frac{1}{2}\pi r^\tau_{ij})$. However, there is no guarantee that this componentwise transformation of the matrix of Kendall's rank correlation coefficients will remain positive definite (although in our experience it very often does). In this case R^* can be transformed by the eigenvalue method given in Algorithm 7.57 to obtain a positive-definite matrix that is close to R^*. The remaining parameter ν of the copula could then be estimated by maximum likelihood, as discussed in Section 7.5.3.

Algorithm 7.57 (eigenvalue method). Let R^* be a so-called *pseudo*-correlation matrix, i.e. a symmetric matrix of pairwise correlation estimates with unit diagonal entries and off-diagonal entries in $[-1, 1]$ that is not positive semidefinite.

(1) Calculate the spectral decomposition $R^* = GLG'$ as in (6.59), where L is the matrix of eigenvalues and G is an orthogonal matrix whose columns are eigenvectors of R^*.

(2) Replace all negative eigenvalues in L by small values $\delta > 0$ to obtain $\tilde{L}$.

(3) Calculate $Q = G\tilde{L}G'$, which will be symmetric and positive definite but not a correlation matrix, since its diagonal elements will not necessarily equal one.

(4) Return the correlation matrix $R = \wp(Q)$, where $\wp$ denotes the correlation matrix operator defined in (6.5).

In Examples 7.55 and 7.56 we saw that it is relatively easy to calibrate the Gauss copula and the correlation parameter matrix P of the t copula to sample rank correlations. This technique is particularly useful when we have limited multivariate data and formal estimation of a full multivariate model is unrealistic. Consider the following hypothetical example.

Example 7.58 (fictitious risk integration situation). Suppose a company is divided into a number of business units that function semi-autonomously. The company management would like to calculate an enterprise-wide P&L distribution for a one-month period. They have historical data on monthly results for each of the business units for the last two years only, i.e. twenty-four observations. However, each business unit believes that through detailed knowledge of their own business going back over a longer period they can specify their own P&L fairly accurately. Rather than attempting to fit a multivariate distribution to twenty-four observations, the risk-management team decides to combine the individual marginal models provided by each of the business units using a matrix of rank correlations estimated from the twenty-four data points.

In this situation we can build multivariate models by combining the known marginal distributions using any copula that can be calibrated to the estimated rank correlations. The Gaussian and t copulas lend themselves to this purpose and can be used to build meta-Gaussian and meta-t models that are consistent with the available information.

Typically, these models could then be used in a Monte Carlo risk analysis; we have seen in Section 7.1.4 that meta-Gaussian and meta-t models are particularly easy to simulate. Because the approach is obviously prone to model risk (24 observations provide very meagre multivariate data) it should be seen as a form of sensitivity analysis performed using detailed marginal information and only vague dependence information; we might choose to compare a meta-Gaussian model with no tail dependence and a meta-t model with, say, three degrees of freedom and very strong tail dependence. In Section 8.4.4 we will have more to say on this problem of risk integration under dependence uncertainty.

7.5.2 Forming a Pseudo-sample from the Copula

We now turn to the estimation of parametric copulas by maximum likelihood (ML). In practical situations we are seldom interested in the copula alone, but also require estimates of the margins to form a full multivariate model; even when the copula is of central interest, as it is for us in this chapter, we are forced to estimate margins in order to estimate the copula, since copula data are almost never observed directly.

While we may attempt to estimate margins and copula in one single optimization, splitting the modelling into two steps can yield more insight and allow a more detailed analysis of the different model components. In this section we describe briefly some general approaches to the first step of estimating margins and constructing a *pseudo-sample* of observations from the copula. In the following section we describe how the copula parameters are estimated by ML from the pseudo-sample.

Let $\hat{F}_1, \dots, \hat{F}_d$ denote estimates of the marginal dfs (possible methods are discussed below). The pseudo-sample from the copula consists of the vectors $\hat{\boldsymbol{U}}_1, \dots, \hat{\boldsymbol{U}}_n$, where

$$\hat{\boldsymbol{U}}_t = (\hat{U}_{t,1}, \dots, \hat{U}_{t,d})' = (\hat{F}_1(X_{t,1}), \dots, \hat{F}_d(X_{t,d}))'. \tag{7.53}$$

Observe that, even if the original data vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are iid, the pseudo-sample data are generally dependent, because the marginal estimates $\hat{F}_i$ will in most cases be constructed from all of the original data vectors through the univariate samples $X_{1,i}, \ldots, X_{n,i}$. Possible methods for obtaining the marginal estimate $\hat{F}_i$ include the following.

(1) Parametric estimation. We choose an appropriate parametric model for the data in question and fit it by ML: for financial risk-factor return data we might consider the GH distribution, or one of its special cases such as Student t or normal inverse Gaussian (NIG); for insurance or operational loss data we might consider a standard actuarial loss distribution such as Pareto or lognormal.

(2) Non-parametric estimation with variant of empirical df. We could estimate F_j using

$$F^*_{i,n}(x) = \frac{1}{n+1} \sum_{t=1}^{n} I_{\{X_{t,i} \leqslant x\}}, \tag{7.54}$$

which differs from the usual empirical df by the use of the denominator $n + 1$ rather than n. This guarantees that the pseudo-copula data in (7.53) lie strictly in the interior of the unit cube; to implement ML we must be able to evaluate the copula density at each $\hat{\boldsymbol{U}}_i$, and in many cases this density is infinite on the boundary of the cube.

(3) Extreme value theory for the tails. Empirical distribution functions are known to be poor estimators of the underlying distribution in the tails. An alternative is to use a technique from extreme value theory, described in Section 5.2.6, whereby the tails are modelled semi-parametrically using a generalized Pareto distribution (GPD); the body of the distribution may be modelled empirically.

Example 7.59. We analyse five years of daily log-return data (1996–2000) for Intel, Microsoft and General Electric stocks. The marginal distributions are estimated empirically (method (2)) and the pseudo-sample from the copula is shown in Figure 7.13. Essentially, the points are plotted at the coordinates $(\text{rank}(X_{t,i})/(n+1), \text{rank}(X_{t,j})/(n+1))$, where $\text{rank}(X_{t,i})$ denotes the rank of $X_{t,i}$ in the sample $X_{1,i}, \ldots, X_{n,i}$.

7.5.3 Maximum Likelihood Estimation

Let $C_{\boldsymbol{\theta}}$ denote a parametric copula, where $\boldsymbol{\theta}$ is the vector of parameters to be estimated. The MLE is obtained by maximizing

$$\ln L(\boldsymbol{\theta}; \hat{\boldsymbol{U}}_1, \ldots, \hat{\boldsymbol{U}}_n) = \sum_{t=1}^{n} \ln c_{\boldsymbol{\theta}}(\hat{\boldsymbol{U}}_t) \tag{7.55}$$

with respect to $\boldsymbol{\theta}$, where $c_{\boldsymbol{\theta}}$ denotes the copula density as in (7.18) and $\hat{\boldsymbol{U}}_t$ denotes a pseudo-observation from the copula.

Obviously, the statistical quality of the estimates of the copula parameters depends very much on the quality of the estimates of the marginal distributions used in

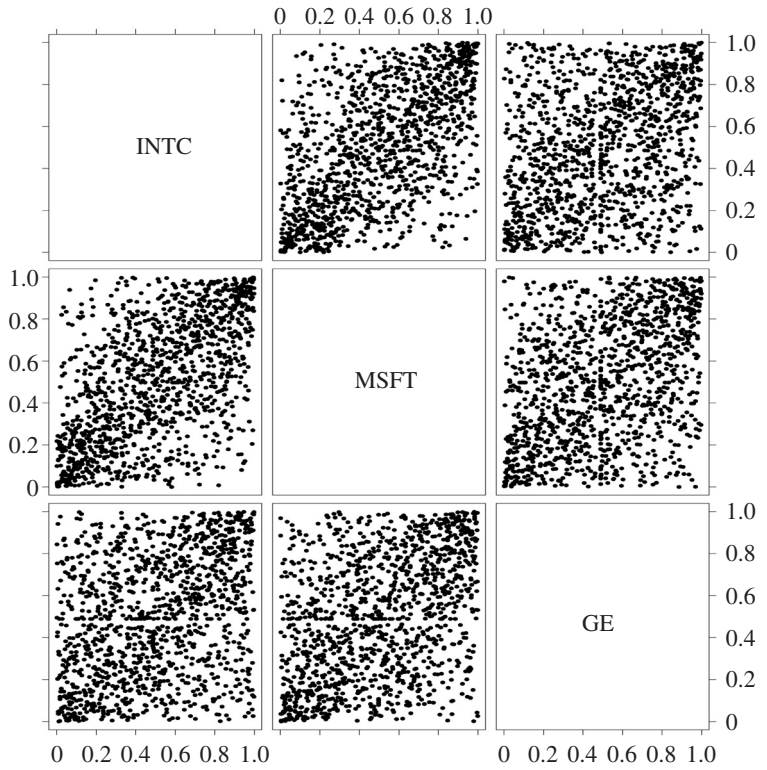

Figure 7.13. Pairwise scatterplots of a pseudo-sample from a copula for trivariate Intel, Microsoft and General Electric log-returns (see Example 7.59).

the formation of the pseudo-sample from the copula. The properties of estimates derived using the marginal estimation methods (1) and (2) in Section 7.5.2 have both been studied in more theoretical detail. When margins are estimated parametrically (method (1)), inference about the copula using (7.55) amounts to what has been termed the inference functions for margins (IFM) approach by Joe (1997). When margins are estimated non-parametrically (method (2)), the estimates of the copula parameters may be regarded as semi-parametric and the approach has been labelled pseudo-maximum likelihood by Genest and Rivest (1993) (see Notes and Comments for more references). One could envisage using the two-stage method to decide on the most appropriate copula family and then estimating all parameters (marginal and copula) in a final fully parametric round of estimation.

In practice, to implement the ML method we need to derive the copula density. This is straightforward, if tedious, for the exchangeable Archimedean copulas of Section 7.4, and these have been popular models in bivariate and trivariate applications to insurance loss data. For implicit copulas like the Gaussian and t copulas we use (7.19). The MLE is generally found by numerical maximization of the resulting log-likelihood (7.55).

Example 7.60 (fitting the Gauss copula). In the case of a Gauss copula we use (7.19) to see that the log-likelihood (7.55) becomes

$$\ln L(P; \hat{\boldsymbol{U}}_1, \ldots, \hat{\boldsymbol{U}}_n) = \sum_{t=1}^{n} \ln f_P(\Phi^{-1}(\hat{U}_{t,1}), \ldots, \Phi^{-1}(\hat{U}_{t,d})) - \sum_{t=1}^{n}\sum_{j=1}^{d} \ln \phi(\Phi^{-1}(\hat{U}_{t,j})),$$

where f_Σ will be used to denote the joint density of a random vector with $N_d(\mathbf{0}, \Sigma)$ distribution. It is clear that the second term is not relevant in the maximization with respect to P, and the MLE is given by

$$\hat{P} = \arg\max_{\Sigma \in \mathcal{P}} \sum_{t=1}^{n} \ln f_\Sigma(\boldsymbol{Y}_t), \tag{7.56}$$

where $Y_{t,j} = \Phi^{-1}(\hat{U}_{t,j})$ for $j = 1, \ldots, d$ and $\mathcal{P}$ denotes the set of all possible linear correlation matrices. To perform this maximization in practice, note that the set $\mathcal{P}$ can be constructed as

$$\mathcal{P} = \{P = \wp(Q) \colon Q = AA',\ A \text{ lower triangular with ones on the diagonal}\},$$

where $\wp$ is defined in (6.5). In other words, we can search over the set of unrestricted lower-triangular matrices with ones on the diagonal. This search is feasible in low dimensions but very slow in high dimensions, since the number of parameters is $O(d^2)$.

An approximate solution to the maximization may be obtained easily as follows. Suppose that instead of maximizing over $\mathcal{P}$ as in (7.56) we maximize over the set of all covariance matrices. This maximization problem has the analytical solution $\hat{\Sigma} = (1/n)\sum_{t=1}^{n} \boldsymbol{Y}_t \boldsymbol{Y}_t'$, which is the MLE of the covariance matrix Σ for iid normal data with $N_d(\mathbf{0}, \Sigma)$ distribution. In practice, $\hat{\Sigma}$ is likely to be *close* to being a correlation matrix. As an approximate solution to the original problem we could take the correlation matrix $\tilde{P} = \wp(\hat{\Sigma})$.

When a Gauss copula is fitted to the trivariate data in Example 7.59 by full ML, the estimated correlation matrix has entries 0.58 (INTC-MSFT), 0.34 (INTC-GE) and 0.40 (MSFT-GE); the value of the log-likelihood at the maximum is 376.65. Using the alternative method gives estimates that are identical to two significant figures and that yield a log-likelihood value of 376.62.

A further alternative would be to use the estimation procedure in Example 7.55, based on Spearman's rank correlations. Using the Spearman method we get, respectively, 0.57, 0.34 and 0.40 for the parameter estimates; the value of the log-likelihood at this value of P is 376.50, which is also not so far from the maximum.

Example 7.61 (fitting the *t* copula). In the case of the t copula, (7.19) implies that the log-likelihood (7.55) is

$$\ln L(\nu, P; \hat{\boldsymbol{U}}_1, \ldots, \hat{\boldsymbol{U}}_n) = \sum_{t=1}^{n} \ln g_{\nu,P}(t_\nu^{-1}(\hat{U}_{t,1}), \ldots, t_\nu^{-1}(\hat{U}_{t,d})) - \sum_{t=1}^{n}\sum_{j=1}^{d} \ln g_\nu(t_\nu^{-1}(\hat{U}_{t,j})),$$

where $g_{\nu,P}$ denotes the joint density of a random vector with $t_d(\nu, \mathbf{0}, P)$ distribution, P is a linear correlation matrix, g_ν is the density of a univariate $t_1(\nu, 0, 1)$ distribution, and t_ν^{-1} is the corresponding quantile function.

Again, in relatively low dimensions we could search over the set of correlation matrices P and degrees of freedom parameter ν for a global maximum. For higher-dimensional work it would be easier to estimate P using Kendall's tau estimates, as in Example 7.56, and to estimate the single parameter ν by maximum likelihood.

When a t copula is fitted to the trivariate data in Example 7.59 by full ML, the estimated matrix P has entries 0.59 (INTC-MSFT), 0.36 (INTC-GE) and 0.42 (MSFT-GE); the estimate of ν is 6.5 and the value of the log-likelihood at the maximum is 420.39. Using the simpler method based on Kendall's tau gives identical parameter estimates to two significant figures and a log-likelihood value of 420.32. Clearly, the t model fits much better than a Gauss copula model; the log-likelihood is increased by over 40. This would be massively significant in a likelihood ratio test (although, strictly speaking, such a test introduces a technical difficulty, since the Gauss copula represents a boundary case of the t copula model ($\nu = \infty$), which violates standard regularity conditions (see Notes and Comments)).

Following standard statistical practice we usually fit a number of copula models to data and compare the quality of the fitted models using tools like the Akaike information criterion (see Appendix A.3.6). We may also carry out goodness-of-fit tests to assess the plausibility that the data come from any given copula. Most of the goodness-of-fit tests that have been suggested for copulas are quite computationally intensive and are limited to applications in relatively low dimensions (see Notes and Comments for some references).

Notes and Comments

The copula estimation procedure based on empirical values of Kendall's tau is discussed in detail for bivariate Archimedean copulas by Genest and Rivest (1993); they explain why the procedure may be considered to be a method-of-moments technique and show how confidence intervals for the copula parameter (in the case of single-parameter copulas) may be derived.

The method of calibrating the Gauss copula with Spearman's rank correlation in Example 7.55 is essentially due to Iman and Conover (1982). The use of this calibration method to build meta-Gaussian models with prescribed margins and the Monte Carlo simulation of data from these models are implemented in the @RISK software (Palisade 1997), which is widely used in insurance. Our Example 7.56 is intended to show that this approach can be extended to meta-t models, which may well be more interesting due to their tail dependence.

The eigenvalue method for correcting the positive definiteness of correlation matrices given in Algorithm 7.57 is described by Rousseeuw and Molenberghs (1993). An empirical comparison of the eigenvalue method with different approaches to this problem, including so-called shrinkage methods, is found in Lindskog (2000).

The inference functions for margins (IFM) approach to the estimation of copulas (method (1) of Section 7.5.2 followed by maximization of (7.55)) is described by

Joe (1997), who gives asymptotic theory; the name of the approach (IFM) follows terminology of McLeish and Small (1988).

The pseudo-likelihood approach to copula estimation (method (2) of Section 7.5.2 followed by maximization of (7.55)) is described in Genest and Rivest (1993), and the consistency and asymptotic normality of the resulting parameter estimates are demonstrated. In Monte Carlo simulations it is found that this method outperforms the Kendall's tau method for a bivariate Clayton copula (see also Genest, Ghoudi and Rivest 1995).

Frees and Valdez (1997) discuss the relevance of copulas in actuarial applications and give an example where copulas are fitted to data using the Kendall's tau method and the IFM method. Also in an insurance context, Klugman and Parsa (1999) discuss ML inference for copulas and bivariate goodness-of-fit tests, while Chen and Fan (2005) describe a likelihood-ratio test for semi-parametric copula selection.

The fitting of the t copula to data and statistical aspects of testing this copula against the Gauss copula are discussed at length in Mashal and Zeevi (2002); the technical problem that the Gauss copula is a boundary case of the t copula is addressed in this paper and a correction is suggested. The authors provide a number of financial examples suggesting that extremal dependence is a feature of financial data. Breymann, Dias and Embrechts (2003) fit various bivariate copulas to high-frequency financial return data at different timescales and provide extensive comparisons with respect to goodness-of-fit.

There is a growing literature on goodness-of-fit tests for copulas: see the survey article by Genest, Rémillard and Beaudoin (2009). For an attractive graphical test, see Hofert and Mächler (2014).

Papers that develop dynamic time-series models for financial return data using copulas include Chen and Fan (2006), Patton (2004, 2006) and Fortin and Kuzmics (2002). A change-point problem for copulas within econometrics is discussed in Dias and Embrechts (2009).

8
Aggregate Risk

This chapter is devoted to a number of theoretical concepts in quantitative risk management that fall under the broad heading of aggregate risk and integrated risk management. We understand aggregate risk as the risk of a portfolio, which could even be the entire position in risky assets of a financial institution. The material builds on general ideas in risk measurement discussed in Section 2.3 and also uses in certain places the copula theory of Chapter 7 and some facts about elliptical distributions from Section 6.3.

In Sections 8.1–8.3 we treat the issue of measuring aggregate risk. We begin with general results and we discuss, in particular, the dual representation of convex risk measures as generalized scenarios (a mathematical extension of the idea of a stress test). Next we consider certain law-invariant risk measures (risk measures that depend only on the loss distribution). Finally, we apply the representation result to the case of linear portfolios and we discuss risk measurement for the special case of portfolios that are linear combinations of elliptically distributed risk factors.

Section 8.4 is concerned with risk aggregation: we assume that risk capital numbers for sub-units of an enterprise have been computed and we discuss methods for aggregating these risk capital numbers into a capital requirement for the entire enterprise. Moreover, we consider the problem of bounding an aggregate risk if we know something about the individual risks that contribute to the whole but have only limited information about their dependence. We discuss specific difficulties that arise when risk is measured with a non-subadditive risk measure like VaR. Finally, in Section 8.5, we treat the subject of allocating risk capital for an aggregate risk back to the individual risks in the portfolio. This issue is relevant for the purposes of performance measurement, loan pricing and capital budgeting.

8.1 Coherent and Convex Risk Measures

In this section we present elements of the modern theory of risk measures. Our exposition is a simplified account of material found in Föllmer and Schied (2011). We begin by recalling the axioms characterizing coherent and convex risk measures. For the economic motivation of these axioms we refer to Section 2.3.5.

Consider a probability space $(\Omega, \mathcal{F}, P)$ and a linear space $\mathcal{M} \subset \mathcal{L}^0(\Omega, \mathcal{F}, P)$, where $\mathcal{L}^0(\Omega, \mathcal{F}, P)$ denotes the set of all random variables on $(\Omega, \mathcal{F}, P)$ that are almost surely (a.s.) finite. Each $L \in \mathcal{M}$ represents the loss incurred on a financial position over some fixed time horizon. We assume throughout that constant random

variables belong to $\mathcal{M}$ and denote them by lowercase letters. In this context a *risk measure* is a mapping $\varrho\colon \mathcal{M} \to \mathbb{R}$ with the interpretation that $\varrho(L)$ gives the total amount of capital that is needed to back a position with loss L. The axioms for ϱ are as follows.

Monotonicity. $L_1 \leqslant L_2 \Rightarrow \varrho(L_1) \leqslant \varrho(L_2)$.

Translation invariance. For $m \in \mathbb{R}$, $\varrho(L + m) = \varrho(L) + m$.

Subadditivity. For $L_1, L_2 \in \mathcal{M}$, $\varrho(L_1 + L_2) \leqslant \varrho(L_1) + \varrho(L_2)$.

Positive homogeneity. For $\lambda \geqslant 0$, $\varrho(\lambda L) = \lambda\varrho(L)$.

Convexity. For $0 \leqslant \gamma \leqslant 1$, $L_1, L_2 \in \mathcal{M}$,

$$\varrho(\gamma L_1 + (1-\gamma)L_2) \leqslant \gamma\varrho(L_1) + (1-\gamma)\varrho(L_2).$$

Definition 8.1. A risk measure that satisfies the monotonicity, translation invariance and convexity axioms is called a *convex* measure of risk; a risk measure that satisfies the monotonicity, translation invariance, subadditivity and positive homogeneity axioms is called a *coherent* measure of risk.

A coherent risk measure is automatically convex; the converse implication is not true, as will be seen below. On the other hand, for a positive-homogeneous risk measure, convexity and coherence are equivalent.

8.1.1 Risk Measures and Acceptance Sets

There is an important relationship between risk measures and so-called *acceptance sets*. For a given risk measure the associated acceptance set contains the positions that are acceptable without any backing capital.

Definition 8.2. For a monotone and translation-invariant risk measure ϱ, the associated acceptance set of ϱ is the set

$$A_\varrho = \{L \in \mathcal{M}\colon \varrho(L) \leqslant 0\}. \tag{8.1}$$

Proposition 8.3. *For a monotone and translation-invariant risk measure ϱ with associated acceptance set A_ϱ, the following statements hold.*

(1) *A_ϱ is nonempty and satisfies the condition*

$$L \in A_\varrho \text{ and } \tilde{L} \leqslant L \Rightarrow \tilde{L} \in A_\varrho. \tag{8.2}$$

(2) *ϱ can be reconstructed from A_ϱ via*

$$\varrho(L) = \inf\{m \in \mathbb{R}\colon L - m \in A_\varrho\}. \tag{8.3}$$

Proof. Statement (1) is obvious. For (2) note that

$$\inf\{m\colon L - m \in A_\varrho\} = \inf\{m\varrho(L-m) \leqslant 0\} = \inf\{m\colon \varrho(L) - m \leqslant 0\},$$

and this is obviously equal to $\varrho(L)$. □

Conversely, it is sometimes useful to start with a set $A \subset \mathcal{M}$ of acceptable positions and to define an associated risk measure ϱ_A using (8.3). The properties of such a risk measure are given in the following proposition.

Proposition 8.4. *Suppose that the set A satisfies (8.2) and define ϱ_A by*

$$\varrho_A(L) = \inf\{m \in \mathbb{R} \colon L - m \in A\}. \tag{8.4}$$

Suppose, moreover, that $\varrho_A(L)$ defined in this way is finite for all $L \in \mathcal{M}$. Then ϱ_A is a monotone and translation-invariant risk measure on $\mathcal{M}$. The associated acceptance A_{ϱ_A} satisfies $A_{\varrho_A} \supseteq A$.

Proof. These properties of ϱ_A are easily checked. □

Remark 8.5. It is natural to enquire when the sets A and A_{ϱ_A} in Proposition 8.4 are equal. One result in that direction is given in Section 4.1 of Föllmer and Schied (2011) for the case where $\mathcal{M}$ contains only bounded random variables: in that case, $A = A_{\varrho_A}$ if and only if the set A is closed in the supremum norm.

In the next proposition we require some further basic ideas from convex analysis. A set $C \subset \mathcal{M}$ is said to be convex if $(1-\gamma)x + \gamma y \in C$ whenever $x \in C$, $y \in C$ and $0 < \gamma < 1$. A convex set is a convex cone if it has the additional property that it is closed under positive scalar multiplication, i.e. $\lambda x \in C$ when $x \in C$ and $\lambda > 0$.

Proposition 8.6.

(a) *Consider a monotone and translation-invariant risk measure ϱ with associated acceptance set A_ϱ defined by (8.1). Then*

  (a1) *ϱ is a convex risk measure if and only if A_ϱ is convex, and*

  (a2) *ϱ is coherent if and only if A_ϱ is a convex cone.*

(b) *More generally, consider a set of acceptable positions A and the associated risk measure ϱ_A defined by (8.4) (whose acceptance set may be larger than A). Then ϱ_A is a convex risk measure if A is convex and ϱ_A is coherent if A is a convex cone.*

Proof. For part (a1) it is clear that A_ϱ is convex if ϱ is convex. For the converse direction, consider arbitrary $L_1, L_2 \in \mathcal{M}$ and $0 \leqslant \gamma \leqslant 1$. Now for $i = 1, 2$, $L_i - \varrho(L_i) \in A_\varrho$ by definition of A_ϱ. Since A_ϱ is convex, we also have that $\gamma(L_1 - \varrho(L_1)) + (1-\gamma)(L_2 - \varrho(L_2)) \in A_\varrho$. By the definition of A_ϱ and translation invariance we have

$$\begin{aligned} 0 &\geqslant \varrho(\gamma(L_1 - \varrho(L_1)) + (1-\gamma)(L_2 - \varrho(L_2))) \\ &= \varrho(\gamma L_1 + (1-\gamma)L_2) - (\gamma\varrho(L_1) + (1-\gamma)\varrho(L_2)), \end{aligned}$$

which implies the convexity of ϱ.

To prove (a2) assume that $L \in A_\varrho$. As ϱ is positive homogeneous, $\varrho(\lambda L) = \lambda\varrho(L) \leqslant 0$ and hence $\lambda L \in A_\varrho$. Conversely, for $L \in \mathcal{M}$ we have $L - \varrho(L) \in A_\varrho$ and, as A_ϱ is a convex cone, also $\lambda(L - \varrho(L)) \in A_\varrho$ for all λ. Hence, by translation invariance,

$$0 \geqslant \varrho(\lambda L - \lambda\varrho(L)) = \varrho(\lambda L) - \lambda\varrho(L). \tag{8.5}$$

For the opposite inequality note that for $m < \varrho(L)$, $L - m \notin A_\varrho$ and hence also $\lambda(L - m) \notin A_\varrho$ for all $\lambda > 0$. Hence

$$0 < \varrho(\lambda L - \lambda m) = \varrho(\lambda L) - \lambda m,$$

i.e. $\varrho(\lambda L) > \lambda m$. By taking the supremum we get $\varrho(\lambda L) \geqslant \sup\{\lambda m : m < \varrho(L)\} = \lambda\varrho(L)$; together with (8.5) the claim follows.

The proof of (b) uses similar arguments to parts (a1) and (a2) and is omitted. □

We now give a number of examples of risk measures and acceptance sets.

Example 8.7 (value-at-risk). Given a confidence level $\alpha \in (0, 1)$, suppose we call an rv $L \in \mathcal{M}$ acceptable if $P(L > 0) \leqslant 1 - \alpha$. The associated risk measure defined by (8.4) is given by

$$\varrho_\alpha(L) := \inf\{m \in \mathbb{R} : P(L - m > 0) \leqslant 1 - \alpha\} = \inf\{m \in \mathbb{R} : P(L \leqslant m) \geqslant \alpha\},$$

which is the VaR at confidence level α.

Example 8.8 (risk measures based on loss functions). Consider a function $\ell : \mathbb{R} \to \mathbb{R}$ that is strictly increasing and convex and some threshold $c \in \mathbb{R}$. Assume that $E(\ell(L))$ is finite for all $L \in \mathcal{M}$ and define an acceptance set by

$$A = \{L \in \mathcal{M} : E(\ell(L)) \leqslant \ell(c)\}$$

and the associated risk measure by

$$\varrho_A = \inf\{m \in \mathbb{R} : E(\ell(L - m)) \leqslant \ell(c)\}.$$

In this context ℓ is called a *loss function* because the convexity of ℓ serves to penalize large losses; a loss function can be derived from a utility function u (a strictly increasing and concave function) by setting $\ell(x) = -u(-x)$.

The set A obviously satisfies (8.2), so that ϱ_A is translation invariant and monotone by Proposition 8.4. Furthermore, A is convex. This can be seen by considering acceptable positions L_1 and L_2 and observing that the convexity of ℓ implies

$$\begin{aligned} E(\ell(\gamma L_1 + (1-\gamma)L_2)) &\leqslant E(\gamma\ell(L_1) + (1-\gamma)\ell(L_2)) \\ &\leqslant \gamma\ell(c) + (1-\gamma)\ell(c) \\ &= \ell(c), \end{aligned}$$

where we have used the fact that $E(\ell(L_i)) \leqslant \ell(c)$ for acceptable positions. Hence $\gamma L_1 + (1-\gamma)L_2 \in A$ as required, and ϱ_A is a convex measure of risk by Proposition 8.6.

As a specific example we take $\ell(x) = \mathrm{e}^{\alpha x}$ for some $\alpha > 0$. We get

$$\begin{aligned}\varrho_{\alpha,c}(L) &:= \inf\{m\colon E(\mathrm{e}^{\alpha(L-m)}) \leqslant \mathrm{e}^{\alpha c}\} = \inf\{m\colon E(\mathrm{e}^{\alpha L}) \leqslant \mathrm{e}^{\alpha c+\alpha m}\}\\ &= \frac{1}{\alpha}\ln\{E(\mathrm{e}^{\alpha L})\} - c.\end{aligned}$$

Note that $\varrho_{\alpha,c}(0) = -c$, which could not be true for a coherent risk measure. Consider the special case where $c = 0$ and write $\varrho_\alpha := \varrho_{\alpha,0}$. We find that for $\lambda > 1$,

$$\varrho_\alpha(\lambda L) = \frac{1}{\alpha}\ln\{E(\mathrm{e}^{\alpha\lambda L})\} \geqslant \frac{1}{\alpha}\ln\{E(\mathrm{e}^{\alpha L})^\lambda\} = \lambda\varrho_\alpha(L),$$

where the inequality is strict if the rv L is non-degenerate. This shows that ϱ_α is convex but not coherent. In insurance mathematics this risk measure is known as the exponential premium principle if the losses L are interpreted as claims.

Example 8.9 (stress-test or worst-case risk measure). Given a set of stress scenarios $S \subset \Omega$, a definition of a stress-test risk measure is

$$\varrho(L) = \sup\{L(\omega)\colon \omega \in S\},$$

i.e. the worst loss when we restrict our attention to those elements of the sample space Ω that belong to S. The associated acceptance set is $A_\varrho = \{L\colon L(\omega) \leqslant 0 \text{ for all } \omega \in S\}$, i.e. the losses that are non-positive for all stress scenarios. The crucial part of defining the risk measure is the choice of the scenario set S, which is often guided by probabilistic considerations involving the underlying measure P.

Example 8.10 (generalized scenarios). Consider a set $\mathcal{Q}$ of probability measures on $(\Omega, \mathcal{F})$ and a mapping $\gamma\colon \mathcal{Q} \to \mathbb{R}$ such that $\inf\{\gamma(Q)\colon Q \in \mathcal{Q}\} > -\infty$. Suppose that $\sup_{Q\in\mathcal{Q}} E_Q(|L|) < \infty$ for all $L \in \mathcal{M}$. Define a risk measure ϱ by

$$\varrho(L) = \sup\{E_Q(L) - \gamma(Q)\colon Q \in \mathcal{Q}\}. \tag{8.6}$$

Note that a measure $Q \in \mathcal{Q}$ such that $\gamma(Q)$ is large is penalized in the maximization in (8.6), so γ can be interpreted as a penalty function specifying the relevance of the various measures in $\mathcal{Q}$. The corresponding acceptance set is given by

$$A_\varrho = \{L \in \mathcal{M}\colon \sup\{E_Q(L) - \gamma(Q)\colon Q \in \mathcal{Q}\} \leqslant 0\}.$$

A_ϱ is obviously convex so that ϱ is a convex risk measure. In fact, every convex risk measure can be represented in the form (8.6), as will be shown in Theorem 8.11 below (at least for the case of finite Ω). In the case where $\gamma(\cdot) \equiv 0$ on $\mathcal{Q}$, ϱ is obviously positive homogeneous and therefore coherent.

The stress-test risk measure of Example 8.9 is a special case of (8.6) in which the penalty function is equal to zero and in which $\mathcal{Q}$ is the set of all Dirac measures $\delta_\omega(\cdot)$, $\omega \in S$, i.e. measures such that $\delta_\omega(B) = I_B(\omega)$ for arbitrary measurable sets $B \subset \Omega$. Since in (8.6) we may choose more general sets of probability measures than simply Dirac measures, risk measures of the form (8.6) are frequently referred to as generalized scenario risk measures.

8.1.2 Dual Representation of Convex Measures of Risk

In this section we show that convex measures of risk have dual representations as generalized scenario risk measures. We state and prove a theorem in the simpler setting of a finite probability space. However, the result can be extended to general probability spaces by imposing additional continuity conditions on the risk measure. See, for instance, Sections 4.2 and 4.3 of Föllmer and Schied (2011).

Theorem 8.11 (dual representation for risk measures). *Suppose that Ω is a finite probability space with $|\Omega| = n < \infty$. Let $\mathcal{F} = \mathcal{P}(\Omega)$, the set of all subsets of Ω, and take $\mathcal{M} := \{L\colon \Omega \to \mathbb{R}\}$. Then the following hold.*

(1) *Every convex risk measure ϱ on $\mathcal{M}$ can be written in the form*

$$\varrho(L) = \max\{E_Q(L) - \alpha_{\min}(Q)\colon Q \in \mathcal{I}^1(\Omega, \mathcal{F})\}, \tag{8.7}$$

where $\mathcal{I}^1(\Omega, \mathcal{F})$ denotes the set of all probability measures on Ω, and where the penalty function $\alpha_{\min}$ is given by

$$\alpha_{\min}(Q) = \sup\{E_Q(L)\colon L \in A_\varrho\}. \tag{8.8}$$

(2) *If ϱ is coherent, it has the representation*

$$\varrho(L) = \max\{E_Q(L)\colon Q \in \mathcal{Q}\} \tag{8.9}$$

for some set $\mathcal{Q} = \mathcal{Q}(\varrho) \subset \mathcal{I}^1(\Omega, \mathcal{F})$.

Proof. The proof is divided into three steps. For simplicity we write $\mathcal{I}^1$ for $\mathcal{I}^1(\Omega, \mathcal{F})$.

Step 1. First we show that for $L \in \mathcal{M}$,

$$\varrho(L) \geqslant \sup\{E_Q(L) - \alpha_{\min}(Q)\colon Q \in \mathcal{I}^1\}. \tag{8.10}$$

To establish (8.10), set $L' := L - \varrho(L)$ and note that $L' \in A_\varrho$ so that

$$\alpha_{\min}(Q) = \sup\{E_Q(L)\colon L \in A_\varrho\} \geqslant E_Q(L') = E_Q(L) - \varrho(L).$$

This gives $\varrho(L) \geqslant E_Q(L) - \alpha_{\min}(Q)$, and taking the supremum over different measures Q gives (8.10).

Step 2. The main step of the proof is now to construct for $L \in \mathcal{M}$ a measure $Q_L \in \mathcal{I}^1$ such that $\varrho(L) \leqslant E_{Q_L}(L) - \alpha_{\min}(Q_L)$; together with (8.10) this establishes (8.7) and (8.8). This is the most technical part of the argument and we give a simple illustration in Example 8.12.

By translation invariance it is enough to construct Q_L for a loss L with $\varrho(L) = 0$; moreover, we may assume without loss of generality that $\varrho(0) = 0$. Since $|\Omega| = n < \infty$ we may identify L with the vector of possible outcomes $\boldsymbol{\ell} = (L(\omega_1), \dots, L(\omega_n))'$ in $\mathbb{R}^n$. Similarly, a probability measure $Q \in \mathcal{I}^1$ can be identified with the vector of corresponding probabilities $\boldsymbol{q} = (q(\omega_1), \dots, q(\omega_n))'$, which is an element of the unit simplex in $\mathbb{R}^n$. We may identify A_ϱ with a convex subset $\mathcal{A}_\varrho$ of $\mathbb{R}^n$. Note that a

loss L with $\varrho(L) = 0$ is not in the interior of A_ϱ. Otherwise $L + \epsilon$ would belong to A_ϱ for $\epsilon > 0$ small enough, which would imply that $0 \geqslant \varrho(L + \epsilon) = \varrho(L) + \epsilon = \epsilon$, which is a contradiction. Hence $\boldsymbol{\ell}$ is not in the interior of $\mathcal{A}_\varrho$. According to the supporting hyperplane theorem (see Proposition A.8 in Appendix A.1.5) there therefore exists a vector $\boldsymbol{u} \in \mathbb{R}^n \setminus \{0\}$ such that

$$\boldsymbol{u}'\boldsymbol{\ell} \geqslant \sup\{\boldsymbol{u}'\boldsymbol{x} : \boldsymbol{x} \in \mathcal{A}_\varrho\}. \tag{8.11}$$

Below we will construct Q_L from the vector $\boldsymbol{u}$. First, we show that $u_j \geqslant 0$ for all $1 \leqslant j \leqslant n$. Since $\varrho(0) = 0$, by monotonicity and translation invariance we have $\varrho(L - 1 - \lambda I_{\{\omega_j\}}) < 0$ for all j and all $\lambda > 0$. This shows that $L - 1 - \lambda I_{\{\omega_j\}} \in A_\varrho$ or, equivalently, $\boldsymbol{\ell} - \mathbf{1} - \lambda \boldsymbol{e}_j \in \mathcal{A}_\varrho$, where we use the notation $\mathbf{1} = (1, \ldots, 1)'$. Applying relation (8.11) we get that $\boldsymbol{u}'\boldsymbol{\ell} \geqslant \boldsymbol{u}'(\boldsymbol{\ell} - \mathbf{1} - \lambda \boldsymbol{e}_j)$ for all $\lambda > 0$, which implies that

$$0 \geqslant -\sum_{j=1}^{n} u_j - \lambda u_j, \quad \forall \lambda > 0.$$

This is possible only for $u_j \geqslant 0$. Since $\boldsymbol{u}$ is different from zero, as least one of the components must be strictly positive, and we can define the vector $\boldsymbol{q} := \boldsymbol{u}/(\sum_{j=1}^n u_j)$. Note that $\boldsymbol{q}$ belongs to the unit simplex in $\mathbb{R}^n$, and we define Q_L to be the associated probability measure. It remains to verify that Q_L satisfies the inequality $\varrho(L) \leqslant E_{Q_L}(L) - \alpha_{\min}(Q_L)$; since we assumed $\varrho(L) = 0$, we need to show that

$$E_{Q_L}(L) \geqslant \alpha_{\min}(Q). \tag{8.12}$$

We have $\alpha_{\min}(Q_L) = \sup\{E_{Q_L}(X) : X \in A_\varrho\} = \sup\{\boldsymbol{q}'\boldsymbol{x} : \boldsymbol{x} \in \mathcal{A}_\varrho\}$. It follows from (8.11) that

$$E_{Q_L}(L) = \boldsymbol{q}'\boldsymbol{\ell} \geqslant \sup\{\boldsymbol{q}'\boldsymbol{x} : \boldsymbol{x} \in \mathcal{A}_\varrho\} = \alpha_{\min}(Q_L),$$

and hence we obtain the desired relation (8.12).

Step 3. In order to establish the representation (8.9) for coherent risk measures, we recall that for a coherent risk measure ϱ the acceptance set A_ϱ is a convex cone. Hence for $\lambda > 0$ we obtain

$$\alpha_{\min}(Q) = \sup_{L \in A_\varrho} E_Q(L) = \sup_{\lambda L \in A_\varrho} E_Q(\lambda L) = \lambda \alpha_{\min}(Q),$$

which is possible only for $\alpha_{\min}(Q) \in \{0, \infty\}$. The representation (8.9) follows if we set $\mathcal{Q}(\varrho) := \{Q \in \mathcal{S}^1 : \alpha_{\min}(Q) = 0\}$. □

Example 8.12. To illustrate the construction in step (2) of the proof of Theorem 8.11 we give a simple example (see Figure 8.1 to visualize the construction).

Consider the case where $d = 2$ and where the risk measure is $\varrho(L) = \ln E(\mathrm{e}^L)$, the exponential premium principle of Example 8.8 with $\alpha = 1$ and $c = 0$. Assume that the probability measure P is given by $\boldsymbol{p} = (0.4, 0.6)$, in which case the losses L

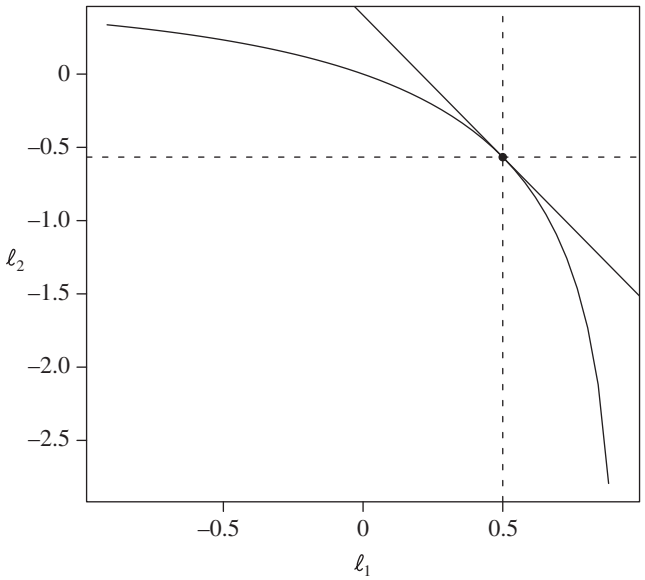

Figure 8.1. Illustration of the measure construction in step (2) of the proof of Theorem 8.11 (see Example 8.12 for details).

for which $\varrho(L) = 0$ are represented by the curve with equation $f_\varrho(\boldsymbol{\ell}) := \ln(0.4\mathrm{e}^{\ell_1} + 0.6\mathrm{e}^{\ell_2}) = 0$; the acceptance set A_ϱ consists of the curve and the region below it. We now construct Q_L for the loss L on the curve with $L(\omega_1) = \ell_1 = 0.5$ and $L(\omega_2) = \ell_2 \approx -0.566$. A possible choice for the vector $\boldsymbol{u}$ in (8.11) (the normal vector of the supporting hyperplane) is to take

$$\boldsymbol{u} = \nabla f_\rho(\boldsymbol{\ell}) = \left(\frac{\partial f_\varrho(\boldsymbol{\ell})}{\partial \ell_1}, \frac{\partial f_\varrho(\boldsymbol{\ell})}{\partial \ell_2} \right)'.$$

By normalization of $\boldsymbol{u}$ the measure Q_L may be identified with $\boldsymbol{q} \approx (0.659, 0.341)$, and for this measure the penalty is $\alpha_{\min}(Q_L) = \boldsymbol{q}'\boldsymbol{\ell} \approx 0.137$.

Properties of $\alpha_{\min}$. Next we discuss properties of the penalty function $\alpha_{\min}$. First we explain that $\alpha_{\min}$ is a minimal penalty function representing ϱ. Suppose that $\alpha\colon \mathcal{I}^1(\Omega, \mathcal{F}) \to \mathbb{R}$ is any function such that (8.7) holds for all $L \in M$. Then for $Q \in \mathcal{I}^1(\Omega, \mathcal{F})$, $L \in \mathcal{M}$ fixed we have $\varrho(L) \geqslant E_Q(L) - \alpha(Q)$ and hence also

$$\begin{aligned} \alpha(Q) \geqslant \sup_{L\in\mathcal{M}} \{E_Q(L) - \varrho(L)\} &\geqslant \sup_{L\in A_\varrho} \{E_Q(L) - \varrho(L)\} \\ &\geqslant \sup_{L\in A_\varrho} E_Q(L) = \alpha_{\min}(Q). \end{aligned} \tag{8.13}$$

Next we give an alternative representation of $\alpha_{\min}$ that will be useful in the analysis of risk measures for linear portfolios. Consider some $L \in \mathcal{M}$. Since $L - \varrho(L) \in A_\varrho$ we get that

$$\sup_{L\in\mathcal{M}} \{E_Q(L) - \varrho(L)\} = \sup_{L\in\mathcal{M}} \{E_Q(L - \varrho(L))\} \leqslant \sup_{L\in A_\varrho} E_Q(L) = \alpha_{\min}(Q).$$

Combining this with the previous estimate (8.13) we obtain the relation

$$\alpha_{\min}(Q) = \sup_{L \in \mathcal{M}} \{E_Q(L) - \varrho(L)\}. \tag{8.14}$$

8.1.3 Examples of Dual Representations

In this section we give a detailed derivation of the dual representation for expected shortfall, and we briefly discuss the dual representation for the risk measure based on the exponential loss function.

Expected shortfall. Recall from Section 2.3.4 that expected shortfall is given by

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha} \int_\alpha^1 q_u(L)\, \mathrm{d}u, \quad \alpha \in [0, 1),$$

for integrable losses L. The following lemma gives alternative expressions for ES_α.

Proposition 8.13. *For $0 < \alpha < 1$ we have*

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha} E((L - q_\alpha(L))^+) + q_\alpha(L) \tag{8.15}$$

$$= \frac{1}{1-\alpha} (E(L; L > q_\alpha(L)) + q_\alpha(L)(1 - \alpha - P(L > q_\alpha(L)))). \tag{8.16}$$

Proof. Recall that for $U \sim U(0, 1)$ the random variable $F^{\leftarrow}(U)$ has distribution function F_L (see Proposition 7.2). Since $q_\alpha(L) = F_L^{\leftarrow}(\alpha)$, (8.15) follows from observing that

$$\begin{aligned}
\frac{1}{1-\alpha} E((L - q_\alpha(L))^+) &= \frac{1}{1-\alpha} \int_0^1 (F_L^{\leftarrow}(u) - F_L^{\leftarrow}(\alpha))^+ \, \mathrm{d}u \\
&= \frac{1}{1-\alpha} \int_\alpha^1 (F_L^{\leftarrow}(u) - F_L^{\leftarrow}(\alpha)) \, \mathrm{d}u \\
&= \frac{1}{1-\alpha} \int_\alpha^1 q_u(L) \, \mathrm{d}u - q_\alpha(L).
\end{aligned}$$

For (8.16) we use $E((L - q_\alpha(L))^+) = E(L; L > q_\alpha(L)) - q_\alpha(L) P(L > q_\alpha(L))$. □

If F_L is continuous at $q_\alpha(L)$, then $P(L > q_\alpha(L)) = 1 - \alpha$ and

$$\mathrm{ES}_\alpha(L) = \frac{E(L; L > q_\alpha(L))}{1-\alpha} = E(L \mid L > \mathrm{VaR}_\alpha)$$

(see also Lemma 2.13).

Theorem 8.14. *For $\alpha \in [0, 1)$, ES_α defines a coherent measure of risk on $\mathcal{M} = \mathcal{L}^1(\Omega, \mathcal{F}, P)$. The dual representation is given by*

$$\mathrm{ES}_\alpha(L) = \max\{E^Q(L) \colon Q \in \mathcal{Q}_\alpha\}, \tag{8.17}$$

where $\mathcal{Q}_\alpha$ is the set of all probability measures on $(\Omega, \mathcal{F})$ that are absolutely continuous with respect to P and for which the measure-theoretic density $\mathrm{d}Q/\mathrm{d}P$ is bounded by $(1-\alpha)^{-1}$.

Proof. For $\alpha = 0$ one has $\mathcal{Q}_\alpha = \{P\}$ and the relation (8.17) is obvious so we consider the case where $\alpha > 0$. By translation invariance we can assume without loss of generality that $q_\alpha(L) > 0$. Since expected shortfall is only concerned with the upper tail, we can also assume that $L \geqslant 0$. (If it were not, we could simply define $\tilde{L} = \max(L, 0)$ and observe that $\mathrm{ES}_\alpha(\tilde{L}) = \mathrm{ES}_\alpha(L)$.)

Define a coherent measure ϱ_α by

$$\varrho_\alpha(L) = \sup\{E^Q(L) \colon Q \in \mathcal{Q}_\alpha\}. \tag{8.18}$$

We want to show that $\varrho_\alpha(L)$ can be written in the form of Proposition 8.13. As a first step we transform the optimization problem in (8.18). The measures in the set $\mathcal{Q}_\alpha$ can alternatively be described in terms of their measure-theoretic density and hence by the set of random variables $\{\psi \colon 0 \leqslant \psi \leqslant 1/(1-\alpha),\ E(\psi) = 1\}$. We therefore have

$$\varrho_\alpha(L) = \sup\left\{E(L\psi) \colon 0 \leqslant \psi \leqslant \frac{1}{1-\alpha},\ E(\psi) = 1\right\}.$$

Transforming these random variables according to $\varphi = (1-\alpha)\psi$ and factoring out the expression $E(L)$ we get

$$\varrho_\alpha(L) = \frac{E(L)}{1-\alpha} \sup\left\{E\left(\frac{L}{E(L)}\varphi\right) \colon 0 \leqslant \varphi \leqslant 1,\ E(\varphi) = 1-\alpha\right\}.$$

It follows that

$$\varrho_\alpha(L) = \frac{E(L)}{1-\alpha} \sup\{\tilde{E}(\varphi) \colon 0 \leqslant \varphi \leqslant 1,\ E(\varphi) = 1-\alpha\}, \tag{8.19}$$

where the measure $\tilde{P}$ is defined by $\mathrm{d}\tilde{P}/\mathrm{d}P = L/E(L)$.

We now show that the supremum in (8.19) is attained by the random variable

$$\varphi_0 = I_{\{L > q_\alpha(L)\}} + \kappa I_{\{L = q_\alpha(L)\}}, \tag{8.20}$$

where $\kappa \geqslant 0$ is chosen such that $E(\varphi_0) = (1-\alpha)$. To verify the optimality of φ_0 consider an arbitrary $0 \leqslant \varphi \leqslant 1$ with $E(\varphi) = (1-\alpha)$. By definition of φ_0, we must have the inequality

$$0 \leqslant (\varphi_0 - \varphi)(L - q_\alpha(L)), \tag{8.21}$$

as the first factor is nonnegative for $L > q_\alpha(L)$ and nonpositive for $L < q_\alpha(L)$. Integration of (8.21) gives

$$0 \leqslant E((\varphi_0 - \varphi)(L - q_\alpha(L))) = E(L(\varphi_0 - \varphi)) - q_\alpha(L)E(\varphi_0 - \varphi).$$

The second term now vanishes as $E(\varphi_0) = E(\varphi) = 1-\alpha$, and the first term equals $E(L)\tilde{E}(\varphi_0 - \varphi)$. It follows that $\tilde{E}(\varphi_0) \geqslant \tilde{E}(\varphi)$, verifying the optimality of φ_0.

Inserting φ_0 in (8.19) gives

$$\begin{aligned}
\varrho_\alpha(L) &= \frac{E(L)}{(1-\alpha)}\tilde{E}(\varphi_0) \\
&= \frac{1}{1-\alpha}E(L\varphi_0) \\
&= \frac{1}{1-\alpha}(E(L;\, L > q_\alpha(L)) + \kappa q_\alpha(L)P(L = q_\alpha(L))).
\end{aligned} \tag{8.22}$$

The condition $E(\varphi_0) = (1-\alpha)$ yields that $P(L > q_\alpha(L)) + \kappa P(L = q_\alpha(L)) = (1-\alpha)$ and hence that

$$\kappa = \frac{(1-\alpha) - P(L > q_\alpha(L))}{P(L = q_\alpha(L))},$$

where we use the convention $0/0 = 0$. Inserting κ in (8.22) gives (8.16), which proves the theorem. □

Remark 8.15. Readers familiar with mathematical statistics may note that the construction of φ_0 in the optimization problem (8.19) is similar to the construction of the optimal test in the well-known Neyman–Pearson Lemma. The proof shows that for $\alpha \in (0, 1)$ and integrable L, a measure $Q_L \in \mathcal{Q}_\alpha$ that attains the supremum in the dual representation of ES_α is given by the measure-theoretic density

$$\frac{\mathrm{d}Q_L}{\mathrm{d}P} = \frac{\varphi_0}{(1-\alpha)} = \frac{1}{1-\alpha}(I_{\{L > q_\alpha(L)\}} + \kappa I_{\{L = q_\alpha(L)\}}),$$

where

$$\kappa = \begin{cases} \dfrac{(1-\alpha) - P(L > q_\alpha(L))}{P(L = q_\alpha(L))}, & P(L = q_\alpha(L)) > 0, \\ 0, & P(L = q_\alpha(L)) = 0. \end{cases}$$

Risk measure for the exponential loss function. We end this section by giving without proof the dual representation for the risk measure derived from the acceptance set of the exponential loss function, given by

$$\varrho_\alpha(L) = \frac{1}{\alpha} \log\{E(\mathrm{e}^{\alpha L})\}$$

(see Example 8.8 for details). In this case we have a non-zero penalty function since ϱ_α is convex but not coherent. It can be shown that

$$\varrho_\alpha(L) = \max_{Q \in \mathcal{Q}_1} \left\{ E_Q(L) - \frac{1}{\alpha} H(Q \mid P) \right\},$$

where

$$H(Q \mid P) = \begin{cases} E_Q\left(\ln \dfrac{\mathrm{d}Q}{\mathrm{d}P} \right) & \text{if } Q \ll P, \\ \infty & \text{otherwise,} \end{cases}$$

is known as *relative entropy* between P and Q. In other words, the penalty function is proportional to the relative entropy between the two measures. The proof of this fact requires a detailed study of the properties of relative entropy and is therefore omitted (see Sections 3.2 and 4.9 of Föllmer and Schied (2011) for details).

Notes and Comments

The classic paper on coherent risk measures is Artzner et al. (1999); a non-technical introduction by the same authors is Artzner et al. (1997). Technical extensions such as the characterization of coherent risk measures on infinite probability spaces are given

in Delbaen (2000, 2002, 2012). Stress testing as an approach to risk measurement is studied by Berkowitz (2000), Kupiec (1998), Breuer et al. (2009) and Rebonato (2010), among others.

The study of convex risk measures in the context of risk management and mathematical finance began with Föllmer and Schied (2002) (see also Frittelli 2002). A good treatment at advanced textbook level is given in Chapter 4 of Föllmer and Schied (2011). Cont (2006) provides an interesting link between convex risk measures and model risk in the pricing of derivatives. An alternative proof of Theorem 8.11 can be based on the duality theorem for convex functions (see, for instance, Remark 4.18 of Föllmer and Schied (2011)).

Different existing notions of expected shortfall are discussed in the very readable paper by Acerbi and Tasche (2002). Expected shortfall has been independently studied by Rockafellar and Uryasev (2000, 2002) under the name *conditional value-at-risk*; in particular, these papers develop the idea that expected shortfall can be obtained as the value of a convex optimization problem.

There has been recent interest in the subject of multi-period risk measures, which take into account the evolution of the final value of a position over several time periods and consider the effect of intermediate information and actions. Important papers in this area include Artzner et al. (2007), Riedel (2004), Weber (2006) and Cheridito, Delbaen and Kupper (2005). A textbook treatment and further references can be found in Chapter 11 of Föllmer and Schied (2011).

8.2 Law-Invariant Coherent Risk Measures

A risk measure ϱ is termed *law invariant* if $\varrho(L)$ depends on L only via its df F_L; examples of law-invariant risk measures are VaR and expected shortfall. On the other hand, the stress-test risk measures of Example 8.9 are typically not law invariant. In this section we discuss a number of law-invariant and coherent risk measures that are frequently used in financial and actuarial studies.

8.2.1 Distortion Risk Measures

The class of distortion risk measures is an important class of coherent risk measures. These risk measures are presented in many different ways in the literature, and a variety of different names are used. We begin by summarizing the more important representations before investigating the properties of distortion risk measures. Finally, we consider certain parametric families of distortion risk measures.

Representations of distortion risk measures We begin with a general definition.

Definition 8.16 (distortion risk measure).

(1) A convex distortion function D is a convex, increasing and absolutely continuous function on $[0, 1]$ satisfying $D(0) = 0$ and $D(1) = 1$.

(2) The distortion risk measure associated with a convex distortion function D is defined by

$$\varrho(L) = \int_0^1 q_u(L)\,\mathrm{d}D(u). \tag{8.23}$$

Note that every convex distortion function is a distribution function on [0, 1]. The simplest example of a distortion risk measure is expected shortfall, which is obtained by taking the convex distortion function $D_\alpha(u) = (1-\alpha)^{-1}(u-\alpha)^+$. Clearly, a distortion risk measure is law invariant (it depends on the rv L only via the distribution of L), as it is defined as an average of the quantiles of L.

Since a convex distortion function D is absolutely continuous by definition, it can be written in the form $D(u) = \int_0^u \phi(s)\,\mathrm{d}s$ for an increasing, positive function ϕ (the right derivative of D). This yields the alternative representation

$$\varrho(L) = \int_0^1 q_u(L)\phi(u)\,\mathrm{d}u. \tag{8.24}$$

A risk measure of the form (8.24) is also known as a *spectral risk measure*, and the function ϕ is called the *spectrum*. It can be thought of as a weighting function applied to the quantiles of the distribution of L. In the case of expected shortfall, the spectrum is $\phi(u) = (1-\alpha)^{-1} I_{\{u \geqslant \alpha\}}$, showing that an equal weight is placed on all quantiles beyond the α-quantile.

A second alternative representation is derived in the following lemma.

Lemma 8.17. *The distortion risk measure ϱ associated with a convex distortion function D can be written in the form*

$$\varrho(L) = \int_{\mathbb{R}} x\,\mathrm{d}D \circ F_L(x), \tag{8.25}$$

where $D \circ F_L$ denotes the composition of the functions D and F_L, that is, $D \circ F_L(x) = D(F_L(x))$.

Proof. Let $G(x) = D \circ F_L(x)$ and note that G is itself a distribution function. The associated quantile function is given by $G^{\leftarrow} = F_L^{\leftarrow} \circ D^{\leftarrow}$, as can be seen by using Proposition A.3 (iv) to write

$$G^{\leftarrow}(v) = \inf\{x\colon D \circ F_L(x) \geqslant v\} = \inf\{x\colon F_L(x) \geqslant D^{\leftarrow}(v)\}$$

and noting that this equals $F_L^{\leftarrow} \circ D^{\leftarrow}(v)$ by definition of the generalized inverse. The right-hand side of (8.25) can therefore be written in the form

$$\int_{\mathbb{R}} x\,\mathrm{d}G(x) = \int_0^1 G^{\leftarrow}(u)\,\mathrm{d}u = \int_0^1 F_L^{\leftarrow} \circ D^{\leftarrow}(u)\,\mathrm{d}u = E(F_L^{\leftarrow} \circ D^{\leftarrow}(U)),$$

where U is a standard uniform random variable. Now introduce the random variable $V = D^{\leftarrow}(U)$, which has df D. We have shown that

$$\int_{\mathbb{R}} x\,\mathrm{d}D \circ F_L(x) = E(F^{\leftarrow}(V)) = \int_0^1 F_L^{\leftarrow}(v)\,\mathrm{d}D(v)$$

and thus established the result. □

The representation (8.25) gives more intuition for the idea of a distortion. The original df F_L is distorted by the function D. Moreover, for $u \in (0, 1)$ we note that $D(u) \leqslant u$, by the convexity of D, so that the distorted df $G = D \circ F$ places more mass on high values of L than the original df F.

Finally, we show that a distortion risk measure can be represented as a weighted average of expected shortfall over different confidence levels. To do this we fix a convex distortion D with associated distortion risk measure ϱ and spectrum ϕ (see (8.24)). From now on we work with the right-continuous version of ϕ. Since ϕ is increasing we can obtain a measure ν on $[0, 1]$ by setting $\nu([0, t]) := \phi(t)$ for $0 \leqslant t \leqslant 1$. Note that for every function $f : [0, 1] \to \mathbb{R}$ we have

$$\int_0^1 f(\alpha)\,\mathrm{d}\nu(\alpha) = f(0)\phi(0) + \int_0^1 f(\alpha)\,\mathrm{d}\phi(\alpha).$$

Moreover, we now define a further measure μ on $[0, 1]$ by setting

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(\alpha) = (1-\alpha), \quad \text{that is,} \quad \int_0^1 f(\alpha)\,\mathrm{d}\mu(\alpha) = f(0) + \int_0^1 f(\alpha)(1-\alpha)\,\mathrm{d}\phi(\alpha). \tag{8.26}$$

Now we may state the representation result for ϱ.

Proposition 8.18. *Let ϱ be a distortion risk measure associated with the convex distortion D and define the measure μ by (8.26). Then μ is a probability measure and we have the representation*

$$\varrho(L) = \int_0^1 \mathrm{ES}_\alpha(L)\,\mathrm{d}\mu(\alpha).$$

Proof. Using integration by parts and (8.26) we check that

$$\begin{aligned}\mu([0,1]) &= \phi(0) + \int_0^1 (1-\alpha)\,\mathrm{d}\phi(\alpha) = \phi(0) + \int_0^1 \phi(\alpha)\,\mathrm{d}\alpha - \phi(0)\\ &= \int_0^1 \phi(\alpha)\,\mathrm{d}\alpha = D(1) - D(0) = 1,\end{aligned}$$

which shows that μ is a probability measure. Next we turn to the representation result for ϱ. By Fubini's Theorem we have that

$$\begin{aligned}\varrho(L) &= \int_0^1 q_u(L)\phi(u)\,\mathrm{d}u = \int_0^1 q_u(L) \int_0^u 1\,\mathrm{d}\nu(\alpha)\,\mathrm{d}u\\ &= \int_0^1 \int_0^1 q_u(L) 1_{\{\alpha \leqslant u\}} \mathrm{d}\nu(\alpha)\,\mathrm{d}u = \int_0^1 \int_\alpha^1 q_u(L)\,\mathrm{d}u\,\mathrm{d}\nu(\alpha)\\ &= \int_0^1 (1-\alpha)\,\mathrm{ES}_\alpha(L)\,\mathrm{d}\nu(\alpha) = \int_0^1 \mathrm{ES}_\alpha(L)\,\mathrm{d}\mu(\alpha).\end{aligned}$$

□

Properties of distortion risk measures. Next we discuss certain properties of distortion risk measures. Distortion risk measures are *comonotone additive* in the following sense.

Definition 8.19 (comonotone additivity). A risk measure ϱ on a space of random variables $\mathcal{M}$ is said to be comonotone additive if

$$\varrho(L_1 + \cdots + L_d) = \varrho(L_1) + \cdots + \varrho(L_d)$$

whenever $(L_1, \ldots, L_d)$ is a vector of comonotonic risks.

The property of comonotonicity was defined in Definition 7.2.1, and it was shown in Proposition 7.20 that the quantile function (or, in other words, the value-at-risk risk measure) is additive for comonotonic risks. The comonotone additivity of the distortion risk measures follows easily from the fact that they can be represented as weighted integrals of the quantile function as in (8.23).

Moreover, distortion risk measure are coherent. Monotonicity, translation invariance and positive homogeneity are obvious. To establish that they are coherent we only need to check subadditivity, which follows immediately from Proposition 8.18 and Theorem 8.14 by observing that

$$\begin{aligned}\varrho(L_1 + L_2) &= \int_0^1 \mathrm{ES}_\alpha(L_1 + L_2)\,\mathrm{d}\mu(\alpha) \\ &\leqslant \int_0^1 \mathrm{ES}_\alpha(L_1)\,\mathrm{d}\mu(\alpha) + \int_0^1 \mathrm{ES}_\alpha(L_2)\,\mathrm{d}\mu(\alpha) = \varrho(L_1) + \varrho(L_2).\end{aligned}$$

In summary, we have verified that distortion risk measures are law invariant, coherent and comonotone additive. In fact, it may also be shown that, on a probability space without atoms (i.e. a space where $P(\{\omega\}) = 0$ for all ω), a law-invariant, coherent, comonotone-additive risk measure must be of the form (8.23) for some convex distortion D.

Example 8.20 (parametric families). A number of useful parametric families of distortion risk measures can be based on convex distortion functions that take the form

$$D_\alpha(u) = \Psi(\Psi^{-1}(u) + \ln(1-\alpha)), \quad 0 \leqslant \alpha < 1, \tag{8.27}$$

where Ψ is a continuous df on $\mathbb{R}$. The distortion function for expected shortfall is obtained when $\Psi(u) = 1 - \mathrm{e}^{-u}$ for $u \geqslant 0$, the standard exponential df.

There has been interest in distortion risk measures that are obtained by considering different dfs Ψ that are strictly increasing on the whole real line $\mathbb{R}$. A natural question concerns the constraints on Ψ that lead to convex distortion functions. It is straightforward to verify, by differentiating $D_\alpha(u)$ in (8.27) twice with respect to u, that a necessary and sufficient condition is that $\ln \psi(u)$ is concave, where ψ denotes the density of Ψ (see Tsukahara 2009).

A family of convex distortion functions of the form (8.27) is strictly decreasing in α for fixed u. Moreover, $D_0(u) = u$ (corresponding to the risk measure $\varrho(L) = E(L)$) and $\lim_{\alpha\to 1} D(u) = 1_{\{u=1\}}$. The fact that for $\alpha_1 < \alpha_2$ and $0 < u < 1$ we have $D_{\alpha_1}(u) > D_{\alpha_2}(u)$ means that, roughly speaking, D_{α_2} distorts the original probability measure more than D_{α_1} and places more weight on outcomes in the tail.

A particular example is obtained by taking $\Psi(u) = 1/(1 + \mathrm{e}^{-u})$, the standard logistic df. This leads to the parametric family of convex distortion functions given by $D_\alpha(u) = (1-\alpha)u(1-\alpha u)^{-1}$ for $0 \leqslant \alpha < 1$. Writing $G_\alpha(x) = D_\alpha \circ F_L(x)$ we can show that

$$\left(\frac{1 - G_\alpha(x)}{G_\alpha(x)}\right) = \frac{1}{1-\alpha}\left(\frac{1 - F_L(x)}{F_L(x)}\right),$$

and this yields an interesting interpretation for this family. For every possibly critical loss level x, the *odds* of the tail event $\{L > x\}$ given by $(1 - F_L(x))/F_L(x)$ are

multiplied by $(1-\alpha)^{-1}$ under the distorted loss distribution. For this reason the family is known as the *proportional odds family*.

8.2.2 The Expectile Risk Measure

Definition 8.21. Let $\mathcal{M} := L^1(\Omega, \mathcal{F}, P)$, the set of all integrable random variables L with $E|L| < \infty$. Then, for $\alpha \in (0,1)$ and $L \in \mathcal{M}$, the α-expectile $e_\alpha(L)$ is given by the unique solution y of the equation

$$\alpha E((L-y)^+) = (1-\alpha)E((L-y)^-), \tag{8.28}$$

where $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$.

Recalling that $x^+ - x^- = x$, we note that $e_{0.5}(L) = E(L)$ since

$$\begin{aligned} E((L-y)^-) = E((L-y)^+) &\iff E((L-y)^+ - (L-y)^-) = 0 \\ &\iff E(L-y) = 0. \end{aligned}$$

For square-integrable losses L, the expectile $e_\alpha(L)$ can also be viewed as the minimizer in an optimization problem of the form

$$\min_{y\in\mathbb{R}} E(S(y, L)) \tag{8.29}$$

for a so-called *scoring function* $S(y, L)$. This could be relevant for the out-of-sample testing of expectile estimates (so-called backtesting), as will be explained in Section 9.3. The particular scoring function that yields the expectile is

$$S^e_\alpha(y, L) = |1_{\{L\leqslant y\}} - \alpha|(L-y)^2. \tag{8.30}$$

In fact, we can compute that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}y} E(S^e_\alpha(y, L)) &= \frac{\mathrm{d}}{\mathrm{d}y} \int_{-\infty}^{\infty} |1_{\{y\geqslant x\}} - \alpha|(y-x)^2 \,\mathrm{d}F_L(x) \\ &= \frac{\mathrm{d}}{\mathrm{d}y} \int_{-\infty}^{y} (1-\alpha)(y-x)^2 \,\mathrm{d}F_L(x) + \frac{\mathrm{d}}{\mathrm{d}y} \int_{y}^{\infty} \alpha(y-x)^2 \mathrm{d}F_L(x) \\ &= 2(1-\alpha) \int_{-\infty}^{y} (y-x) \,\mathrm{d}F_L(x) + 2\alpha \int_{y}^{\infty} (y-x) \,\mathrm{d}F_L(x) \\ &= 2(1-\alpha)E((L-y)^-) - 2\alpha E((L-y)^+), \end{aligned} \tag{8.31}$$

and setting this equal to zero yields the equation (8.28) that defines an expectile.

Remark 8.22. In Section 9.3.3 we will show that the α-quantile $q_\alpha(L)$ is also a minimizer in an optimization problem of the form (8.29) if we consider the scoring function

$$S^q_\alpha(y, L) = |1_{\{L\leqslant y\}} - \alpha| \, |L-y|. \tag{8.32}$$

We now show that the α-expectile of an arbitrary df F_L can be represented as the α-quantile of a related df $\tilde{F}_L$ that is strictly increasing on its support. This also shows the uniqueness of the α-expectile of a distribution. Moreover, we obtain a formula that can he helpful for computing expectiles of certain distributions and we illustrate this with a simple example.

Proposition 8.23. *Let $\alpha \in (0,1)$ and L be an rv such that $\mu := E(L) < \infty$. Then the solution $e_\alpha(L)$ of (8.28) may be written as $e_\alpha(L) = \tilde{F}_L^{-1}(\alpha)$, where*

$$\tilde{F}_L(y) = \frac{yF_L(y) - \mu(y)}{2(yF_L(y) - \mu(y)) + \mu - y} \tag{8.33}$$

is a continuous df that is strictly increasing on its support and $\mu(y) := \int_{-\infty}^{y} x \, \mathrm{d}F_L(x)$ is the lower partial moment of F_L.

Proof. Since $x^+ + x^- = |x|$, equation (8.31) shows that the expectile y must also solve

$$\alpha E(|L - y|) = E((L - y)^-) = \int_{-\infty}^{y} (y - x) \, \mathrm{d}F_L(x) = yF_L(y) - \mu(y).$$

Moreover,

$$\begin{aligned} E(|L - y|) &= \int_{-\infty}^{y} (y - x) \, \mathrm{d}F_L(x) + \int_{y}^{\infty} (x - y) \, \mathrm{d}F_L(x) \\ &= 2\int_{-\infty}^{y} (y - x) \, \mathrm{d}F_L(x) + \int_{-\infty}^{\infty} (x - y) \, \mathrm{d}F_L(x) \\ &= 2(yF_L(y) - \mu(y)) + \mu - y, \end{aligned}$$

and hence $\alpha = \tilde{F}_L(y)$ with $\tilde{F}_L$ as defined in (8.33).

Next we show that $\tilde{F}_L$ is indeed a distribution function. The derivative of $\tilde{F}_L$ can be easily computed to be

$$\tilde{f}_L(y) = \frac{\mu F_L(y) - \mu(y)}{(2(yF_L(y) - \mu(y)) + \mu - y)^2} = \frac{F_L(y)(\mu - E(L \mid L \leqslant y))}{(2(yF_L(y) - \mu(y)) + \mu - y)^2}.$$

Clearly, $\tilde{f}_L$ is nonnegative for all y and strictly positive on $D = \{y \colon 0 < F_L(y) < 1\}$, so that $\tilde{F}_L$ is increasing for all y and strictly increasing on D. Let $y_0 = \inf D$ and $y_1 = \sup D$ denote the left and right endpoints of F_L. It is then easy to check that $[y_0, y_1]$ is the support of $\tilde{F}_L$ and $\lim_{y \to y_0} \tilde{F}_L(y) = 0$ and $\lim_{y \to y_1} \tilde{F}_L(y) = 1$. □

In the following example we consider a Bernoulli distribution where the quantile is an unsatisfactory risk measure that can only take the values zero and one. In contrast, the expectile can take any value between zero and one.

Example 8.24. Let $L \sim \mathrm{Be}(p)$ be a Bernoulli-distributed loss. Then

$$F_L(y) = \begin{cases} 0, & y < 0, \\ 1 - p, & 0 \leqslant y < 1, \\ 1, & y \geqslant 1, \end{cases} \qquad \mu(y) = \begin{cases} 0, & y < 1, \\ p, & y \geqslant 1, \end{cases}$$

from which it follows that

$$\tilde{F}_L(y) = \frac{y(1-p)}{y(1-2p) + p}, \quad 0 \leqslant y \leqslant 1 \quad \text{and} \quad e_\alpha(L) = \frac{\alpha p}{(1-\alpha) + p(2\alpha - 1)}.$$

Properties of the expectile.

Proposition 8.25. *Provided $\alpha \geqslant 0.5$, the expectile risk measure $\varrho = e_\alpha$ is a coherent risk measure on $\mathcal{M} = L^1(\Omega, \mathcal{F}, P)$.*

Proof. For $L \in \mathcal{M}$ and $y \in \mathbb{R}$ define the function

$$\begin{aligned} g(L, y, \alpha) &= \alpha E((L-y)^+) - (1-\alpha)E((L-y)^-) \\ &= (2\alpha - 1)E((L-y)^+) + (1-\alpha)E(L-y) \end{aligned}$$

and note that, for fixed L, it is a decreasing function of y and, for fixed y, it is monotonic in L, so $L_1 \leqslant L_2 \Rightarrow g(L_1, y, \alpha) \leqslant g(L_2, y, \alpha)$.

Translation invariance and positive homogeneity follow easily from the fact that if $g(L, y, \alpha) = 0$ (i.e. if $e_\alpha(L) = y$), then $g(L + m, y + m, \alpha) = 0$ for $m \in \mathbb{R}$ and $g(\lambda L, \lambda y, \alpha) = 0$ for $\lambda > 0$.

For monotonicity, fix α and let $y_1 = e_\alpha(L_1)$, $y_2 = e_\alpha(L_2)$. If $L_2 \geqslant L_1$ then $g(L_2, y_1, \alpha) \geqslant g(L_1, y_1, \alpha) = g(L_2, y_2, \alpha) = 0$. Since g is decreasing in y, it must be the case that $y_2 \geqslant y_1$.

For subadditivity, again let $y_1 = e_\alpha(L_1)$, $y_2 = e_\alpha(L_2)$. We have that

$$\begin{aligned} g(L_1 + L_2, y_1 + y_2, \alpha) &= (2\alpha - 1)E((L_1 + L_2 - y_1 - y_2)^+) \\ &\quad + (1-\alpha)E(L_1 + L_2 - y_1 - y_2) \\ &= (2\alpha - 1)E((L_1 + L_2 - y_1 - y_2)^+) \\ &\quad + (1-\alpha)E(L_1 - y_1) + (1-\alpha)E(L_2 - y_2), \end{aligned}$$

and, since $(2\alpha - 1)E((L_i - y_i)^+) + (1-\alpha)E(L_i - y_i) = 0$ for $i = 1, 2$, we get

$$\begin{aligned} g(L_1 + L_2, y_1 + y_2, \alpha) = (2\alpha - 1)(&E((L_1 + L_2 - y_1 - y_2)^+) \\ &- E((L_1 - y_1)^+) - E((L_2 - y_2)^+)) \leqslant 0, \end{aligned}$$

where we have used the fact that $(x_1 + x_2)^+ \leqslant x_1^+ + x_2^+$. Since $g(L, y, \alpha)$ is decreasing in y it must be the case that $e_\alpha(L_1 + L_2) \leqslant y_1 + y_2$. □

The expectile risk measure is a law-invariant, coherent risk measure for $\alpha \geqslant 0.5$. However, it is not comonotone additive and therefore does not belong to the class of distortion risk measures described in Section 8.2.1.

If L_1 and L_2 are comonotonic random variables of the same type (so that $L_2 = kL_1 + m$ for some $m \in \mathbb{R}$ and $k > 0$), then we do have comonotone additivity (by the properties of translation invariance and positive homogeneity), but for comonotonic variables that are not of the same type we can find examples where $e_\alpha(L_1 + L_2) < e_\alpha(L_1) + e_\alpha(L_2)$ for $\alpha > 0.5$.

Notes and Comments

For distortion risk measures we use the definition of Tsukahara (2009) but restrict our attention to convex distortion functions. Using this definition, distortion risk measures are equivalent to the spectral risk measures of Kusuoka (2001), Acerbi (2002) and Tasche (2002). A parallel notion of distortion risk measures (or premium

principles) has been developed in the insurance mathematics literature, where they are also known as Wang measures (see Wang 1996), although the ideas are also found in Denneberg (1990). A good discussion of these risk measures is given by Denuit and Charpentier (2004). The characterization of distortion risk measures on atomless probability spaces as law-invariant, comonotone-additive, coherent risk measures is due to Kusuoka (2001). The representation as averages of expected shortfalls is found in Föllmer and Schied (2011).

Adam, Houkari and Laurent (2008) gives examples of distortion risk measures used in the context of portfolio optimization. Further examples can be found in Tsukahara (2009), which discusses different choices of distortion function, the properties of the resulting risk measures and statistical estimation. The concept of a distortion also plays an important role in mathematical developments within prospect theory and behavioural finance: see, for instance, Zhou (2010) and He and Zhou (2011).

Expectiles have emerged as risk measures around the recent discussion related to elicitability, a statistical notion used for the comparison of forecasts (see Gneiting 2011; Ziegel 2015). This issue will be discussed in more detail in Chapter 9. Early references on expectiles include the papers by Newey and Powell (1987), Jones (1994) and Abdous and Remillard (1995); a textbook treatment is given in Remillard (2013).

8.3 Risk Measures for Linear Portfolios

In this section we consider linear portfolios in the set

$$\mathcal{M} = \{L\colon L = m + \boldsymbol{\lambda}'\boldsymbol{X},\ m \in \mathbb{R},\ \boldsymbol{\lambda} \in \mathbb{R}^d\}, \tag{8.34}$$

where $\boldsymbol{X}$ is a fixed d-dimensional random vector of risk factors defined on some probability space $(\Omega, \mathcal{F}, P)$. The case of linear portfolios is interesting for a number of reasons. To begin with, many standard approaches to risk aggregation and capital allocation are explicitly or implicitly based on the assumption that portfolio losses have a linear relationship to underlying risk factors. Moreover, as we observed in Chapter 2, it is common to use linear approximations for losses due to market risks over short time horizons.

In Section 8.3.1 we apply the dual representation of coherent risk measures to the case of linear portfolios. We show that every coherent risk measure on the set $\mathcal{M}$ in (8.34) can be viewed as a stress test in the sense of Example 8.9. In Section 8.3.2 we consider the important case where the factor vector $\boldsymbol{X}$ has an elliptical distribution, and in Section 8.3.3 we consider briefly the case of non-elliptical distributions. As well as deriving the form of the stress test in Section 8.3.2 we also collect a number of important related results concerning risk measurement for linear portfolios of elliptically distributed risks.

8.3.1 Coherent Risk Measures as Stress Tests

Given a positive-homogeneous risk measure $\varrho\colon \mathcal{M} \to \mathbb{R}$ it is convenient to define a risk-measure function $r_\varrho(\boldsymbol{\lambda}) = \varrho(\boldsymbol{\lambda}'\boldsymbol{X})$, which can be thought of as a function of

portfolio weights. There is a one-to-one relationship between ϱ and r_ϱ given by

$$\varrho(m + \boldsymbol{\lambda}'\boldsymbol{X}) = m + r_\varrho(\boldsymbol{\lambda}).$$

Properties of ϱ therefore carry over to r_ϱ, and vice versa. We summarize the results in the following lemma.

Lemma 8.26. *Consider some translation-invariant risk measure $\varrho\colon \mathcal{M} \to \mathbb{R}$ with associated risk-measure function r_ϱ.*

(1) *ϱ is a positive-homogeneous risk measure if and only if r_ϱ is a positive-homogeneous function on $\mathbb{R}^d$, that is, $r_\varrho(t\boldsymbol{\lambda}) = t r_\varrho(\boldsymbol{\lambda})$ for all $t > 0$, $\boldsymbol{\lambda} \in \mathbb{R}^d$.*

(2) *Suppose that ϱ is positive homogeneous. Then ϱ is subadditive if and only if r_ϱ is a convex function on $\mathbb{R}^d$.*

The result follows easily from the definitions; a formal proof is therefore omitted.

The main result of this section shows that a coherent risk measure on the set of linear portfolios can be viewed as a stress test of the kind described in Example 8.9, where the scenario set is given by the set

$$S_\varrho := \{\boldsymbol{x} \in \mathbb{R}^d\colon \boldsymbol{u}'\boldsymbol{x} \leqslant r_\varrho(\boldsymbol{u}) \text{ for all } \boldsymbol{u} \in \mathbb{R}^d\}. \tag{8.35}$$

Proposition 8.27. *ϱ is a coherent risk measure on the set of linear portfolios $\mathcal{M}$ in (8.34) if and only if for every $L = m + \boldsymbol{\lambda}'\boldsymbol{X} \in \mathcal{M}$ we have the representation*

$$\varrho(L) = m + r_\varrho(\boldsymbol{\lambda}) = \sup\{m + \boldsymbol{\lambda}'\boldsymbol{x}\colon \boldsymbol{x} \in S_\varrho\}. \tag{8.36}$$

Proof. The risk measure given in (8.36) can be viewed as a generalized scenario: $\rho(L) = \sup\{E_Q(L)\colon Q \in \mathcal{Q}\}$, where $\mathcal{Q}$ is the set of Dirac measures $\{\delta_{\boldsymbol{x}}\colon \boldsymbol{x} \in S_\varrho\}$ (see Example 8.10). Such a risk measure is automatically coherent.

Conversely, suppose that ϱ is a coherent risk measure on the linear portfolio set $\mathcal{M}$. Since ϱ is translation invariant we can set $m = 0$ and consider random variables $L = \boldsymbol{\lambda}'\boldsymbol{X} \in \mathcal{M}$. Since $E_Q(\boldsymbol{\lambda}'\boldsymbol{X}) = \boldsymbol{\lambda}' E_Q(\boldsymbol{X})$, Theorem 8.11 shows that

$$\varrho(\boldsymbol{\lambda}'\boldsymbol{X}) = \sup\{\boldsymbol{\lambda}' E_Q(\boldsymbol{X})\colon Q \in \mathcal{I}^1(\Omega, \mathcal{F}),\ \alpha_{\min}(Q) = 0\}. \tag{8.37}$$

According to relation (8.14), the measures $Q \in \mathcal{I}^1(\Omega, \mathcal{F})$ for which $\alpha_{\min}(Q) = 0$ are those for which $E_Q(L) \leqslant \varrho(L)$ for all $L \in \mathcal{M}$. Hence we get

$$\begin{aligned}\{Q \in \mathcal{I}^1(\Omega, \mathcal{F})\colon \alpha_{\min}(Q) = 0\}& \\ &= \{Q \in \mathcal{I}^1(\Omega, \mathcal{F})\colon \boldsymbol{u}' E_Q(\boldsymbol{X}) \leqslant r_\varrho(\boldsymbol{u}) \text{ for all } \boldsymbol{u} \in \mathbb{R}^d\} \\ &= \{Q \in \mathcal{I}^1(\Omega, \mathcal{F})\colon E_Q(\boldsymbol{X}) \in S_\varrho\}. \end{aligned} \tag{8.38}$$

Now define the set $C := \{\boldsymbol{\mu} \in \mathbb{R}^d\colon \exists Q \in \mathcal{I}^1(\Omega, \mathcal{F}) \text{ with } \boldsymbol{\mu} = E_Q(\boldsymbol{X})\}$, and denote the closure of C by $\bar{C}$. Note that C and hence also $\bar{C}$ are convex subsets of $\mathbb{R}^d$. By combining (8.37) and (8.38) we therefore obtain

$$\varrho(\boldsymbol{\lambda}'\boldsymbol{X}) = \sup\{\boldsymbol{\lambda}'\boldsymbol{\mu}\colon \boldsymbol{\mu} \in C \cap S_\varrho\} = \sup\{\boldsymbol{\lambda}'\boldsymbol{\mu}\colon \boldsymbol{\mu} \in \bar{C} \cap S_\varrho\}.$$

If $S_\varrho \subset \bar{C}$, the last equation is equivalent to $\varrho(\boldsymbol{\lambda}'\boldsymbol{X}) = \sup\{\boldsymbol{\lambda}'\boldsymbol{\mu}\colon \boldsymbol{\mu} \in S_\varrho\}$, which is the result we require. Note that the key insight in this argument is the fact that a

probability measure on the linear portfolio space $\mathcal{M}$ can be identified by its mean vector; it is here that the special structure of $\mathcal{M}$ enters.

To verify that $S_\varrho \subset \bar{C}$, suppose to the contrary that there is some $\boldsymbol{\mu}_0 \in S_\varrho$ that does not belong to $\bar{C}$. According to Proposition A.8 (b) (the strict separation part of the separating hyperplane theorem), there would exist some $\boldsymbol{u}^* \in \mathbb{R}^d \setminus \{0\}$ such that $\boldsymbol{\mu}_0' \boldsymbol{u}^* > \sup\{\boldsymbol{\mu}' \boldsymbol{u}^* \colon \boldsymbol{\mu} \in \bar{C}\}$. It follows that

$$\begin{aligned} r_\varrho(\boldsymbol{u}^*) = \varrho((\boldsymbol{u}^*)' \boldsymbol{X}) &\leqslant \sup\{E_Q((\boldsymbol{u}^*)' \boldsymbol{X}) \colon Q \in \mathcal{I}^1(\Omega, \mathcal{F})\} \\ &= \sup\{\boldsymbol{\mu}' \boldsymbol{u}^* \colon \boldsymbol{\mu} \in \bar{C}\} < \boldsymbol{\mu}_0' \boldsymbol{u}^*. \end{aligned}$$

This contradicts the fact that $\boldsymbol{\mu}_0 \in S_\varrho$, which requires $\boldsymbol{\mu}_0' \boldsymbol{u}^* \leqslant r_\varrho(\boldsymbol{u}^*)$. □

The scenario set S_ϱ in (8.35) is an intersection of the *half-spaces* $H_u = \{\boldsymbol{x} \in \mathbb{R}^d \colon \boldsymbol{u}' \boldsymbol{x} \leqslant r_\varrho(\boldsymbol{u})\}$, so S_ϱ is a closed convex set. The precise form of S_ϱ depends on the distribution of $\boldsymbol{X}$ and on the risk measure ϱ. In the case of the quantile risk measure $\varrho = \mathrm{VaR}_\alpha$, the set S_ϱ has a probabilistic interpretation as a so-called *depth set*. Suppose that $\boldsymbol{X}$ is such that for all $\boldsymbol{u} \in \mathbb{R}^d \setminus \{0\}$ the random variable $\boldsymbol{u}' \boldsymbol{X}$ has a continuous distribution function. Then, for $H_u := \{\boldsymbol{x} \in \mathbb{R}^d \colon \boldsymbol{u}' \boldsymbol{x} \leqslant \mathrm{VaR}_\alpha(\boldsymbol{u}' \boldsymbol{X})\}$ we have that $P(\boldsymbol{u}' \boldsymbol{X} \in H_u) = \alpha$, so that the set S_{VaR_α} is the intersection of all half-spaces with probability α.

8.3.2 Elliptically Distributed Risk Factors

We have seen in Chapter 6 that an elliptical model may be a reasonable approximate model for various kinds of risk-factor data, such as stock or exchange-rate returns. The next result summarizes some key results for risk measurement on linear spaces when the underlying distribution of the risk factors is elliptical.

Theorem 8.28 (risk measurement for elliptical risk factors). *Suppose that $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ and let $\mathcal{M}$ be the space of linear portfolios (8.34). For any positive-homogeneous, translation-invariant and law-invariant risk measure ϱ on $\mathcal{M}$ the following properties hold.*

(1) *For any $L = m + \boldsymbol{\lambda}' \boldsymbol{X} \in \mathcal{M}$ we have*

$$\varrho(L) = \sqrt{\boldsymbol{\lambda}' \Sigma \boldsymbol{\lambda}} \varrho(Y) + \boldsymbol{\lambda}' \boldsymbol{\mu} + m, \tag{8.39}$$

where $Y \sim S_1(\psi)$, i.e. a univariate spherical distribution (a distribution that is symmetric around 0) with generator ψ.

(2) *If $\varrho(Y) \geqslant 0$ then ϱ is subadditive on $\mathcal{M}$. In particular, VaR_α is subadditive if $\alpha \geqslant 0.5$.*

(3) *If $\boldsymbol{X}$ has a finite-mean vector, then, for any $L = m + \boldsymbol{\lambda}' \boldsymbol{X} \in \mathcal{M}$, we have*

$$\varrho(L - E(L)) = \sqrt{\sum_{i=1}^d \sum_{j=1}^d \rho_{ij} \lambda_i \lambda_j \varrho(X_i - E(X_i)) \varrho(X_j - E(X_j))}, \tag{8.40}$$

where the ρ_{ij} are elements of the correlation matrix $\wp(\Sigma)$.

(4) *If $\varrho(Y) > 0$ and $\boldsymbol{X}$ has a finite covariance matrix, then, for every $L \in \mathcal{M}$,*

$$\varrho(L) = E(L) + k_\varrho \sqrt{\operatorname{var}(L)} \tag{8.41}$$

for some constant $k_\varrho > 0$ that depends on the risk measure.

(5) *If $\varrho(Y) > 0$ and Σ is invertible, then the scenario set S_ϱ in the stress-test representation (8.36) of ϱ is given by the ellipsoid*

$$S_\varrho = \{\boldsymbol{x} \colon (\boldsymbol{x} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \leqslant \varrho(Y)^2\}.$$

Proof. For any $L \in \mathcal{M}$ it follows from Definition 6.25 that we can write

$$L = m + \boldsymbol{\lambda}' \boldsymbol{X} \overset{\mathrm{d}}{=} \boldsymbol{\lambda}' A \boldsymbol{Y} + \boldsymbol{\lambda}' \boldsymbol{\mu} + m$$

for a spherical random vector $\boldsymbol{Y} \sim S_k(\psi)$, a matrix $A \in \mathbb{R}^{d \times k}$ satisfying $AA' = \Sigma$ and a constant vector $\boldsymbol{\mu} \in \mathbb{R}^d$. By Theorem 6.18 (3) we have

$$L \overset{\mathrm{d}}{=} \|A'\boldsymbol{\lambda}\| Y + \boldsymbol{\lambda}' \boldsymbol{\mu} + m, \tag{8.42}$$

where Y is a component of the random vector $\boldsymbol{Y}$ that has the symmetric distribution $Y \sim S_1(\psi)$. Every $L \in \mathcal{M}$ is therefore an rv of the same type, and the translation invariance and homogeneity of ϱ imply that

$$\varrho(L) = \|A'\boldsymbol{\lambda}\| \varrho(Y) + \boldsymbol{\lambda}' \boldsymbol{\mu} + m, \tag{8.43}$$

so that claim (1) follows.

For (2) set $L_1 = m_1 + \boldsymbol{\lambda}_1' \boldsymbol{X}$ and $L_2 = m_2 + \boldsymbol{\lambda}_2' \boldsymbol{X}$. Since $\|A'(\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2)\| \leqslant \|A'\boldsymbol{\lambda}_1\| + \|A'\boldsymbol{\lambda}_2\|$ and since $\varrho(Y_1) \geqslant 0$, the subadditivity of the risk measure follows easily. Since $E(L) = \boldsymbol{\lambda}' \boldsymbol{\mu} + m$ and ϱ is translation invariant, formula (8.39) implies that

$$\varrho(L - E(L)) = \left(\sum_{i=1}^{d} \sum_{j=1}^{d} \lambda_i \lambda_j \rho_{ij} \sigma_i \sigma_j \right)^{1/2} \varrho(Y), \tag{8.44}$$

where $\sigma_i = \sqrt{\Sigma_{ii}}$ for $i = 1, \ldots, d$. As a special case of (8.44) we observe that

$$\varrho(X_i - E(X_i)) = \varrho(\boldsymbol{e}_i' \boldsymbol{X} - E(\boldsymbol{e}_i' \boldsymbol{X})) = \sigma_i \varrho(Y). \tag{8.45}$$

Combining (8.44) and (8.45) yields formula (8.40) and proves part (3).

For (4) assume that $\operatorname{cov}(\boldsymbol{X}) = c\Sigma$ for some positive constant c. It follows easily from (8.44) that $\varrho(L) = E(L) + \sqrt{\operatorname{var}(L)}\varrho(Y)/\sqrt{c}$ and $k_\varrho = \varrho(Y)/\sqrt{c}$.

For (5) note that part (2) implies that the risk-measure function $r_\varrho(\boldsymbol{\lambda})$ takes the form $r_\varrho(\boldsymbol{\lambda}) = \|A'\boldsymbol{\lambda}\| \varrho(Y) + \boldsymbol{\lambda}' \boldsymbol{\mu}$, so that the set S_ϱ in (8.36) is

$$\begin{aligned} S_\varrho &= \{\boldsymbol{x} \in \mathbb{R}^d \colon \boldsymbol{u}'\boldsymbol{x} \leqslant \boldsymbol{u}'\boldsymbol{\mu} + \|A'\boldsymbol{u}\| \varrho(Y),\ \forall \boldsymbol{u} \in \mathbb{R}^d\} \\ &= \{\boldsymbol{x} \in \mathbb{R}^d \colon \boldsymbol{u}' A A^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \leqslant \|A'\boldsymbol{u}\| \varrho(Y),\ \forall \boldsymbol{u} \in \mathbb{R}^d\} \\ &= \left\{ \boldsymbol{x} \in \mathbb{R}^d \colon \boldsymbol{v}' \frac{A^{-1}(\boldsymbol{x} - \boldsymbol{\mu})}{\varrho(Y)} \leqslant \|\boldsymbol{v}\|,\ \forall \boldsymbol{v} \in \mathbb{R}^d \right\}, \end{aligned}$$

where the last line follows because $\mathbb{R}^d = \{A'\boldsymbol{u} : \boldsymbol{u} \in \mathbb{R}^d\}$. By observing that the Euclidean unit ball $\{\boldsymbol{y} \in \mathbb{R}^d : \boldsymbol{y}'\boldsymbol{y} \leqslant 1\}$ can be written as the set $\{\boldsymbol{y} \in \mathbb{R}^d : \boldsymbol{v}'\boldsymbol{y} \leqslant \|\boldsymbol{v}\|, \ \forall \boldsymbol{v} \in \mathbb{R}^d\}$, we conclude that, for $\boldsymbol{x} \in S_\varrho$, the vectors $\boldsymbol{y} = A^{-1}(\boldsymbol{x} - \boldsymbol{\mu})/\varrho(Y)$ describe the unit ball and therefore

$$S_\varrho = \{\boldsymbol{x} \in \mathbb{R}^d : (\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \leqslant \varrho(Y)^2\}.$$

□

The various parts of Theorem 8.28 have a number of important implications. Part (2) gives a special case where the VaR risk measure is subadditive and therefore coherent; we recall from Section 2.25 that this is not the case in general. Part (3) gives a useful interpretation of risk measures on $\mathcal{M}$ in terms of the aggregation of stress tests, as will be seen later in Section 8.4.

Part (4) is relevant to portfolio optimization. If we consider only the portfolio losses $L \in \mathcal{M}$ for which $E(L)$ is fixed at some level, then the portfolio weights that minimize ϱ also minimize the variance. The portfolio minimizing the risk measure ϱ is the same as the Markowitz variance-minimizing portfolio.

Part (5) shows that the scenario sets in the stress-test representation of coherent risk measures are ellipsoids when the distribution of risk-factor changes is elliptical. Moreover, for different examples of law-invariant coherent risk measures, we simply obtain ellipsoids of differing radius $\varrho(Y)$. Scenario sets of ellipsoidal form are often used in practice and this result provides a justification for this practice in the case of linear portfolios of elliptical risk factors.

8.3.3 Other Risk Factor Distributions

We now turn briefly to the application of Proposition 8.27 in situations where the risk factors do not have an elliptical distribution. The VaR risk measure is not coherent on the linear space $\mathcal{M}$ in general. Consider the simple case where we have two independent standard exponentially distributed risk factors X_1 and X_2.

Here it may happen that VaR_α is not coherent on $\mathcal{M}$ for some values of α. In such situations, (8.36) does not hold in general and we may find vectors of portfolio weights $\boldsymbol{\lambda}$ such that

$$\mathrm{VaR}_\alpha(\boldsymbol{\lambda}'\boldsymbol{X}) > \sup\{\boldsymbol{\lambda}'\boldsymbol{x} : \boldsymbol{x} \in S_\alpha\},$$

where

$$S_\alpha := S_{\mathrm{VaR}_\alpha} = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{u}'\boldsymbol{x} \leqslant \mathrm{VaR}_\alpha(\boldsymbol{u}'\boldsymbol{X}), \ \forall \boldsymbol{u} \in \mathbb{R}^d\}.$$

Such a situation is shown in Figure 8.2 (a) for two independent standard exponential risk factors. Each line bounds a half-space with probability $\alpha = 0.7$, and the intersection of these half-spaces (the empty area in the centre) is the set $S_{0.7}$. Some lines are not supporting hyperplanes of $S_{0.7}$, meaning they do not touch it; an example is the bold diagonal line in the upper right corner of the picture. In such situations we can construct vectors of portfolio weights $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ such that $\mathrm{VaR}_\alpha((\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2)'\boldsymbol{X}) > \mathrm{VaR}_\alpha(\boldsymbol{\lambda}_1'\boldsymbol{X}) + \mathrm{VaR}_\alpha(\boldsymbol{\lambda}_2'\boldsymbol{X})$. In fact, for $\alpha = 0.7$ we simply have $\mathrm{VaR}_\alpha(X_1 + X_2) > \mathrm{VaR}_\alpha(X_1) + \mathrm{VaR}_\alpha(X_2)$, as may

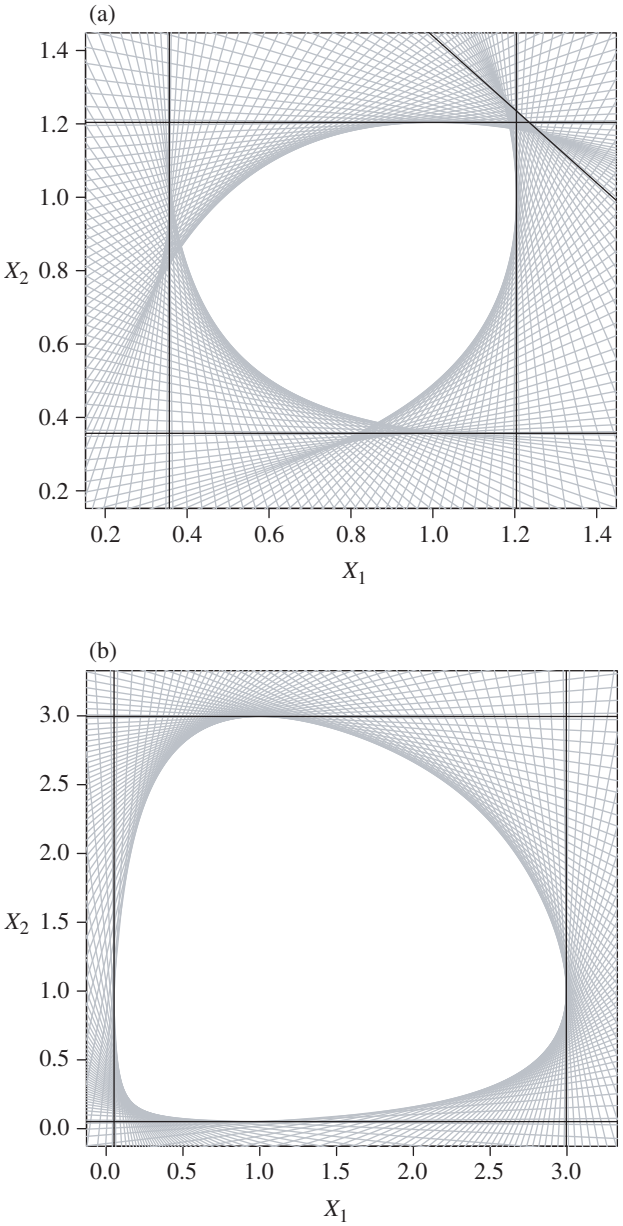

Figure 8.2. Illustration of scenario sets S_α when X_1 and X_2 are two independent standard exponential variates and the risk measure is VaR_α. (a) The case $\alpha = 0.7$, where VaR_α is not coherent on $\mathcal{M}$. (b) The case $\alpha = 0.95$, where VaR_α is coherent.

be deduced from the picture by noting that the black vertical line on the right is $x_1 = \mathrm{VaR}_\alpha(X_1)$, the upper black horizontal line is $x_2 = \mathrm{VaR}_\alpha(X_2)$, and the bold diagonal line is $x_1 + x_2 = \mathrm{VaR}_\alpha(X_1 + X_2)$. This agrees with our observation in Example 7.30.

For the value $\alpha = 0.95$, the depth set is shown in Figure 8.2 (b). In this case the depth set has a smooth boundary and there are supporting hyperplanes bounding half-spaces with probability α in every direction. We can apply Proposition 8.27 to conclude that VaR_α is a coherent risk measure on $\mathcal{M}$ for $\alpha = 0.95$.

These issues do not arise for the expected shortfall risk measure or any other coherent risk measure. The scenario set S_ϱ for these risk measures would have supporting hyperplanes with equation $\boldsymbol{u}'\boldsymbol{x} = r_\varrho(\boldsymbol{u})$ for every direction $\boldsymbol{u} \neq \mathbf{0}$.

Notes and Comments

The presentation of the relationship between coherent risk measures and stress tests on linear portfolio spaces is based on McNeil and Smith (2012). Our definition of a stress test coincides with the concept of the *maximum loss* risk measure introduced by Studer (1997, 1999), who also considers ellipsoidal scenario sets. Breuer et al. (2009) describe the problem of finding scenarios that are "plausible, severe and useful" and propose a number of refinements to Studer's approach based on ellipsoidal sets.

The depth sets obtained when the risk measure is VaR have an interesting history in statistics and have been studied by Massé and Theodorescu (1994) and Rousseeuw and Ruts (1999), among others. The concept has its origins in an empirical concept of *data depth* introduced by Tukey (1975) as well as in theoretical work on multivariate analogues of the quantile function by Eddy (1984) and Nolan (1992).

The treatment of the implications of elliptical distributions for risk measurement follows Embrechts, McNeil and Straumann (2002). Chapter 9 of Hult et al. (2012) contains an interesting discussion of elliptical distributions in risk management. There is an extensive body of economic theory related to the use of elliptical distributions in finance. The papers by Owen and Rabinovitch (1983), Chamberlain (1983) and Berk (1997) provide an entry to the area. Landsman and Valdez (2003) discuss the explicit calculation of the quantity $E(L \mid L > q_\alpha(L))$ for portfolios of elliptically distributed risks. This coincides with expected shortfall for continuous loss distributions (see Proposition 2.13).

8.4 Risk Aggregation

The need to aggregate risk can arise in a number of situations. Suppose that capital amounts $\mathrm{EC}_1, \ldots, \mathrm{EC}_d$ (EC stands for economic capital) have been computed for each of d subsidiaries or business lines making up an enterprise and a method for computing the aggregate capital for the whole enterprise is required. Or, in a similar vein, suppose that capital amounts $\mathrm{EC}_1, \ldots, \mathrm{EC}_d$ have been computed for d different asset classes on the balance sheet of an enterprise and a method is required to compute the overall capital required to back all assets.

A *risk-aggregation rule* is a mapping

$$f : \mathbb{R}^d \to \mathbb{R}, \quad f(\mathrm{EC}_1, \ldots, \mathrm{EC}_d) = \mathrm{EC},$$

which takes as input the individual capital amounts and gives as output the aggregate capital EC. Examples of commonly used rules are *simple summation*

$$\mathrm{EC} = \mathrm{EC}_1 + \cdots + \mathrm{EC}_d \tag{8.46}$$

and *correlation adjusted summation*

$$\mathrm{EC} = \sqrt{\sum_{i=1}^{d}\sum_{j=1}^{d}\rho_{ij}\,\mathrm{EC}_i\,\mathrm{EC}_j}, \tag{8.47}$$

where the ρ_{ij} are a set of parameters satisfying $0 \leqslant \rho_{ij} \leqslant 1$, which are usually referred to as correlations. Of course, (8.46) is a special case of (8.47) when $\rho_{ij} = 1$, $\forall i, j$, and the aggregate capital given by (8.46) is an upper bound for the aggregate capital given by (8.47).

The application of rules like (8.46) and (8.47) in the absence of any deeper consideration of multivariate models for the enterprise or the use of risk measures is referred to as *rules-based aggregation*. By contrast, the use of aggregation rules that can be theoretically justified by relating capital amounts to risk measures and multivariate models for losses is referred to as *principles-based aggregation*. In the following sections we give examples of the latter approach.

8.4.1 Aggregation Based on Loss Distributions

In this section we suppose that the overall loss of the enterprise over a fixed time interval is given by $L_1 + \cdots + L_d$, where $L_1, \ldots, L_d$ are the losses arising from sub-units of the enterprise (such as business units or asset classes on the balance sheet). We consider a translation-invariant risk measure ϱ and define a mean-adjusted version of the risk measure by

$$\varrho^{\mathrm{mean}}(L) = \varrho(L - E(L)) = \varrho(L) - E(L). \tag{8.48}$$

ϱ^{mean} can be thought of as the capital required to cover unexpected losses.

The capital requirements for the sub-units are given by $\mathrm{EC}_i = \varrho^{\mathrm{mean}}(L_i)$ for $i = 1, \ldots, d$, and the aggregate capital should be given by $\mathrm{EC} = \varrho^{\mathrm{mean}}(L_1 + \cdots + L_d)$. We require an aggregation rule f such that $\mathrm{EC} = f(\mathrm{EC}_1, \ldots, \mathrm{EC}_d)$.

As an example, suppose that we take the risk measure $\varrho(L) = k\,\mathrm{sd}(L) + E(L)$, where sd denotes the standard deviation, k is some positive constant, and second moments of the loss distributions are assumed to be finite. Regardless of the underlying distribution of $L_1, \ldots, L_d$ the standard deviation satisfies

$$\mathrm{sd}(L) = \sqrt{\sum_{i=1}^{d}\sum_{j=1}^{d}\rho_{ij}\,\mathrm{sd}(L_i)\,\mathrm{sd}(L_j)}, \tag{8.49}$$

where the ρ_{ij} are the elements of the correlation matrix of $(L_1, \ldots, L_d)$ and the aggregation rule (8.47) therefore follows in this case.

When the losses are elements of the linear space $\mathcal{M}$ in (8.34) and the distribution of the underlying risk-factor changes $\boldsymbol{X}$ is elliptical with finite covariance matrix, then (8.49) and Theorem 8.28 (4) imply that (8.47) is justified for any positive-homogeneous, translation-invariant and law-invariant risk measures. We now give a more elegant proof of this fact that does not require us to assume finite second moments of $\boldsymbol{X}$.

Proposition 8.29. *Let $\boldsymbol{X} \sim E_k(\boldsymbol{\mu}, \Sigma, \psi)$ with $E(\boldsymbol{X}) = \boldsymbol{\mu}$. Let $\mathcal{M} = \{L\colon L = m + \boldsymbol{\lambda}'\boldsymbol{X},\ \boldsymbol{\lambda} \in \mathbb{R}^k,\ m \in \mathbb{R}\}$ be the space of linear portfolios and let ϱ be a positive-homogeneous, translation-invariant and law-invariant risk measure on $\mathcal{M}$. For $L_1, \ldots, L_d \in \mathcal{M}$ let $\mathrm{EC}_i = \varrho^{\mathrm{mean}}(L_i)$ and $\mathrm{EC} = \varrho^{\mathrm{mean}}(L_1 + \cdots + L_d)$. The capital amounts $\mathrm{EC}, \mathrm{EC}_1, \ldots, \mathrm{EC}_d$ then satisfy the aggregation rule (8.47), where the ρ_{ij} are elements of the correlation matrix $P = \wp(\tilde{\Sigma})$ and where $\tilde{\Sigma}$ is the dispersion matrix of the (elliptically distributed) random vector $(L_1, \ldots, L_d)$.*

Proof. Let $L_i = \boldsymbol{\lambda}_i'\boldsymbol{X} + m_i$ for $i = 1, \ldots, d$. It follows from Theorem 8.28 (1) that

$$\mathrm{EC}_i = \varrho(L_i) - E(L_i) = \sqrt{\boldsymbol{\lambda}_i'\Sigma\boldsymbol{\lambda}_i}\varrho(Y),$$

where $Y \sim S_1(\psi)$, and that

$$\begin{aligned}\mathrm{EC} &= \sqrt{(\boldsymbol{\lambda}_1 + \cdots + \boldsymbol{\lambda}_d)'\Sigma(\boldsymbol{\lambda}_1 + \cdots + \boldsymbol{\lambda}_d)}\varrho(Y)\\ &= \sqrt{\sum_{i=1}^{d}\sum_{j=1}^{d}\boldsymbol{\lambda}_i'\Sigma\boldsymbol{\lambda}_j}\varrho(Y)\\ &= \sqrt{\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\boldsymbol{\lambda}_i'\Sigma\boldsymbol{\lambda}_j}{\sqrt{(\boldsymbol{\lambda}_i'\Sigma\boldsymbol{\lambda}_i)(\boldsymbol{\lambda}_j'\Sigma\boldsymbol{\lambda}_j)}}\mathrm{EC}_i\,\mathrm{EC}_j}.\end{aligned}$$

The dispersion matrix $\tilde{\Sigma}$ of $(L_1, \ldots, L_d)$ is now given by $\tilde{\Sigma} = \Lambda\Sigma\Lambda'$, where $\Lambda \in \mathbb{R}^{d\times k}$ is the matrix with rows given by the vectors $\boldsymbol{\lambda}_i$. The correlation matrix $P = \wp(\tilde{\Sigma})$ clearly has elements given by

$$\boldsymbol{\lambda}_i'\Sigma\boldsymbol{\lambda}_j/\sqrt{(\boldsymbol{\lambda}_i'\Sigma\boldsymbol{\lambda}_i)(\boldsymbol{\lambda}_j'\Sigma\boldsymbol{\lambda}_j)}$$

and the result follows. □

Proposition 8.29 implies that the aggregation rule (8.47) can be justified when we work with the mean-adjusted value-at-risk or expected shortfall risk measures if we are prepared to make the strong assumption that the underlying multivariate loss distribution is elliptical.

Clearly the elliptical assumption is unlikely to hold in practice, so the theoretical support that allows us to view (8.47) as a principles-based approach will generally be lacking. However, even if the formula is used as a pragmatic rule, there are also practical problems with the approach.

- The formula requires the specification of pairwise correlations between the losses $L_1, \ldots, L_d$. It will be difficult to obtain estimates of these correlations, since empirical data is generally available at the level of the underlying risk factors rather than the level of resulting portfolio losses.
- If, instead, the parameters are chosen by *expert judgement*, then there are compatibility requirements. In order to make sense, the ρ_{ij} must form the elements of a positive-semidefinite correlation matrix. When a correlation matrix is pieced together from pairwise estimates it is quite easy to violate this condition, and the risk of this happening increases with dimension.

- If $L_1, \ldots, L_d$ are believed to have a non-elliptical distribution, then the limited range of attainable correlations for each pair (L_i, L_j), as discussed in connection with Fallacy 2 in Section 7.2.2, is also a relevant constraint
- The use of (8.47) offers no obvious way to incorporate tail dependence between the losses into the calculation of aggregate capital.

It might be supposed that use of the summation formula (8.46) would avoid these issues with correlation and yield a conservative upper bound for aggregate capital for any possible underlying distribution of $(L_1, \ldots, L_d)$. While this is true if the risk measure ϱ is a coherent risk measure, it is not true in general if ϱ is a non-subadditive risk measure, such as VaR. This is an example of Fallacy 3 in Section 7.2.2.

It is possible that the underlying multivariate loss model is one where, for some value of α, $\mathrm{VaR}_\alpha(L_1+\cdots+L_d) > \mathrm{VaR}_\alpha(L_1)+\cdots+\mathrm{VaR}_\alpha(L_d)$. In this case, if we set $\mathrm{EC}_i = \mathrm{VaR}_\alpha(L_i) - E(L_i)$ and take the sum $\mathrm{EC}_1 + \cdots + \mathrm{EC}_1$, this will underestimate the actual required capital $\mathrm{EC} = \mathrm{VaR}_\alpha(L) - E(L)$, where $L = L_1 + \cdots + L_d$.

In Section 8.4.4 we examine the problem of putting upper and lower bounds on aggregate capital when marginal distributions are known and marginal capital requirements are determined by the value-at-risk measure.

8.4.2 Aggregation Based on Stressing Risk Factors

Another situation where aggregation rules of the form (8.47) are used in practice is in the aggregation of capital contributions computed by stressing individual risk factors. An example of such an application is the standard formula approach to Solvency II (see, for example, CEIOPS 2006). Capital amounts $\mathrm{EC}_1, \ldots, \mathrm{EC}_d$ are computed by examining the effects on the balance sheet of extreme changes in a number of key risk factors, and (8.47) is used to compute an overall capital figure that takes into account the dependence of the risk factors.

To understand when the use of (8.47) may be considered to be a principles-based approach to aggregation, suppose we write $\boldsymbol{x} = \boldsymbol{X}(\omega)$ for a scenario defined in terms of changes in fundamental risk factors and $L(\boldsymbol{x})$ for the corresponding loss. We assume that $L(\boldsymbol{x})$ is a known function and, for simplicity, that it is increasing in each component of $\boldsymbol{x}$. Following common practice, the d risk factors are stressed one at a time by predetermined amounts $k_1, \ldots, k_d$. Capital contributions for each risk factor are set by computing

$$\mathrm{EC}_i = L(k_i \boldsymbol{e}_i) - L(E(X_i)\boldsymbol{e}_i), \tag{8.50}$$

where $\boldsymbol{e}_i$ denotes the ith unit vector and where $k_i > E(X_i)$ so that $\mathrm{EC}_i > 0$. The value EC_i can be thought of as the loss incurred by stressing risk factor i by an amount k_i relative to the impact of stressing it by its expected change, while all other risk factors are held constant. One possibility is that the size of the stress event is set at the level of the α-quantile of the distribution of X_i, so that $k_i = q_\alpha(X_i)$ for α close to 1. We now prove a simple result that justifies the use of the aggregation rule (8.47) to combine the contributions $\mathrm{EC}_1, \ldots, \mathrm{EC}_d$ defined in (8.50) into an aggregate capital EC.

Proposition 8.30. *Let $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$, with $E(\boldsymbol{X}) = \boldsymbol{\mu}$. Let $\mathcal{M}$ be the space of linear portfolios (8.34) and let ϱ be a positive-homogeneous, translation-invariant and law-invariant risk measure on $\mathcal{M}$. Then, for any $L = L(\boldsymbol{X}) = m + \boldsymbol{\lambda}'\boldsymbol{X} \in \mathcal{M}$ we have*

$$\varrho(L - E(L)) = \sqrt{\sum_{i=1}^{d}\sum_{j=1}^{d} \rho_{ij}\,\mathrm{EC}_i\,\mathrm{EC}_j}, \tag{8.51}$$

where $\mathrm{EC}_i = L(\varrho(X_i)\boldsymbol{e}_i) - L(E(X_i)\boldsymbol{e}_i)$ and ρ_{ij} is an element of $\wp(\Sigma)$.

Proof. We observe that $\mathrm{EC}_i = \lambda_i \varrho(X_i) - \lambda_i E(X_i) = \lambda_i \varrho(X_i - E(X_i))$, and (8.51) follows by application of Theorem 8.28 (3). □

Proposition 8.30 shows that, under a strong set of assumptions, we can aggregate the effects of single-risk-factor stresses to obtain an aggregate capital requirement that corresponds to application of any positive-homogeneous, translation-invariant and law-invariant risk measure to the distribution of the unexpected loss; this would apply to VaR, expected shortfall or one of the distortion risk measures of Section 8.2.1. It is this idea that underscores the use of (8.47) in Solvency II. However, the key assumptions are, once again, the linearity of losses in the risk-factor changes and the elliptical distribution of risk-factor changes, both of which are simplistic in real-world applications.

We can of course regard the use of (8.47) as a pragmatic, rules-based approach. The correlation parameters are defined at the level of the risk factors and, for typical market-risk factors such as returns on prices or rates, the data may be available to permit estimation of these parameters. For other risk factors, such as mortality and policy lapse rates in Solvency II applications, parameters may be set by expert judgement and the same issues mentioned in Section 8.4.1 apply. In particular, the matrix with components ρ_{ij} must be positive definite in order for the procedure to make any kind of sense.

The summation rule may once again appear to be a conservative rule that avoids the problems related to estimating and setting correlations. However, it should be noted that in the presence of non-linear relationships between losses and risk factors, there can be complex interactions between risk factors that would require even higher capital than indicated by the sum of losses due to single-risk-factor stresses (see Notes and Comments).

8.4.3 Modular versus Fully Integrated Aggregation Approaches

The approaches discussed in Sections 8.4.1 and 8.4.2 can be described as *modular* approaches to risk capital. The risk is computed in modules or *silos* and then aggregated. In Section 8.4.1 the modules are defined in terms of business units or asset classes; in Section 8.4.2 the modules are defined in terms of individual risk factors. The former approach is arguably more natural because the losses across asset classes and business units are additive and it is possible to remove risks from the enterprise by selling parts of the business. The risks due to fundamental underlying risk factors are more pervasive and may manifest themselves in different parts of the balance

sheet; typically, their effects can be non-linear and they can only be reduced by hedging.

Regardless of the nature of the underlying silos the aggregation approaches we have described involve the specification of correlations and the use of (8.47) or its special case (8.46). We have observed that there are practical problems associated with choosing correlations, and in Chapter 7 we have argued that correlation gives only a partial description of a multivariate distribution and that copulas are a better approach to multivariate dependence modelling. It is natural to consider using copulas in aggregation.

In the set-up of Section 8.4.1, where the total loss is given by $L = L_1 + \cdots + L_d$ and the L_i are losses due to business units, suppose that we know, or can accurately estimate, the marginal distributions $F_1, \ldots, F_d$ for each of the modules. This is a necessary prerequisite for computing the marginal capital requirements $\mathrm{EC}_i = \varrho(L_i) - E(L_i)$. Instead of aggregating these marginal capital figures with correlation, we could attempt to choose a suitable copula C and build a multivariate loss distribution $F(\boldsymbol{x}) = C(F_1(x_1), \ldots, F_d(x_d))$ using the converse of Sklar's Theorem (7.3). This is referred to as the *margins-plus-copula* approach. Computation of aggregate capital would typically proceed by generating large numbers of multivariate losses from F, summing them to obtain simulated overall losses L, and then applying empirical quantile or shortfall estimation techniques.

The problem with this approach is the specification of C. Multivariate loss data from the business units may be sparse or non-existent, and expert judgment may have to be employed. This might involve deciding whether the copula should have a degree of tail dependence, taking a view on plausible levels of rank correlation between the pairs (L_i, L_j) and then using the copula calibration methods based on rank correlation described in Section 7.5.1. Clearly, this approach has as many, if not more, problems than choosing a correlation matrix to use in (8.47). It remains a modular approach in which we start with models for the individual L_i and add dependence assumptions as an overlay. In Section 8.4.4 we will address the issue of *dependence uncertainty* in such a margins-plus-copula approach; in particular, we will quantify the "best-to-worst" gaps in VaR and ES estimation if only the marginal dfs of the losses are known.

A more appealing approach, which we describe as a *fully integrated* approach, is to build multivariate models for the changes in underlying risk factors $\boldsymbol{X} = (X_1, \ldots, X_k)'$ and for the functionals $g_i : \mathbb{R}^k \mapsto \mathbb{R}$ that give the losses $L_i = g_i(\boldsymbol{X})$, $i = 1, \ldots, d$, for the different portfolios, desks or business units that make up the enterprise. It is generally easier to build multivariate models for underlying risk factors because more data exist at the level of the risk factors. The models for $\boldsymbol{X}$ may range in sophistication from margins-plus-copula distributional models to more dynamic, financial econometric models. They are often referred to as *economic scenario generators*. In the fully integrated approach, aggregate capital is derived by applying risk measures to the distribution of $L = g_1(\boldsymbol{X}) + \cdots + g_d(\boldsymbol{X})$, and the losses in business units L_i and L_j are implicitly dependent through their mutual dependence on $\boldsymbol{X}$.

8.4.4 Risk Aggregation and Fréchet Problems

In the margins-plus-copula approach to risk aggregation described in Section 8.4.3 a two-step procedure for the construction of a model for the total loss $L = L_1 + \cdots + L_d$ is followed.

(1) Find appropriate models (dfs) $F_1, \ldots, F_d$ for the marginal risks $L_1, \ldots, L_d$. These can be obtained by statistical fitting to historical data, or postulated a priori in a stress-testing exercise.

(2) Choose a suitable copula C resulting in a joint model $C(F_1, \ldots, F_d)$ for the random vector $\boldsymbol{L} = (L_1, \ldots, L_d)'$ from which the df for the total portfolio loss L can be derived.

Based on steps (1) and (2), any law-invariant risk measure $\varrho(L)$ can, in principle, be calculated. The examples we will concentrate on in this section are $\varrho = \mathrm{VaR}_\alpha$ and $\varrho = \mathrm{ES}_\alpha$. Note that there is nothing special about the sum structure of the portfolio L; more general portfolios (or financial positions) $L = \Psi(L_1, \ldots, L_d)$ for suitable functions $\Psi : \mathbb{R}^d \to \mathbb{R}$ could also be considered.

As mentioned in Section 8.4.3, there is a lot of model uncertainty surrounding the choice of the appropriate copula in step (2). In this section we will therefore drop step (2) while retaining step (1). Clearly, this means that quantities such as $\mathrm{VaR}_\alpha(L) = \mathrm{VaR}_\alpha(L_1 + \cdots + L_d)$ can no longer be computed precisely due to the lack of a fully specified model for the vector $\boldsymbol{L}$ and hence the aggregate loss L.

Instead we will try to find bounds for $\mathrm{VaR}_\alpha(L)$ given *only* the marginal information from step (1). Problems of this type are known as *Fréchet problems* in the literature. In this section we refer to the situation where only marginal information is available as *dependence uncertainty*.

In order to derive bounds we introduce the class of rvs

$$\begin{aligned}\mathscr{S}_d &= \mathscr{S}_d(F_1, \ldots, F_d) \\ &= \left\{ L = \sum_{i=1}^{d} L_i : L_1, \ldots, L_d \text{ rvs with } L_i \sim F_i,\ i = 1, \ldots, d \right\}.\end{aligned}$$

Clearly, every element of $\mathscr{S}_d$ is a feasible risk position satisfying step (1). The problem of finding VaR bounds under dependence uncertainty now reduces to finding

$$\overline{\mathrm{VaR}}_\alpha(\mathscr{S}_d) = \sup\{\mathrm{VaR}_\alpha(L) : L \in \mathscr{S}_d(F_1, \ldots, F_d)\}$$

and

$$\underline{\mathrm{VaR}}_\alpha(\mathscr{S}_d) = \inf\{\mathrm{VaR}_\alpha(L) : L \in \mathscr{S}_d(F_1, \ldots, F_d)\}.$$

We will use similar notation if VaR_α is replaced by another risk measure ϱ; for instance, we will write $\overline{\mathrm{ES}}_\alpha$ and $\underline{\mathrm{ES}}_\alpha$. In our main case of interest, when $L = L_1 + \cdots + L_d$ with the L_i variables as in step (1), we will often write $\varrho(\mathscr{S}_d) = \varrho(L)$ and use similar notation for the corresponding upper and lower bounds. For

expected shortfall, which is a coherent and comonotone-additive risk measure (see Definition 8.19), we have

$$\overline{\mathrm{ES}}_\alpha(L) = \sum_{i=1}^{d} \mathrm{ES}_\alpha(L_i),$$

and we see that the upper bound is achieved under comonotonicity. We often refer to $\underline{\varrho}$ as the *best* and $\overline{\varrho}$ as the *worst* ϱ; this interpretation depends, of course, on the context.

The calculation of $\overline{\mathrm{VaR}}_\alpha(\mathcal{S}_d)$, $\underline{\mathrm{VaR}}_\alpha(\mathcal{S}_d)$ and $\underline{\mathrm{ES}}_\alpha(\mathcal{S}_d)$ is difficult in general. We will review some of the main results without proof; further references on this very active research area can be found in Notes and Comments. The available results very much depend on the dimension ($d = 2$ versus $d > 2$) and whether the portfolio is homogeneous ($F_1 = \cdots = F_d$) or not. We begin with a result for the case $d = 2$.

Proposition 8.31 (VaR, $d = 2$). *Under the set-up above, $\forall \alpha \in (0, 1)$,*

$$\overline{\mathrm{VaR}}_\alpha(\mathcal{S}_2) = \inf_{x \in [0, 1-\alpha]} \{F_1^{-1}(\alpha + x) + F_2^{-1}(1 - x)\}$$

and

$$\underline{\mathrm{VaR}}_\alpha(\mathcal{S}_2) = \inf_{x \in [0, \alpha]} \{F_1^{-1}(x) + F_2^{-1}(\alpha - x)\}.$$

Proof. See Makarov (1981) and Rüschendorf (1982). □

From the above proposition we already see that the optimal *couplings*—the dependence structures achieving the VaR bounds—combine large outcomes in one risk with small outcomes in the other. Next we give VaR bounds for higher dimensions, assuming a homogeneous portfolio.

Proposition 8.32 (VaR, $d \geqslant 2$, homogeneous case). *Suppose that $F := F_1 = \cdots = F_d$ and that for some $b \in \mathbb{R}$ the density function f of F (assumed to exist) is decreasing on $[b, \infty)$. Then, for $\alpha \in [F(b), 1)$ and $X \sim F$,*

$$\overline{\mathrm{VaR}}_\alpha(\mathcal{S}_d) = d\,E(X \mid X \in [F^{-1}(\alpha + (d-1)c, F^{-1}(1-c)]), \tag{8.52}$$

where c is the smallest number in $[0, (1-\alpha)/d]$ such that

$$\int_{\alpha+(d-1)c}^{1-c} F^{-1}(t)\,\mathrm{d}t \geqslant \frac{1-\alpha-dc}{d}((d-1)F^{-1}(\alpha+(d-1)c) + F^{-1}(1-c)).$$

If the density f of F is decreasing on its support, then for $\alpha \in (0, 1)$ and $X \sim F$,

$$\underline{\mathrm{VaR}}_\alpha(\mathcal{S}_d) = \max\{(d-1)F^{-1}(0) + F^{-1}(\alpha), d\,E(X \mid X \leqslant F^{-1}(\alpha))\}. \tag{8.53}$$

Proof. For the proof of (8.52) see Wang, Peng and Yang (2013). The case (8.53) follows by symmetry arguments (see Bernard, Jiang and Wang 2014). □

Remark 8.33. First of all note the extra condition on the density f of F that is needed to obtain the sharp bound (8.53) for $\underline{\mathrm{VaR}}_\alpha(\mathcal{S}_d)$: we need f to be decreasing on its *full* support rather than only on a certain tail region, which is sufficient for (8.52). As a consequence, both (8.52) and (8.53) can be applied, for instance, to the case where F is Pareto, but for lognormal rvs only (8.52) applies.

Though the results (8.52) and (8.53) look rather involved, they exhibit an interesting structure. As in the case where $d = 2$, the extremal couplings combine large and small values of the underlying df F. More importantly, if $c = 0$, then (8.52) reduces to

$$\overline{\mathrm{VaR}}_\alpha(\mathcal{S}_d) = \overline{\mathrm{ES}}_\alpha(\mathcal{S}_d). \tag{8.54}$$

The extremal coupling for VaR is rather special and differs from the extremal coupling for ES, which is of course comonotonicity. The condition $c = 0$ corresponds to the crucial notion of d-mixability (see Definition 8.35).

The observation (8.54) is relevant to a discussion of the pros and cons of value-at-risk versus expected shortfall and the regulatory debate surrounding these risk measures. Although the upper bounds coincide in the case $c = 0$, it is much easier to compute $\overline{\mathrm{ES}}_\alpha(\mathcal{S}_d)$ due to the comonotone additivity of expected shortfall (see Embrechts et al. (2014) and Notes and Comments).

Similar to Proposition 8.32, a sharp bound for the best ES case for a homogeneous portfolio can be given, and this also requires a strong monotonicity condition for the underlying density. Here the *lower expected shortfall* risk measure LES_α enters; for $\alpha \in (0, 1)$ this is defined to be

$$\mathrm{LES}_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_u(X)\,\mathrm{d}u = -\,\mathrm{ES}_{1-\alpha}(-X).$$

Proposition 8.34 (ES, $d \geqslant 2$, homogeneous case). *Suppose that $F = F_1 = \cdots = F_d$, that F has a finite first moment and that the density function of F (which is assumed to exist) is decreasing on its support. Then, for $\alpha \in [1 - dc, 1)$, $\beta = (1 - \alpha)/d$ and $X \sim F$,*

$$\begin{aligned}\underline{\mathrm{ES}}_\alpha(\mathcal{S}_d) &= \frac{1}{\beta}\int_0^\beta ((d-1)F^{-1}((d-1)t) + F^{-1}(1-t))\,\mathrm{d}t \\ &= (d-1)^2\,\mathrm{LES}_{(d-1)\beta}(X) + \mathrm{ES}_{1-\beta}(X),\end{aligned} \tag{8.55}$$

where c is the smallest number in $[0, 1/d]$ such that

$$\int_{(d-1)c}^{1-c} F^{-1}(d)\,\mathrm{d}t \geqslant \frac{1-dc}{d}((d-1)F^{-1}((d-1)c) + F^{-1}(1-c)).$$

Proof. See Bernard, Jiang and Wang (2014). □

An important tool in proofs of these results is a general concept of multivariate *negative dependence* known as *mixability*, which is introduced next.

Definition 8.35. A df F on $\mathbb{R}$ is called *d-completely mixable* (d-CM) if there exist d rvs $X_1, \dots, X_d \sim F$ such that for some $k \in \mathbb{R}$,

$$P(X_1 + \cdots + X_d = dk) = 1. \tag{8.56}$$

F is *completely mixable* if F is d-CM for all $d \geqslant 2$. The dfs $F_1, \dots, F_d$ on $\mathbb{R}$ are called *jointly mixable* if there exist d rvs $X_i \sim F_i$, $i = 1, \dots, d$, such that for some $c \in \mathbb{R}$,

$$P(X_1 + \cdots + X_d = c) = 1.$$

Clearly, if F has finite-mean μ, we must have $k = \mu$ in (8.56). Complete mixability is a concept of strong negative dependence. It is indeed this dependence structure that yields the extremal couplings in Proposition 8.32. The above definition of d-complete mixability and its link to dependence-uncertainty problems can be found in Wang and Wang (2011). Examples of completely mixable dfs are the normal, Student t, Cauchy and uniform distributions. In Rüschendorf and Uckelmann (2002) it was shown that any continuous distribution function with a symmetric and unimodal density is d-completely mixable for any $d \geqslant 2$. See Notes and Comments for a historical perspective and further references.

In contrast to the above analytic results for the homogeneous case, very little is known for non-homogeneous portfolios, i.e. for portfolios where the condition $F_1 = \cdots = F_d$ does not hold. In general, however, there is a fast and efficient numerical procedure for solving dependence-uncertainty problems that is called the *rearrangement algorithm* (RA) (see Embrechts, Puccetti and Rüschendorf 2013). The RA was originally worked out for the calculation of best/worst VaR bounds; it can be generalized to other risk measures like expected shortfall. Mathematically, the RA is based on the above idea of mixability. For instance, for the calculation of $\overline{\mathrm{VaR}}_\alpha(L)$, one discretizes the $(1-\alpha)100\%$ upper tail of the underlying factor dfs $F_1, \dots, F_d$, using $N = 100\,000$ bins, say. For the dimension d, values around and above 1000, say, can easily be handled by the RA. It can similarly be used for the calculation of $\underline{\mathrm{VaR}}_\alpha(L)$ and $\underline{\mathrm{ES}}_\alpha(L)$ in both the homogeneous and non-homogeneous cases.

With the results discussed above, including the RA, we can now calculate several quantities related to diversification, (non-)coherence, and model and dependence uncertainty. We restrict our attention to the additive portfolio $L = L_1 + \cdots + L_d$ under the set-up in step (1). It will be useful to consider several functions $\mathcal{X}\colon \mathbb{R}^2 \to \mathbb{R}$ that compare risk measures under different dependence assumptions. Examples encountered in the literature are the following.

Super/subadditivity indices. $a = \mathrm{VaR}_\alpha(L)$, $b = \mathrm{VaR}_\alpha^+(L)$, and $\mathcal{X}_1(a,b) = a/b$, $\mathcal{X}_2(a,b) = 1 - (a/b)$, $\mathcal{X}_3(a,b) = b - a$. $\mathrm{VaR}_\alpha^+(L)$ denotes the comonotonic case, i.e. $\mathrm{VaR}_\alpha^+(L) = \sum_{i=1}^d \mathrm{VaR}_\alpha(L_i)$.

Worst superadditivity ratio. $a = \overline{\mathrm{VaR}}_\alpha(L)$, $b = \mathrm{VaR}_\alpha^+(L)$, and $\mathcal{X}_4(a,b) = a/b$; the case is similar for the *best superadditivity ratio*, replacing $\overline{\mathrm{VaR}}_\alpha(L)$ by $\underline{\mathrm{VaR}}_\alpha(L)$.

Dependence-uncertainty spread. Either $a = \underline{\mathrm{VaR}}_\alpha(L)$, $b = \overline{\mathrm{VaR}}_\alpha(L)$ or $a = \underline{\mathrm{ES}}_\alpha(L)$, $b = \overline{\mathrm{ES}}_\alpha(L)$ and $\mathcal{X}_5(a,b) = b - a$. A further interesting measure compares the VaR dependence-uncertainty spread with the ES dependence-uncertainty spread.

Best/worst (VaR, ES) ratios. Either $a = \underline{\mathrm{VaR}}_\alpha(L), b = \underline{\mathrm{ES}}_\alpha(L)$ or $a = \overline{\mathrm{VaR}}_\alpha(L)$, $b = \overline{\mathrm{ES}}_\alpha(L)$ and $\mathcal{X}_6(a,b) = b/a$, say.

As we already observed in (8.54), whenever $c = 0$ in Proposition 8.32 we have that $\overline{\mathrm{VaR}}_\alpha(\mathcal{I}_d) = \overline{\mathrm{ES}}_\alpha(\mathcal{I}_d)$. The following result extends this observation in an asymptotic way.

Proposition 8.36 (asymptotic equivalence of $\overline{\mathrm{ES}}$ and $\overline{\mathrm{VaR}}$). *Suppose that $L_i \sim F_i$, $i \geqslant 1$, and that*

(i) for some $k > 1$, $E(|L_i - E(L_i)|^k)$ is uniformly bounded, and

(ii) for some $\alpha \in (0,1)$,

$$\liminf_{d\to\infty} \frac{1}{d} \sum_{i=1}^{d} \mathrm{ES}_\alpha(L_i) > 0.$$

Then, as $d \to \infty$,

$$\frac{\overline{\mathrm{ES}}_\alpha(\mathcal{I}_d)}{\overline{\mathrm{VaR}}_\alpha(\mathcal{I}_d)} = 1 + O(d^{(1/k)-1}). \tag{8.57}$$

Proof. See Embrechts, Wang and Wang (2014). □

Proposition 8.36 shows that under very general assumptions typically encountered in QRM practice, we have that for d large, $\overline{\mathrm{VaR}}_\alpha(L) \approx \overline{\mathrm{ES}}_\alpha(L)$. The proposition also provides a rate of convergence. From numerical examples it appears that these asymptotic results hold fairly accurately even for small to medium values of the portfolio dimension d (see Example 8.40). From the same paper (Embrechts, Wang and Wang 2014) we add a final result related to the VaR and ES dependence-uncertainty spreads.

Proposition 8.37 (dependence-uncertainty spread of VaR versus ES). *Take $0 < \alpha_1 \leqslant \alpha_2 < 1$ and assume that the dfs F_i, $i \geqslant 1$, satisfy condition (i) of Proposition 8.36 as well as*

(iii) $\liminf_{d\to\infty} \frac{1}{d} \sum_{i=1}^{d} \mathrm{LES}_{\alpha_1}(X_i) > 0$ *and*

(iv) $\limsup_{d\to\infty} \frac{\sum_{i=1}^{d} E(X_i)}{\sum_{i=1}^{d} \mathrm{ES}_{\alpha_1}(X_i)} < 1.$

Then

$$\liminf_{d\to\infty} \frac{\overline{\mathrm{VaR}}_{\alpha_2}(\mathcal{I}_d) - \underline{\mathrm{VaR}}_{\alpha_2}(\mathcal{I}_d)}{\overline{\mathrm{ES}}_{\alpha_1}(\mathcal{I}_d) - \underline{\mathrm{ES}}_{\alpha_1}(\mathcal{I}_d)} \geqslant 1. \tag{8.58}$$

Proof. See Embrechts, Wang and Wang (2014). □

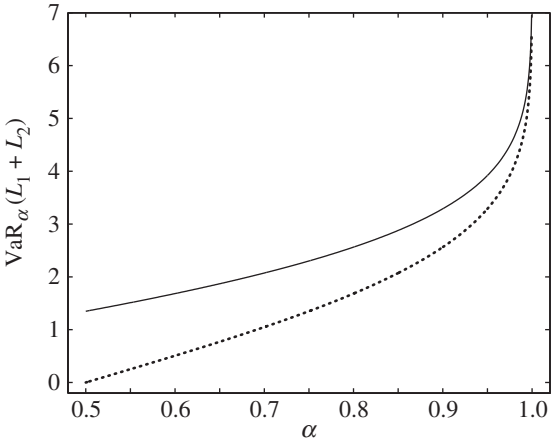

Figure 8.3. The worst-case $\overline{\mathrm{VaR}}_\alpha(L)$ (solid line) plotted against α for two standard normal risks; the case of comonotonic risks $\mathrm{VaR}_\alpha^+(L)$ is shown as a dotted line for comparison.

Remark 8.38. Propositions 8.36 and 8.37 are relevant to the ongoing discussion of risk measures for the calculation of regulatory capital. Recall that under the Basel framework for banking and also the Solvency II framework for insurance, VaR-based capital requirements are to be compared and contrasted with those based on expected shortfall. In particular, comparisons are made between VaR and ES at different quantiles, e.g. between $\mathrm{VaR}_{0.99}$ and $\mathrm{ES}_{0.975}$. The above propositions add a component of dependence uncertainty to these discussions. In particular, the dependence-uncertainty spread of VaR is generally larger than that of ES. For a numerical illustration of this, consider Example 8.40 below.

Examples. We consider examples where the aggregate loss is given by $L = L_1 + \cdots + L_d$. For any risk measure ϱ we denote by $\varrho^+(L)$ the value of ϱ when $L_1, \ldots, L_d$ are comonotonic, and we write $\varrho^\perp(L)$ when they are independent. In a first example we consider the case when $d = 2$ and $F_1 = F_2 = \Phi$, the standard normal df. In Example 8.40, higher-dimensional portfolios with Pareto margins are considered.

Example 8.39 (worst VaR for a portfolio with normal margins). For $i = 1, 2$ let $F_i = \Phi$. In Figure 8.3 we have plotted $\overline{\mathrm{VaR}}_\alpha(L)$ calculated using Proposition 8.31 as a function of α together with the curve corresponding to the comonotonic case $\mathrm{VaR}^+(L)$ calculated using Proposition 7.20. The fact that the former lies above the latter implies the existence of portfolios with normal margins for which VaR is not subadditive. For example, for $\alpha = 0.95$, the upper bound is 3.92, whereas $\mathrm{VaR}_\alpha(L_i) = 1.645$, so, for the worst VaR portfolio, $\overline{\mathrm{VaR}}_{0.95}(L) = 3.92 > 3.29 = \mathrm{VaR}_{0.95}(L_1) + \mathrm{VaR}_{0.95}(L_2)$. The density function of the distribution of (L_1, L_2) that leads to the $\overline{\mathrm{VaR}}_\alpha(L)$ is shown in Figure 8.4 (see Embrechts, Höing and Puccetti (2005) for further details).

Example 8.40 (VaR and ES bounds for Pareto margins). In Tables 8.1 and 8.2 we have applied the various results to a homogeneous Pareto case where $L_i \sim \mathrm{Pa}(\theta, 1)$,

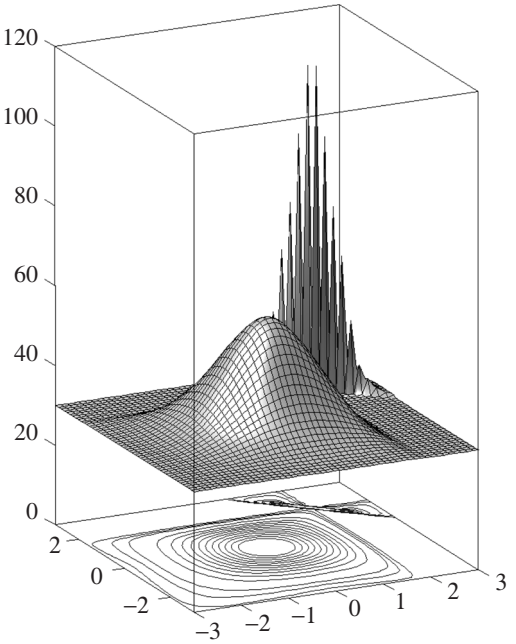

Figure 8.4. Contour and perspective plots of the density function of the distribution of (L_1, L_2) leading to the worst-case $\overline{\mathrm{VaR}}_\alpha(L)$ for $L = L_1 + L_2$ at the $\alpha = 0.95$ level when the L_i are standard normal.

$i = 1, \ldots, d$, so that the common df is $F(x) = 1 - (1 + x)^{-\theta}$, $x \geqslant 0$ (see also Section A.2.8 in the appendix). In Table 8.1 we consider cases where $\theta \geqslant 3$, corresponding to finite-variance distributions; in Table 8.2 we consider cases where $\theta \leqslant 2$, corresponding to infinite-variance distributions. The values $d = 8$ and $d = 56$ are chosen with applications to operational risk in mind. In that context $d = 8$ corresponds to a relatively low-dimensional aggregation problem and $d = 56$ to a moderately high-dimensional aggregation problem (see Chapter 13).

We use the analytic results from Propositions 8.32 and 8.34, as well as the RA. For the independent case, simulation is used. The figures given are appropriately rounded. For the homogeneous case we only report the analytic bounds (with a numerical root search for c in the propositions). The RA bounds are close to identical to their analytical counterparts, with only very small deviations for heavy-tailed dfs, i.e. for small θ. We note that for heavy-tailed risks, the RA requires a fine discretization, and hence considerably more time is needed to calculate $\underline{\mathrm{ES}}_\alpha(L)$.

Both tables confirm the results discussed above, i.e. $\overline{\mathrm{ES}}_\alpha(L)/\overline{\mathrm{VaR}}_\alpha(L)$ is close to 1 even for $d = 8$; the dependence-uncertainty spreads behave as stated in Proposition 8.37, and finally, in the $\mathrm{Pa}(\theta, 1)$, $\theta > 1$, case we have that

$$\lim_{\alpha \uparrow 1} \frac{\overline{\mathrm{ES}}_\alpha(L)}{\mathrm{VaR}^+_\alpha(L)} = \frac{\theta}{\theta - 1},$$

which can be observed in the examples given above, though the convergence is much slower here. The latter result holds more generally for distributions with regularly

Table 8.1. ES and VaR bounds for $L = L_1 + \cdots + L_d$, where $L_i \sim \mathrm{Pa}(\theta, 1)$, $i = 1, \ldots, d$, with df $F(x) = 1 - (1 + x)^{-\theta}$, $x \geqslant 0$, for $\theta \geqslant 3$. See Example 8.40 for a discussion.

	$d = 8$			$d = 56$		
$\theta = 10$	$\alpha = 95\%$	$\alpha = 99\%$	$\alpha = 99.9\%$	$\alpha = 95\%$	$\alpha = 99\%$	$\alpha = 99.9\%$
$\overline{\mathrm{VaR}}_\alpha$	4.0	6.1	9.7	28.0	42.6	68.1
VaR_α^+	2.8	4.7	8.0	19.6	32.8	55.7
$\mathrm{VaR}_\alpha^\perp$	1.5	1.9	2.5	7.8	8.6	9.6
$\underline{\mathrm{VaR}}_\alpha$	0.7	0.8	1.0	5.1	5.9	6.2
$\overline{\mathrm{ES}}_\alpha$	4.0	6.1	9.7	28.0	42.6	68.1
$\mathrm{ES}_\alpha^\perp$	1.8	2.2	2.8	8.3	9.1	10.0
$\underline{\mathrm{ES}}_\alpha$	0.9	1.2	1.7	6.2	6.2	6.2
	$d = 8$			$d = 56$		
$\theta = 5$	$\alpha = 95\%$	$\alpha = 99\%$	$\alpha = 99.9\%$	$\alpha = 95\%$	$\alpha = 99\%$	$\alpha = 99.9\%$
$\overline{\mathrm{VaR}}_\alpha$	10.2	17.1	31.7	71.4	119.8	222.7
VaR_α^+	6.6	12.1	23.8	46.0	84.7	166.9
$\mathrm{VaR}_\alpha^\perp$	3.7	5.0	7.2	18.3	20.7	24.1
$\underline{\mathrm{VaR}}_\alpha$	1.6	1.8	3.0	11.0	12.9	13.8
$\overline{\mathrm{ES}}_\alpha$	10.2	17.1	31.8	71.4	119.8	222.7
$\mathrm{ES}_\alpha^\perp$	4.5	5.9	8.7	19.8	22.2	26.0
$\underline{\mathrm{ES}}_\alpha$	2.5	3.8	6.5	14.0	14.0	14.3
	$d = 8$			$d = 56$		
$\theta = 3$	$\alpha = 95\%$	$\alpha = 99\%$	$\alpha = 99.9\%$	$\alpha = 95\%$	$\alpha = 99\%$	$\alpha = 99.9\%$
$\overline{\mathrm{VaR}}_\alpha$	24.1	46.9	110.2	171.9	333.7	783.7
VaR_α^+	13.7	29.1	72.0	96.0	203.9	504.0
$\mathrm{VaR}_\alpha^\perp$	8.1	12.3	23.0	39.1	47.6	67.2
$\underline{\mathrm{VaR}}_\alpha$	2.9	3.6	9.0	20.4	24.9	27.2
$\overline{\mathrm{ES}}_\alpha$	24.6	47.7	112.0	172.0	333.9	784.0
$\mathrm{ES}_\alpha^\perp$	11.0	17.0	32.9	44.9	56.2	85.5
$\underline{\mathrm{ES}}_\alpha$	7.2	12.9	29.0	28.6	31.3	56.4

varying tails (see Definition 5.7 and Karamata's Theorem (Appendix A.1.4) and recall that this includes distributions like the Student t and loggamma distributions).

Table 8.2 also includes the case $\theta = 0.8$, i.e. an infinite-mean case (for which ES is not defined). Here we note that $\mathrm{VaR}_\alpha^\perp(L) > \mathrm{VaR}_\alpha^+(L)$, so this gives an example of superadditivity of VaR_α in the case of independence (see the discussion following Example 2.25).

Notes and Comments

The use of a standard formula approach based on the kind of aggregation embodied in (8.47) is permitted under Solvency II; see CEIOPS (2006), a document produced

Table 8.2. ES and VaR bounds for $L = L_1 + \cdots + L_d$, where $L_i \sim \mathrm{Pa}(\theta, 1)$, $i = 1, \ldots, d$, with df $F(x) = 1 - (1 + x)^{-\theta}$, $x \geqslant 0$, for $\theta \leqslant 2$. See Example 8.40 for a discussion.

	$d = 8$			$d = 56$		
$\theta = 2$	$\alpha = 95\%$	$\alpha = 99\%$	$\alpha = 99.9\%$	$\alpha = 95\%$	$\alpha = 99\%$	$\alpha = 99.9\%$
$\overline{\mathrm{VaR}}_\alpha$	59	142	465	440	1 054	3 454
VaR_α^+	28	72	245	194	504	1 715
$\mathrm{VaR}_\alpha^\perp$	18	35	96	89	132	293
$\underline{\mathrm{VaR}}_\alpha$	5	9	31	36	46	53
$\overline{\mathrm{ES}}_\alpha$	64	152	498	445	1 064	3 486
$\mathrm{ES}_\alpha^\perp$	31	63	184	123	205	518
$\underline{\mathrm{ES}}_\alpha$	24	56	178	75	149	472
	$d = 8$			$d = 56$		
$\theta = 1.5$	$\alpha = 95\%$	$\alpha = 99\%$	$\alpha = 99.9\%$	$\alpha = 95\%$	$\alpha = 99\%$	$\alpha = 99.9\%$
$\overline{\mathrm{VaR}}_\alpha$	135	409	1 928	1 100	3 323	15 629
VaR_α^+	51	164	792	357	1 150	5 544
$\mathrm{VaR}_\alpha^\perp$	39	98	413	207	421	1 574
$\underline{\mathrm{VaR}}_\alpha$	8	21	99	56	77	99
$\overline{\mathrm{ES}}_\alpha$	169	509	2 392	1 182	3 563	16 744
$\mathrm{ES}_\alpha^\perp$	98	265	1 159	419	1 016	4 126
$\underline{\mathrm{ES}}_\alpha$	88	258	1 199	323	945	4 390
	$d = 8$			$d = 56$		
$\theta = 0.8$	$\alpha = 95\%$	$\alpha = 99\%$	$\alpha = 99.9\%$	$\alpha = 95\%$	$\alpha = 99\%$	$\alpha = 99.9\%$
$\overline{\mathrm{VaR}}_\alpha$	2 250	16 873	300 182	35 168	263 301	4 683 172
VaR_α^+	330	2 522	44 979	2 313	17 653	314 855
$\mathrm{VaR}_\alpha^\perp$	620	4 349	75 877	7 318	49 858	862 855
$\underline{\mathrm{VaR}}_\alpha$	41	315	5 622	207	433	5 622

by the Committee of European Insurance and Operational Pensions Supervisors (now EIOPA).

In the banking context the summation approach in (8.46) is commonly used, particularly for the aggregation of capital requirements for market and credit risk. As explained by Breuer et al. (2010) this is commonly justified by assuming that credit risk arises from the banking book and market risk from the trading book, but they point out that, for derivative instruments depending on both market and credit risks, it can potentially lead to underestimation of risk. In contrast, Alessandri and Drehmann (2010) study integration of credit risk and interest-rate risk in the banking book and conduct simulations suggesting that summation of capital for the two risk types is likely to be too conservative; Drehmann, Sorenson and Stringa (2010) argue that credit and interest-rate risk must be assessed jointly in the banking book.

Kretzschmar, McNeil and Kirchner (2010) make similar points about the importance of developing fully integrated models rather than modular approaches.

A summary of methodological practice in economic capital models before the 2007 crisis is presented in a comprehensive survey by the International Financial Risk Institute that included both banks and insurance companies. In this survey, the prevailing approach to integration is reported to be the use of correlation matrices (see IFRI Foundation and CRO Forum 2007). This approach was favoured by over 75% of the surveyed banks, with the others using simulation approaches based on scenario generation or hybrid approaches. In the insurance industry there was more diversity in the approaches used for integration: around 35% of respondents used the correlation approach and about the same number used simulation; the remainder reported the use of copulas or hybrid approaches.

There is a large literature on Fréchet problems: see, for instance, Chapter 2 in Rüschendorf (2013). From a QRM perspective, Embrechts and Puccetti (2006) gave the field a considerable boost. The latter paper also contains the most important references to the early literature. Historically, the question of bounding the df of a sum of rvs with given marginals goes back to Kolmogorov and was answered by Makarov (1981) for $d = 2$. Frank, Nelsen and Schweizer (1987) restated Makarov's result using the notion of a copula. Independently, Rüschendorf (1982) gave a very elegant proof of the same result using duality. Williamson and Downs (1990) introduced the use of dependence information. Numerous other authors (especially in analysis and actuarial mathematics) have contributed to this area. Besides the comprehensive book by Müller and Stoyan (2002), several other texts in actuarial mathematics contain interesting contributions on dependence modelling: for an introduction, see Chapter 10 in Kaas et al. (2001). A rich set of optimization problems within an actuarial context are to be found in De Vylder (1996); see especially "Part II: Optimization Theory", where the author "shows how to obtain best upper and lower bounds on functionals $T(F)$ of the df F of a risk, under moment or other integral constraints". An excellent account is to be found in Denuit and Charpentier (2004). The definitive account from an actuarial point of view is Denuit et al. (2005). A wealth of actuarial examples is to be found in the two extensive articles Hürlimann (2008a) and Hürlimann (2008b).

The rearrangement algorithm (RA) for VaR appeared in Embrechts, Puccetti and Rüschendorf (2013) and was based on earlier work by Puccetti and Rüschendorf (2012). Full details on the RA are collected by Giovanni Puccetti at https://sites.google.com/site/rearrangementalgorithm/. The interested reader may also search the literature for probability box (or p-box) and the related *Dempster–Shafer Theory*. These search items lead to well-established theory and numerous examples in the realm of engineering, computer science and economics. For expected shortfall, the RA was worked out in Puccetti (2013). For the analytical results and a discussion of the sharpness of the various bounds for VaR and ES, the papers cited for the corresponding propositions give an excellent introduction. Some further interesting papers are Bernard, Jiang and Wang (2014), Bernard et al. (2013) and Bernard, Rüschendorf and Vanduffel (2013).

The notion of complete mixability leads to a condition of negative dependence for multivariate ($d \geqslant 2$) random vectors. Recent developments in this field are summarized in Puccetti and Wang (2015), Puccetti and Wang (2014) and Wang and Wang (2014).

Rosenberg and Schuermann (2006) gives some idea of the applicability of aggregation ideas used in this chapter. The authors construct the joint risk distribution for a typical, large, internationally active bank using the method of copulas and aggregate risk measures across the categories of market, credit and operational risk.

For an illustration of the ideas and results of Section 8.4.4 in the practical environment of a Norwegian financial group, see Dimakos and Aas (2004) and Aas and Puccetti (2014). The latter paper contains a discussion of the best/worst couplings. See also Embrechts, Puccetti and Rüschendorf (2013) on this topic.

8.5 Capital Allocation

The final section of this chapter essentially looks at the converse problem to Section 8.4. Given a model for aggregate losses we now consider how the overall capital requirement may be disaggregated into additive contributions attributable to the different sub-portfolios or assets that make up the overall portfolio.

8.5.1 The Allocation Problem

As in Section 8.4.1 let the rvs $L_1, \ldots, L_d$ represent the losses (or negative P&Ls) arising from d different lines of business, or the losses corresponding to d different asset classes on the balance sheet of a firm. In this section we will refer to these sub-units of a larger portfolio simply as investments. The allocation problem can be motivated by considering the question of how we might measure the risk-adjusted performance of different investments within a portfolio.

The performance of investments is usually measured using a RORAC (return on risk-adjusted capital) approach, i.e. by considering a ratio of the form

$$\frac{\text{expected profit of investment } i}{\text{risk capital for investment } i}. \tag{8.59}$$

The general approach embodied in (8.59) raises the question of how we should calculate the risk capital for an investment that is part of a larger portfolio. It should not simply be the stand-alone risk capital for that investment considered in isolation; this would neglect the issue of diversification and give an inaccurate measure of the performance of an investment within the larger portfolio. Instead, the risk capital for an investment within a portfolio should reflect the contribution of that investment to the overall riskiness of the portfolio. A two-step procedure for determining these contributions is used in practice.

(1) Compute the overall risk capital $\varrho(L)$, where $L = \sum_{i=1}^{d} L_i$ and ϱ is a particular risk measure such as VaR, ES or a mean-adjusted version of one of these (see (8.48)); note that at this stage we are not stipulating that ϱ must be coherent.

(2) Allocate the capital $\varrho(L)$ to the individual investments according to some mathematical *capital allocation principle* such that, if AC_i denotes the capital allocated to the investment with potential loss L_i (the so-called *risk contribution* of unit i), the sum of the risk contributions corresponds to the overall risk capital $\varrho(L)$.

In this section we are interested in step (2) of the procedure; loosely speaking, we require a mapping that takes as input the individual losses $L_1, \dots, L_d$ and the risk measure ϱ and yields as output the vector of risk contributions $(\mathrm{AC}_1, \dots, \mathrm{AC}_d)$ such that

$$\varrho(L) = \sum_{i=1}^{d} \mathrm{AC}_i, \tag{8.60}$$

and such a mapping will be called a capital allocation principle. The relation (8.60) is sometimes called the *full allocation property* since all of the overall risk capital $\varrho(L)$ (not more, not less) is allocated to the investments; we consider this property to be an integral part of the definition of an allocation principle. Of course, there are other properties of a capital allocation principle that are desirable from an economic viewpoint; we first make some formal definitions and give examples of allocation principles before discussing further properties.

The formal set-up. Let $L_1, \dots, L_d$ be rvs on a common probability space $(\Omega, \mathcal{F}, P)$ representing losses (or profits) for d investments. For our discussion it will be useful to consider portfolios where the weights of the individual investments are varied with respect to our basic portfolio $(L_1, \dots, L_d)$, which is regarded as a fixed random vector. That is, we consider an open set $\Lambda \subset \mathbb{R}^d \setminus \{\mathbf{0}\}$ of portfolio weights such that $\mathbf{1} \in \Lambda$ and define for $\boldsymbol{\lambda} \in \Lambda$ the loss $L(\boldsymbol{\lambda}) = \sum_{i=1}^{d} \lambda_i L_i$; the loss of our actual portfolio is of course $L(\mathbf{1})$. Let ϱ be some risk measure defined on a set $\mathcal{M}$ that contains the rvs $\{L(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$. As in Section 8.3.1 we use the associated risk-measure function $r_\varrho : \Lambda \to \mathbb{R}$ with $r_\varrho(\boldsymbol{\lambda}) = \varrho(L(\boldsymbol{\lambda}))$.

8.5.2 The Euler Principle and Examples

From now on we restrict our attention to risk measures that are positive homogeneous. This may be a coherent risk measure, or a mean-adjusted version of a coherent risk measure as in (8.48); it may also be VaR (or a mean-corrected version of VaR) or the standard deviation risk measure. Obviously, the associated risk-measure function must satisfy $r_\varrho(t\boldsymbol{\lambda}) = t r_\varrho(\boldsymbol{\lambda})$ for all $t > 0$, $\boldsymbol{\lambda} \in \Lambda$, so $r_\varrho : \Lambda \to \mathbb{R}$ is a positive-homogeneous function of a vector argument. Recall Euler's well-known rule that states that if r_ϱ is positive homogeneous and differentiable at $\boldsymbol{\lambda} \in \Lambda$, we have

$$r_\varrho(\boldsymbol{\lambda}) = \sum_{i=1}^{d} \lambda_i \frac{\partial r_\varrho}{\partial \lambda_i}(\boldsymbol{\lambda}). \tag{8.61}$$

If we apply this at $\boldsymbol{\lambda} = \mathbf{1}$, we get, using that $\varrho(L) = r_\varrho(\mathbf{1})$,

$$\varrho(L) = \sum_{i=1}^{d} \frac{\partial r_\varrho}{\partial \lambda_i}(\mathbf{1}).$$

This suggests the following definition.

Definition 8.41 (Euler capital allocation principle). If r_ϱ is a positive-homogeneous risk-measure function, which is differentiable at $\boldsymbol{\lambda} = \mathbf{1}$, then the Euler capital allocation principle associated with ϱ has risk contributions

$$\mathrm{AC}_i^\varrho = \frac{\partial r_\varrho}{\partial \lambda_i}(\mathbf{1}), \quad 1 \leqslant i \leqslant d.$$

The Euler principle is sometimes called *allocation by the gradient*, and it obviously gives a full allocation of the risk capital. We now look at a number of specific examples of Euler allocations corresponding to different choices of risk measure ϱ.

Standard deviation and the covariance principle. Consider the risk-measure function $r_{\mathrm{SD}}(\boldsymbol{\lambda}) = \sqrt{\mathrm{var}(L(\boldsymbol{\lambda}))}$ and write Σ for the covariance matrix of $(L_1, \dots, L_d)$. Then we have $r_{\mathrm{SD}}(\boldsymbol{\lambda}) = (\boldsymbol{\lambda}' \Sigma \boldsymbol{\lambda})^{1/2}$, from which it follows that

$$\mathrm{AC}_i^\varrho = \frac{\partial r_{\mathrm{SD}}}{\partial \lambda_i}(\mathbf{1}) = \frac{(\Sigma \mathbf{1})_i}{r_{\mathrm{SD}}(\mathbf{1})} = \frac{\sum_{j=1}^d \mathrm{cov}(L_i, L_j)}{r_{\mathrm{SD}}(\mathbf{1})} = \frac{\mathrm{cov}(L_i, L)}{\sqrt{\mathrm{var}(L)}}.$$

This formula is known as the *covariance principle*. If we consider more generally a risk measure of the form $\varrho(L) = E(L) + \kappa\, \mathrm{SD}(L)$ for some $\kappa > 0$, we get $r_\varrho(\boldsymbol{\lambda}) = \boldsymbol{\lambda}' E(\boldsymbol{L}) + \kappa r_{\mathrm{SD}}(\boldsymbol{\lambda})$ and hence

$$\mathrm{AC}_i^\varrho = E(L_i) + \kappa \frac{\mathrm{cov}(L_i, L)}{\sqrt{\mathrm{var}(L)}}.$$

VaR and VaR contributions. Suppose that $r_{\mathrm{VaR}}^\alpha(\boldsymbol{\lambda}) = q_\alpha(L(\boldsymbol{\lambda}))$. In this case it can be shown that, subject to technical conditions,

$$\mathrm{AC}_i^\varrho = \frac{\partial r_{\mathrm{VaR}}^\alpha}{\partial \lambda_i}(\mathbf{1}) = E(L_i \mid L = q_\alpha(L)), \quad 1 \leqslant i \leqslant d. \tag{8.62}$$

The derivation of (8.62) is more involved than that of the covariance principle, and we give a justification following Tasche (2000) under the simplifying assumption that the loss distribution of $(L_1, \dots, L_d)$ has a joint density f. In the following lemma we denote by $\phi(u, l_2, \dots, l_d) = f_{L_1 | L_2, \dots, L_d}(u \mid l_2, \dots, l_d)$ the conditional density of L_1.

Lemma 8.42. *Assume that $d \geqslant 2$ and that $(L_1, \dots, L_d)$ has a joint density. Then, for any vector $(\lambda_1, \dots, \lambda_d)$ of portfolio weights such that $\lambda_1 \neq 0$, we find that*

(i) *$L(\boldsymbol{\lambda})$ has density*

$$f_{L(\boldsymbol{\lambda})}(t) = |\lambda_1|^{-1} E\left(\phi\left(\lambda_1^{-1}\left(t - \sum_{j=2}^d \lambda_j L_j\right), L_2, \dots, L_d\right)\right);$$

and

(ii) *for $i = 2, \dots, d$,*

$$E(L_i \mid L(\boldsymbol{\lambda}) = t) = \frac{E(L_i \phi(\lambda_1^{-1}(t - \sum_{j=2}^d \lambda_j L_j), L_2, \dots, L_d))}{E(\phi(\lambda_1^{-1}(t - \sum_{j=2}^d \lambda_j L_j), L_2, \dots, L_d))}, \quad \text{a.s.}$$

Proof. For (i) consider the case $\lambda_1 > 0$ and observe that we can write

$$\begin{aligned} P(L(\boldsymbol{\lambda}) \leqslant t) &= E(P(L(\boldsymbol{\lambda}) \leqslant t \mid L_2, \dots, L_d)) \\ &= E\bigg(P\bigg(L_1 \leqslant \lambda_1^{-1}\bigg(t - \sum_{j=2}^{d} \lambda_j L_j\bigg) \;\bigg|\; L_2, \dots, L_d\bigg)\bigg) \\ &= E\bigg(\int_{-\infty}^{\lambda_1^{-1}(t-\sum_{j=2}^{d}\lambda_j L_j)} \phi(u, L_2, \dots, L_d)\,\mathrm{d}u\bigg). \end{aligned}$$

The assertion follows by differentiating under the expectation.

For (ii) observe that we can write

$$E(L_i \mid L(\boldsymbol{\lambda}) = t) = \lim_{\delta \to 0} \frac{\delta^{-1} E(L_i I_{\{t < L(\boldsymbol{\lambda}) \leqslant t+\delta\}})}{\delta^{-1} P(t < L(\boldsymbol{\lambda}) \leqslant t + \delta)} = \frac{(\partial/\partial t) E(L_i I_{\{L(\boldsymbol{\lambda}) \leqslant t\}})}{f_{L(\boldsymbol{\lambda})}(t)},$$

provided $f_{L(\boldsymbol{\lambda})}(t) \neq 0$. The result follows by applying a similar conditioning technique to the ones used in the proof of (i) to the numerator. □

We now explain why (8.62) follows from Lemma 8.42. Since the rv $L(\boldsymbol{\lambda})$ has a density, we have $P(L(\boldsymbol{\lambda}) \leqslant q_\alpha(L(\boldsymbol{\lambda}))) = \alpha$. Writing $k(t) = \lambda_1^{-1}(t - \sum_{j=2}^{d} \lambda_j L_j)$, we have

$$\alpha = P(L(\boldsymbol{\lambda}) \leqslant r^\alpha_{\mathrm{VaR}}(\boldsymbol{\lambda})) = E\bigg(\int_{-\infty}^{k(r^\alpha_{\mathrm{VaR}}(\boldsymbol{\lambda}))} \phi(u, L_2, \dots, L_d)\,\mathrm{d}u\bigg). \tag{8.63}$$

We take derivatives of (8.63) with respect to λ_i for $i = 2, \dots, d$ to get

$$0 = \lambda_1^{-1} E\bigg(\bigg(\frac{\partial r^\alpha_{\mathrm{VaR}}(\boldsymbol{\lambda})}{\partial \lambda_i} - L_i\bigg)\phi(k(r^\alpha_{\mathrm{VaR}}(\boldsymbol{\lambda})), L_2, \dots, L_d)\bigg).$$

Solving this expression for $\partial r^\alpha_{\mathrm{VaR}}(\boldsymbol{\lambda})/\partial \lambda_i$, using part (ii) of Lemma 8.42 and substituting $\boldsymbol{\lambda} = \mathbf{1}$ yields (8.62), as desired. Analogous calculations can be done for $i = 1$ and $\lambda_1 < 0$. Tasche (2000) makes the derivations mathematically rigorous by using the implicit function theorem and giving all necessary conditions.

Expected shortfall and shortfall contributions. Now consider using the risk-measure function $r^\alpha_{\mathrm{ES}}(\boldsymbol{\lambda}) = E(L \mid L \geqslant q_\alpha(L(\boldsymbol{\lambda})))$ corresponding to expected shortfall. It follows from Definition 2.12 that we can write

$$r^\alpha_{\mathrm{ES}}(\boldsymbol{\lambda}) = \frac{1}{1-\alpha} \int_\alpha^1 r^u_{\mathrm{VaR}}(\boldsymbol{\lambda})\,\mathrm{d}u,$$

where we make use of the notation $r^\alpha_{\mathrm{VaR}}(\boldsymbol{\lambda}) = q_\alpha(L(\boldsymbol{\lambda}))$ as above. We apply the Euler principle by again computing the derivative with respect to λ_i. Assuming the differentiability of $r^u_{\mathrm{VaR}}(\boldsymbol{\lambda})$, we have, with $L = L(\mathbf{1})$,

$$\frac{\partial r^\alpha_{\mathrm{ES}}}{\partial \lambda_i}(\mathbf{1}) = \frac{1}{1-\alpha} \int_\alpha^1 \frac{\partial r^u_{\mathrm{VaR}}}{\partial \lambda_i}(\mathbf{1})\,\mathrm{d}u = \frac{1}{1-\alpha} \int_\alpha^1 E(L_i \mid L = q_u(L)\,\mathrm{d}u.$$

Now we assume that the density f_L of L is strictly positive so that the df of L has a differentiable inverse and we can make the change of variables $v = q_u(L) = F_L^{\leftarrow}(u)$. Since $\mathrm{d}v/\mathrm{d}u = (f_L(v))^{-1}$, we get

$$\frac{\partial r_{\mathrm{ES}}^{\alpha}}{\partial \lambda_i}(\mathbf{1}) = \frac{1}{1-\alpha}\int_{q_\alpha(L)}^{\infty} E(L_i \mid L = v) f_L(v)\,\mathrm{d}v = \frac{1}{1-\alpha} E(L_i; L \geqslant q_\alpha(L))).$$

Hence the Euler capital allocation takes the form

$$\mathrm{AC}_i^{\varrho} = E(L_i \mid L \geqslant \mathrm{VaR}_\alpha(L)), \quad L := L(\mathbf{1}), \tag{8.64}$$

where AC_i^{ϱ} is known as the *expected shortfall contribution* of investment possibility (or line of business) i. This is a popular allocation principle in practice, and is often considered to be preferable to the covariance principle and the principle based on VaR contributions. See Notes and Comments for literature on its use in practice in the context of credit portfolios.

Euler allocation for elliptical loss distributions. In the following corollary to Theorem 8.28 we consider the special case of an elliptical loss distribution for the vector $(L_1, \ldots, L_d)$. We consider this distribution to be centred at zero so that it really represents fluctuations of the loss around its mean; centring $(L_1, \ldots, L_d)$ is of course equivalent to working with the mean-adjusted version of some translation-invariant risk measure ϱ. We find that the relative amounts of capital allocated to each investment opportunity are always the same, regardless of whether we base an Euler allocation on the standard deviation, VaR or expected shortfall risk measures, or indeed any positive-homogeneous risk measure. Allocation is therefore very simple in this case: depending on our choice of risk measure we calculate the total risk capital to be allocated and then use a simple partitioning formula given in (8.65) below.

Corollary 8.43. *Assume that $r_\varrho : \Lambda \to \mathbb{R}$ is the risk-measure function of a positive-homogeneous and law-invariant risk measure ϱ. Let $\boldsymbol{L} \sim E_d(\mathbf{0}, \Sigma, \psi)$. Then, under an Euler allocation, the relative capital allocation is given by*

$$\frac{\mathrm{AC}_i^{\varrho}}{\mathrm{AC}_j^{\varrho}} = \frac{\sum_{k=1}^d \Sigma_{ik}}{\sum_{k=1}^d \Sigma_{jk}}, \quad 1 \leqslant i, j \leqslant d. \tag{8.65}$$

Proof. From the proof of Theorem 8.28 we deduce that, by the positive homogeneity of the risk measure, we have

$$r_\varrho(\boldsymbol{\lambda}) = \varrho(L(\boldsymbol{\lambda})) = \varrho\Bigg(\sum_{i=1}^d \lambda_i L_i\Bigg) = \sqrt{\boldsymbol{\lambda}'\Sigma\boldsymbol{\lambda}}\varrho(Y_1),$$

where Y_1 is the first component of a spherical random vector with characteristic generator ψ. For the Euler allocation we get

$$\mathrm{AC}_i^{\varrho} = \frac{\partial r_\varrho}{\partial \lambda_i}(\mathbf{1}) = \frac{\sum_{k=1}^d \Sigma_{ik}}{\sqrt{\mathbf{1}'\Sigma\mathbf{1}}}\varrho(Y_1),$$

from which the result follows. □

8.5.3 Economic Properties of the Euler Principle

In this section we show that the Euler principle has a number of good economic properties. As in the previous section we consider a positive-homogeneous risk measure ϱ and we assume that the corresponding risk-measure function r_ϱ is continuously differentiable in $\mathbb{R}^d \setminus \{0\}$ (a positive-homogeneous function is typically not differentiable in $\boldsymbol{\lambda} = 0$). By $\mathrm{AC}_i^\varrho = \partial r_\varrho(\mathbf{1})/\partial\lambda_i$ we then denote the associated risk contributions under the Euler principle.

Compatibility with a RORAC approach. We define the RORAC of the overall loss by $\mathrm{RORAC}(L) := E(-L)/\rho(L)$; the portfolio-related RORAC of investment unit i is defined as

$$\mathrm{RORAC}(L_i \mid L) := \frac{E(-L_i)}{\mathrm{AC}_i^\varrho},$$

where it is tacitly assumed that the denominator is strictly positive. The Euler principle is then compatible with a RORAC approach in the following sense: if investment opportunity i performs better than the overall portfolio L in the RORAC metric, then the RORAC of the overall portfolio is increased if one increases slightly the weight of unit i. The Euler principle therefore gives correct signals for investment decisions. In mathematical terms, RORAC compatibility means that there is some $\varepsilon > 0$ such that for all $0 < h \leqslant \varepsilon$ it holds that

$$(\mathrm{RORAC}(L_i \mid L) > \mathrm{RORAC}(L)) \Rightarrow (\mathrm{RORAC}(L + hL_i) > \mathrm{RORAC}(L)). \quad (8.66)$$

In order to establish (8.66) it suffices to show that $\mathrm{RORAC}(L_i \mid L) > \mathrm{RORAC}(L)$ implies that $(\mathrm{d}/\mathrm{d}h)\,\mathrm{RORAC}(L + hL_i)|_{h=0} > 0$. Denote by $\boldsymbol{e}_i$ the ith unit vector. Then it holds that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}h}\,\mathrm{RORAC}(L + hL_i)\bigg|_{h=0} &= \frac{\mathrm{d}}{\mathrm{d}h}\,\frac{E(-(L + hL_i))}{r_\varrho(\mathbf{1} + h\boldsymbol{e}_i)}\bigg|_{h=0} \\ &= \frac{1}{r_\varrho(\mathbf{1})^2}\left(E(-L_i)r_\varrho(\mathbf{1}) - E(-L)\frac{\partial r_\varrho(\mathbf{1})}{\partial\lambda_i}\right).\end{aligned}$$

Recall that $\varrho(L) = r_\varrho(\mathbf{1})$ and that $\mathrm{AC}_i^\varrho = \partial r_\varrho(\mathbf{1})/\partial\lambda_i$. Hence the last expression is strictly positive if $E(-L_i)/\mathrm{AC}_i^\varrho > E(-L)/\varrho(L)$, as claimed.

In fact, it can be shown that for a positive-homogeneous ϱ the Euler principle is the only capital allocation principle that satisfies the RORAC compatibility (8.66) (see Tasche (1999) for details).

Diversification benefit. Suppose that the risk measure ϱ is positive homogeneous and subadditive, as is the case for a coherent risk measure or a mean-adjusted version thereof. In that case, since $\varrho(L) \leqslant \sum_{i=1}^d \varrho(L_i)$, the overall risk capital required for the portfolio is smaller than the sum of the risk capital required for the business units on a stand-alone basis. In practice, the difference $\sum_{i=1}^d \varrho(L_i) - \varrho(L)$ is known as the *diversification benefit*. It is reasonable to require that each business unit profits from the diversification benefit in the sense that the individual risk contribution of unit i does not exceed the stand-alone capital charge $\varrho(L_i)$ (otherwise there would

be an incentive for unit i to leave the firm, at least in theory). We now show that the Euler principle does indeed satisfy the inequality

$$\mathrm{AC}_i^{\varrho} \leqslant \varrho(L_i), \quad 1 \leqslant i \leqslant d. \tag{8.67}$$

The key is the following inequality: for a convex and positive-homogeneous function $f : \mathbb{R}^d \to \mathbb{R}$ that is continuously differentiable in $\mathbb{R}^d \setminus \{0\}$, it holds for all $\boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}}$ with $\boldsymbol{\lambda} \neq -\tilde{\boldsymbol{\lambda}}$ that

$$f(\boldsymbol{\lambda}) \geqslant \sum_{i=1}^{d} \lambda_i \frac{\partial f(\boldsymbol{\lambda} + \tilde{\boldsymbol{\lambda}})}{\partial \lambda_i}. \tag{8.68}$$

If we apply this inequality with $f = r_{\varrho}$ (which is convex as ϱ is positive homogeneous and subadditive), $\boldsymbol{\lambda} = \boldsymbol{e}_i$ and $\tilde{\boldsymbol{\lambda}} = \mathbf{1} - \boldsymbol{e}_i$, we get the inequality $r_{\varrho}(\boldsymbol{e}_i) \geqslant \partial r_{\varrho}(\mathbf{1})/\partial \lambda_i$ and hence (8.67).

It remains to establish the inequality (8.68). Since f is convex it holds for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, $\boldsymbol{x} \neq 0$, that

$$f(\boldsymbol{y}) \geqslant f(\boldsymbol{x}) + \sum_{i=1}^{d} (y_i - x_i) \frac{\partial f(\boldsymbol{x})}{\partial x_i}.$$

Moreover, by Euler's rule we have $f(\boldsymbol{x}) = \sum_{i=1}^{d} x_i \partial f(\boldsymbol{x})/\partial x_i$ and hence

$$f(\boldsymbol{y}) \geqslant \sum_{i=1}^{d} y_i \frac{\partial f(\boldsymbol{x})}{\partial x_i}.$$

Substituting $\boldsymbol{y} = \boldsymbol{\lambda}$, $\boldsymbol{x} = \boldsymbol{\lambda} + \tilde{\boldsymbol{\lambda}}$ gives the result.

The work of Kalkbrener (2005) and Denault (2001) takes this analysis one step further. In these papers it is shown that under suitable technical conditions the Euler principle is the only capital allocation principle that satisfies a slight strengthening of the diversification-benefit inequality (8.67). Obviously, this gives additional support for using the Euler principle if one works in the realm of coherent risk measures. From a practical point of view, the use of expected shortfall and expected shortfall contributions might be a reasonable choice in many application areas, particularly for credit risk management and loan pricing (see Notes and Comments, where this issue is discussed further).

Notes and Comments

A broad, non-technical discussion of capital allocation and performance measurement is to be found in Matten (2000) (see also Klaassen and van Eeghen 2009). The term "Euler principle" seems to have first been used in Patrik, Bernegger and Rüegg (1999). The result (8.62) is found in Gouriéroux, Laurent and Scaillet (2000) and Tasche (2000); the former paper assumes that the losses have a joint density and the latter gives a slightly more general result as well as technical details concerning the differentiability of the VaR and ES risk measures with respect to the portfolio composition. Differentiability of the coherent premium principle of Section 2.3.5 is discussed in Fischer (2003). The derivation of allocation principles from properties

of risk measures is also to be found in Goovaerts, Dhaene and Kaas (2003) and Goovaerts, van den Borre and Laeven (2005).

For the arguments concerning suitability of risk measures for performance measurement, see Tasche (1999) and Tasche (2008). An axiomatic approach to capital allocation is found in Kalkbrener (2005) and Denault (2001). For an early contribution on game theory applied to cost allocation in an insurance context, see Lemaire (1984).

Applications to credit risk are found in Kalkbrener, Lotter and Overbeck (2004) and Merino and Nyfeler (2004); these make strong arguments in favour of the use of expected shortfall contributions. On the other hand, Pfeifer (2004) contains some compelling examples to show that expected shortfall as a risk measure and expected shortfall contributions as an allocation method may have some serious deficiencies when used in non-life insurance. The existence of rare, extreme events may lead to absurd capital allocations when based on expected shortfall. The reader is therefore urged to reflect carefully before settling on a specific risk measure and allocation principle. It may also be questionable to base a "coherent" risk-sensitive capital allocation on formal criteria only; for further details on this from a non-life insurance perspective see Koryciorz (2004).

Risk-adjusted performance measures are widely used in industry in the context of capital budgeting and performance measurement. A good overview of current practice is given in Chapter 14 of Crouhy, Galai and Mark (2001) (see also Klaassen and van Eeghen 2009). An analysis of risk management and capital budgeting for financial institutions from an economic viewpoint is given in Froot and Stein (1998).

Part III

Applications

9
Market Risk

In this chapter we look at methods for measuring the market risk in portfolios of traded instruments. We emphasize the use of statistical models and techniques introduced in Part II of the book. While we draw on material from most of the foregoing chapters, essential prerequisites are Chapter 2, in which the basic risk measurement problem was introduced, and Chapter 4 on financial time series. The material is divided into three sections.

In Section 9.1 we revisit the topic of risk factors and mappings, first described in very general terms in Section 2.2. We develop the modelling framework in more detail in this chapter for the specific problem of modelling market risk in a bank's trading book, where derivative positions are common and the regulator requires risk to be measured over short time horizons such as one day or two trading weeks.

Section 9.2 is devoted to the topic of market-risk measurement. Assuming that the portfolio has been mapped to risk factors, we describe the various statistical approaches that are used in industry to estimate loss distributions and risk measures like VaR or expected shortfall. These methods include the variance–covariance (delta-normal), historical simulation and Monte Carlo methods.

The subject of backtesting the performance of such methods is treated in Section 9.3. We describe commonly used model-validation procedures based on VaR violations as well as more recent proposals for comparing methods using scoring functions based on elicitability theory.

9.1 Risk Factors and Mapping

The key idea in this section is that of a *loss operator*, which is introduced in Section 9.1.1. This is a function that relates portfolios losses to changes in the risk factors and is effectively the function that a bank must evaluate in order to determine the P&L of its trading book under scenarios for future risk-factor changes. Since the time to maturity or expiry has an impact on the value of many market instruments, we consider the issue of different timescales for risk measurement and valuation in detail. In Section 9.1.2 we show how the typically non-linear loss operator can be approximated over short time intervals by linear (delta) and quadratic (delta–gamma) functions.

The methodology is applied to a portfolio of zero-coupon bonds in Section 9.1.3, and it is shown that the linear and quadratic approximations to the loss operator have

interpretations in terms of the classical bond pricing concepts of duration and convexity. Since the mapping of fixed-income portfolios is typically a high-dimensional problem, we consider factor modelling strategies for reducing the complexity of the mapping exercise in Section 9.1.4.

9.1.1 The Loss Operator

Consider a portfolio of assets subject to market risk, such as a collection of stocks and bonds or a book of derivatives. The value of the portfolio is given by the continuous-time stochastic process $(V(t))_{t\in\mathbb{R}}$, where it is assumed that $V(t)$ is *known* at time t; this means that the instruments in the portfolio can either be marked-to-market or marked to an appropriate model (see Section 2.2.2 for discussion of these concepts).

For a given time horizon Δt, such as one or ten days in a typical market-risk application, the P&L of the portfolio over the period $[t, t+\Delta t]$ is given by $V(t+\Delta t)-V(t)$. We find it convenient to consider the negative P&L $-(V(t+\Delta t)-V(t))$ and to represent the risk by the right tail of this quantity, which we refer to simply as the loss. It is assumed that the portfolio composition remains constant over this period and that there is no intermediate income or fees (the so-called clean or no-action P&L).

In transforming the problem of analysing the loss distribution to a problem in financial time-series analysis, it is convenient to measure time in units of Δt and to introduce appropriate time-series notation. In a number of places in this chapter we move from a generic continuous-time process $Y(t)$ to the time series $(Y_t)_{t\in\mathbb{Z}}$ by setting

$$Y_t := Y(\tau_t), \quad \tau_t := t(\Delta t). \tag{9.1}$$

Using this notation the loss is written as

$$L_{t+1} := -(V(\tau_{t+1}) - V(\tau_t)) = -(V_{t+1} - V_t). \tag{9.2}$$

In market-risk management we often work with valuation models (such as Black–Scholes) where calendar time is measured in years and interest rates and volatilities are quoted on an annualized basis. In this case, if we are interested in daily losses, we set $\Delta t = 1/365$ or $\Delta t \approx 1/250$; the latter convention is mainly used in markets for equity derivatives since there are approximately 250 trading days per year. The rvs V_t and V_{t+1} then represent the portfolio value on days t and $t+1$, respectively, and L_{t+1} is the loss from day t to day $t+1$.

As explained in Section 2.2.1 the value V_t is modelled as a function of time and a d-dimensional random vector $\boldsymbol{Z}_t = (Z_{t,1}, \ldots, Z_{t,d})'$ of risk factors. This procedure is referred to as *mapping*. Using the canonical units of time for the valuation model (typically years), mapping leads to an equation of the form

$$V_t = g(\tau_t, \boldsymbol{Z}_t) \tag{9.3}$$

for some measurable function $g\colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ and some vector of appropriate risk factors $\boldsymbol{Z}_t$. The choice of the function g and risk factors $\boldsymbol{Z}_t$ reflects the structure of the portfolio and also the desired level of precision in the modelling of risk.

Note that by introducing $f(t, \mathbf{Z}_t) := g(\tau_t, \mathbf{Z}_t)$ we can get the simpler version of the mapping formula used in (2.2). However, the use of the $g(\tau_t, \mathbf{Z}_t)$ notation allows us more flexibility to map positions while preserving market conventions with respect to timescale; see, for example, the mapping of a zero-coupon bond portfolio in Section 9.1.3.

Recall that the risk-factor changes $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ are given by $\boldsymbol{X}_t := \mathbf{Z}_t - \mathbf{Z}_{t-1}$. Using the mapping (9.3) the portfolio loss can be written as

$$L_{t+1} = -(g(\tau_{t+1}, \mathbf{Z}_t + \boldsymbol{X}_{t+1}) - g(\tau_t, \mathbf{Z}_t)). \tag{9.4}$$

Since $\mathbf{Z}_t$ is known at time t, the loss distribution at time t is determined by the distribution of the risk-factor change $\boldsymbol{X}_{t+1}$. We therefore introduce a new piece of notation in this chapter, namely the *loss operator* at time t, written $l_{[t]}\colon \mathbb{R}^d \to \mathbb{R}$, which maps risk-factor changes into losses. It is defined by

$$l_{[t]}(\boldsymbol{x}) := -(g(\tau_{t+1}, z_t + \boldsymbol{x}) - g(\tau_t, z_t)), \quad \boldsymbol{x} \in \mathbb{R}^d, \tag{9.5}$$

where z_t denotes the realized value of $\mathbf{Z}_t$ at time t; we obviously have $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$ at time t. The loss operator will facilitate our discussion of statistical approaches to measuring market risk in Section 9.2.

Note that, while we use lowercase z_t in (9.5) to emphasize that the loss operator is a function of known risk-factor values at time t, we will not apply this convention strictly in later examples.

9.1.2 Delta and Delta–Gamma Approximations

If the mapping function g is differentiable and Δt is relatively small, we can approximate g with a first-order Taylor series approximation

$$g(\tau_t + \Delta t, z_t + \boldsymbol{x}) \approx g(\tau_t, z_t) + g_\tau(\tau_t, z_t)\Delta t + \sum_{i=1}^{d} g_{z_i}(\tau_t, z_t)x_i, \tag{9.6}$$

where the τ subscript denotes the partial derivative with respect to the time argument of the mapping and the z_i subscripts denote the partial derivatives with respect to the risk factors. This allows us to approximate the loss operator in (9.5) by the *linear loss operator* at time t, which is given by

$$l^{\Delta}_{[t]}(\boldsymbol{x}) := -\left(g_\tau(\tau_t, z_t)\Delta t + \sum_{i=1}^{d} g_{z_i}(\tau_t, z_t)x_i\right). \tag{9.7}$$

Note that, when working with a short time horizon Δt, the term $g_\tau(\tau_t, z_t)\Delta t$ is very small and is sometimes omitted in practice.

We can also develop a second-order Taylor series, or so-called *delta–gamma*, approximation. Suppose we introduce vector notation

$$\boldsymbol{\delta}(\tau_t, z_t) = (g_{z_1}(\tau_t, z_t), \ldots, g_{z_d}(\tau_t, z_t))'$$

for the first-order partial derivatives of the mapping with respect to the risk factors. For the second-order partial derivatives let

$$\boldsymbol{\omega}(\tau_t, z_t) = (g_{z_1\tau}(\tau_t, z_t), \ldots, g_{z_d\tau}(\tau_t, z_t))'$$

denote the vector of mixed partial derivatives with respect to time and the risk factors and let $\Gamma(\tau_t, z_t)$ denote the matrix with (i, j)th element given by $g_{z_i z_j}(\tau_t, z_t)$; this matrix contains *gamma sensitivities* to individual risk factors on the diagonal and *cross gamma sensitivities* to pairs of risk factors off the diagonal. The full second-order approximation of g is

$$g(\tau_t + \Delta t, z_t + \boldsymbol{x}) \approx g(\tau_t, z_t) + g_\tau(\tau_t, z_t)\Delta t + \boldsymbol{\delta}(\tau_t, z_t)'\boldsymbol{x} + \tfrac{1}{2}(g_{\tau\tau}(\tau_t, z_t)(\Delta t)^2 + 2\boldsymbol{\omega}(\tau_t, z_t)'\boldsymbol{x}\Delta t + \boldsymbol{x}'\Gamma(\tau_t, z_t)\boldsymbol{x}). \tag{9.8}$$

In practice, we would usually omit terms of order $o(\Delta t)$ (terms that tend to zero faster than Δt). In the above expression this is the term in $(\Delta t)^2$ and, if we assume that risk factors follow a standard continuous-time financial model such as Black–Scholes or many generalizations thereof, the term in $\boldsymbol{x}\Delta t$.

To understand better why the last statement is true, consider the case of the Black–Scholes model. The log stock price at time t is given by $\ln S_t = (\mu - \frac{1}{2}\sigma^2)\tau_t + \sigma W_{\tau_t}$, where μ is the drift, σ is the volatility, $\tau_t = t(\Delta t)$ as usual and W_{τ_t} denotes Brownian motion. It follows that the risk-factor change satisfies

$$X_{t+1} = \ln\left(\frac{S_{t+1}}{S_t}\right) \sim N((\mu - \tfrac{1}{2}\sigma^2)\Delta t, \sigma^2\Delta t).$$

Clearly, $X_{t+1}/(\sigma\sqrt{\Delta t})$ converges in distribution to a standard normal variable as $\Delta t \to 0$. Risk-factor changes x in this model are therefore of order $O(\sqrt{\Delta t})$, meaning they tend to zero at the same rate as $\sqrt{\Delta t}$. It follows that the term $x\Delta t$ tends to zero at the same rate as $(\Delta t)^{3/2}$ and is therefore a term of order $o(\Delta t)$.

Omitting terms of order $o(\Delta t)$ in (9.8) leaves us with the *quadratic loss operator*

$$l^{\Delta,\Gamma}_{[t]}(\boldsymbol{x}) := -(g_\tau(\tau_t, z_t)\Delta t + \boldsymbol{\delta}(\tau_t, z_t)'\boldsymbol{x} + \tfrac{1}{2}\boldsymbol{x}'\Gamma(\tau_t, z_t)\boldsymbol{x}), \tag{9.9}$$

and this typically provides a more accurate approximation to (9.5) than the linear loss operator. In Example 9.1 below we give an application of the delta–gamma approximation (9.9).

Example 9.1 (European call option). The set-up and notation in this example are similar to those of Example 2.2 but we now consider a European call option that has been sold by a bank and *delta-hedged* to remove some of the risk. This means that the bank has bought a quantity of stock equivalent to the delta of the option so that the first-order sensitivity of the hedged position to stock price changes is 0. To simplify the analysis of risk factors, we assume that the interest rate r is constant.

Using the time-series notation in (9.1), the value of the hedged position at time t is

$$V_t = S_t h_t - C^{\mathrm{BS}}(\tau_t, S_t; r, \sigma_t, K, T), \tag{9.10}$$

where S_t and σ_t are the stock price and implied volatility at t, K is the strike price, T is the maturity and $h_t = C^{\mathrm{BS}}_S(\tau_t, S_t; r_t, \sigma_t, K, T)$ is the delta of the option. The time horizon of interest is one day and the natural time unit in the Black–Scholes formula is years, so $\Delta t = 1/250$ and $\tau_t = t/250$.

The valuation formula (9.10) is of the form (9.3) with risk factors $\boldsymbol{Z}_t = (\ln S_t, \sigma_t)'$. The linear loss operator (9.7) is given by

$$l^{\Delta}_{[t]}(\boldsymbol{x}) = C^{\mathrm{BS}}_{\tau}\Delta t + C^{\mathrm{BS}}_{\sigma}x_2,$$

since $g_{z_1}(\tau_t, \boldsymbol{z}_t) = (h_t - C^{\mathrm{BS}}_S)S_t = 0$.

Consider the situation where the time to expiry is $T - \tau_t = 1$, the strike price is 100 and the interest rate is $r = 0.02$. Moreover, assume that the current stock price is $S_t = 110$, so that the option is in the money, and the current implied volatility is $\sigma_t = 0.2$. The values of the Greeks in the Black–Scholes model may be calculated using well-known formulas (see Notes and Comments): they are $C^{\mathrm{BS}}_{\tau} \approx -4.83$ and $C^{\mathrm{BS}}_{\sigma} \approx 34.91$. Suppose we consider the effect of risk-factor changes $\boldsymbol{x} = (0.05, 0.02)'$ representing a stock return of (approximately) 5% and an increase in implied volatility of 2%. The stock return obviously makes no contribution to the linearized loss, which is given by

$$l^{\Delta}_{[t]}(\boldsymbol{x}) = C^{\mathrm{BS}}_{\tau} \cdot (1/250) + C^{\mathrm{BS}}_{\sigma} \cdot 0.02 \approx -0.019 + 0.698 = 0.679.$$

On the other hand, if we use full revaluation of the option at time $t + 1$, the loss would be given by $l_{[t]}(\boldsymbol{x}) \approx 0.812$. So, for risk-factor changes of this order of $\boldsymbol{x}$, there is a 16% underestimate involved in linearization.

To make a second-order approximation in this case we need to compute *gamma*, the second derivative C^{BS}_{SS} with respect to stock price, the second derivative $C^{\mathrm{BS}}_{\sigma\sigma}$ with respect to volatility, and the mixed derivative $C^{\mathrm{BS}}_{S\sigma}$ with respect to stock price and volatility. This gives the quadratic loss operator

$$l^{\Delta\Gamma}_{[t]}(\boldsymbol{x}) = C^{\mathrm{BS}}_{\tau}\Delta t + C^{\mathrm{BS}}_{\sigma}x_2 + \tfrac{1}{2}C^{\mathrm{BS}}_{SS}S_t^2x_1^2 + C^{\mathrm{BS}}_{S\sigma}S_tx_1x_2 + \tfrac{1}{2}C^{\mathrm{BS}}_{\sigma\sigma}x_2^2,$$

where we note that the S_t^2 and S_t factors enter the third and fourth terms because the risk factor is $\ln S_t$ rather than S_t. In the numerical example,

$$\begin{aligned} l^{\Delta\Gamma}_{[t]}(\boldsymbol{x}) &= l^{\Delta}_{[t]}(\boldsymbol{x}) + \tfrac{1}{2}C^{\mathrm{BS}}_{SS}S_t^2x_1^2 + C^{\mathrm{BS}}_{S\sigma}S_tx_1x_2 + \tfrac{1}{2}C^{\mathrm{BS}}_{\sigma\sigma}x_2^2 \\ &\approx 0.679 + 0.218 - 0.083 + 0.011 = 0.825. \end{aligned}$$

This is less than a 2% overestimate of the true loss, which is a substantially more accurate assessment of the impact of $\boldsymbol{x}$. The inclusion of the gamma of the option C^{BS}_{SS} is particularly important.

This example shows that the additional complexity of second-order approximations may often be warranted. Note, however, that delta–gamma approximations can give very poor results when applied to longer time horizons with large risk-factor changes.

9.1.3 Mapping Bond Portfolios

In this section we apply the ideas of Section 9.1.2 to the mapping of a portfolio of bonds and relate this to the classical concepts of duration and convexity in the risk management of bond portfolios.

Basic definitions for bond pricing. In standard bond pricing notation, $p(t, T)$ denotes the price at time t of a default-free zero-coupon bond with maturity T. While zero-coupon bonds of long maturities are relatively rare in practice, many other fixed-income instruments such as coupon bonds or standard swaps can be viewed as portfolios of zero-coupon bonds, and zero-coupon bonds are therefore fundamental building blocks for studying interest-rate risk. We follow a standard convention in modern interest-rate theory and normalize the face value $p(T, T)$ of the bond to 1, and we measure time in years.

The mapping $T \to p(t, T)$ for different maturities is one way of describing the so-called *term structure* of interest rates at time t. An alternative description is based on yields. The *continuously compounded yield* of a zero-coupon bond is defined to be $y(t, T) = -(1/(T-t)) \ln p(t, T)$, so that we have the relationship

$$p(t, T) = \exp(-(T-t)y(t, T)).$$

The mapping $T \mapsto y(t, T)$ is referred to as the continuously compounded *yield curve* at time t. Yields are a popular way of describing the term structure because they are comparable across different times to maturity due to the rescaling by $(T-t)$; they are generally expressed on an annualized basis.

We now consider the mapping of a portfolio of zero-coupon bonds. Note that the same mapping structure would be obtained for a single coupon bond, a portfolio of coupon bonds or any portfolio of promised cash flows at fixed future times.

Detailed mapping of a bond portfolio. We consider a portfolio of d default-free zero-coupon bonds with maturities T_i and prices $p(t, T_i)$, $1 \leqslant i \leqslant d$. By λ_i we denote the number of bonds with maturity T_i in the portfolio.

In a detailed analysis of the change in value of the bond portfolio, one takes all yields $y(t, T_i)$, $1 \leqslant i \leqslant d$, as risk factors. The value of the portfolio at time t is given by

$$V(t) = \sum_{i=1}^{d} \lambda_i p(t, T_i) = \sum_{i=1}^{d} \lambda_i \exp(-(T_i - t)y(t, T_i)). \tag{9.11}$$

Switching to a discrete-time set-up using (9.1), the mapping (9.3) of the bond portfolio can be written as

$$V_t = g(\tau_t, \mathbf{Z}_t) = \sum_{i=1}^{d} \lambda_i \exp(-(T_i - \tau_t)Z_{t,i}), \tag{9.12}$$

where $\tau_t = t(\Delta t)$, Δt is the time horizon expressed in years, and the risk factors are the yields $Z_{t,i} = y(\tau_t, T_i)$, $1 \leqslant i \leqslant d$. The risk-factor changes are the changes in yields $X_{t+1,i} = y(\tau_{t+1}, T_i) - y(\tau_t, T_i)$, $1 \leqslant i \leqslant d$.

From (9.12) the loss operator $l_{[t]}$ and its linear and quadratic approximations can easily be computed. The first derivatives of the mapping function are

$$g_\tau(\tau_t, z_t) = \sum_{i=1}^{d} \lambda_i p(\tau_t, T_i) z_{t,i},$$

$$g_{z_i}(\tau_t, z_t) = -\lambda_i (T_i - \tau_t) \exp(-(T_i - \tau_t)z_{t,i}).$$

Inserting these into (9.7) and reverting to standard bond pricing notation we obtain

$$l^{\Delta}_{[t]}(\boldsymbol{x}) = -\sum_{i=1}^{d} \lambda_i \, p(\tau_t, T_i)(y(\tau_t, T_i)\Delta t - (T_i - \tau_t)x_i), \tag{9.13}$$

where x_i represents the change in yield of the ith bond.

For the second-order approximation we need the second derivatives with respect to yields, which are

$$g_{z_i z_i}(\tau_t, \boldsymbol{z}_t) = \lambda_i (T_i - \tau_t)^2 \exp(-(T_i - \tau_t) z_{t,i})$$

and $g_{z_i z_j}(\tau_t, \boldsymbol{z}_t) = 0$ for $i \neq j$. Using standard bond pricing notation, the quadratic loss operator in (9.9) is

$$l^{\Delta\Gamma}_{[t]}(\boldsymbol{x}) = -\sum_{i=1}^{d} \lambda_i \, p(\tau_t, T_i)(y(\tau_t, T_i)\Delta t - (T_i - \tau_t)x_i + \tfrac{1}{2}(T_i - \tau_t)^2 x_i^2). \tag{9.14}$$

Relationship to duration and convexity. The approximations (9.13) and (9.14) can be interpreted in terms of the classical notions of the duration and convexity of bond portfolios. To make this connection consider a very simple model for the yield curve at time t in which

$$y(\tau_{t+1}, T_i) = y(\tau_t, T_i) + x \tag{9.15}$$

for all maturities T_i. In this model we assume that a *parallel shift in level* takes place along the entire yield curve, an assumption that is unrealistic but that is frequently made in practice.

Obviously, when (9.15) holds, the loss operators in (9.13) and (9.14) are functions of a scalar variable x (the size of the shift). We can express (9.13) in terms of the classical concept of the *duration* of a bond portfolio by writing

$$l^{\Delta}_{[t]}(x) = -V_t(A_t \Delta t - D_t x), \tag{9.16}$$

where

$$D_t := \sum_{i=1}^{d} \frac{\lambda_i \, p(\tau_t, T_i)}{V_t}(T_i - \tau_t), \qquad A_t := \sum_{i=1}^{d} \frac{\lambda_i \, p(\tau_t, T_i)}{V_t} y(\tau_t, T_i).$$

The term that interests us here is D_t, which is usually called the (Macaulay) *duration* of the bond portfolio. It is a weighted sum of the times to maturity of the different cash flows in the portfolio, the weights being proportional to the discounted values of the cash flows.

Over short time intervals the Δt term in (9.16) will be negligible and losses of value in the bond portfolio will be determined by $l_{[t]}(x) \approx v_t D_t x$, so that increases in the level of the yield curve lead to losses and decreases lead to gains (assuming all positions are long so that $\lambda_i > 0$ for all i). The duration D_t can be thought of as the bond pricing analogue of the delta of an option; to a first-order approximation, losses will be governed by D_t. Any two bond portfolios with equal value and duration will be subject to similar losses when there is a small parallel shift of the yield curve, regardless of differences in the exact composition of the portfolios.

Duration is an important tool in traditional bond-portfolio or asset-liability management. The standard duration-based strategy to manage the interest-rate risk of a bond portfolio is called *immunization*. Under this strategy an asset manager, who has a certain amount of funds to invest in various bonds and who needs to make certain known payments in the future, allocates these funds to various bonds in such a way that the duration of the overall portfolio consisting of bond investments and liabilities is equal to zero. As we have just seen, duration measures the sensitivity of the portfolio value with respect to shifts in the level of the yield curve. A zero duration therefore means that the position has been immunized against changes in level. However, the portfolio is still exposed to other types of yield-curve changes, such as changes in slope and curvature.

It is possible to get more accurate approximations for the loss in a bond portfolio by considering second-order effects. The analogue of the gamma of an option is the concept of *convexity*. Under our model (9.15) for changes in the level of yields, the expression for the quadratic loss operator in (9.14) becomes

$$l_{[t]}^{\Delta\Gamma}(x) = -V_t(A_t\Delta t - D_t x + \tfrac{1}{2}C_t x^2), \tag{9.17}$$

where

$$C_t := \sum_{i=1}^{d} \frac{\lambda_i\, p(\tau_t, T_i)}{V_t}(T_i - \tau_t)^2$$

is the convexity of the bond portfolio. The convexity is a weighted average of the squared times to maturity and is the negative of the derivative of the duration with respect to yield. Consider two portfolios (1) and (2) with identical values V_t and durations D_t. Assume that the convexity of portfolio (1) is greater than that of portfolio (2), so that $C_t^{(1)} > C_t^{(2)}$. Ignoring terms in Δt, the difference in loss operators satisfies

$$l_{[t]}^{\Delta\Gamma(1)}(x) - l_{[t]}^{\Delta\Gamma(2)}(x) \approx -\tfrac{1}{2}V_t(C_t^{(1)} - C_t^{(2)})x^2 < 0.$$

In other words, an increase in the level of yields will lead to smaller losses for portfolio (1), and a decrease in the level of yields will lead to larger gains (since $-l_{[t]}^{\Delta\Gamma(1)}(x) > -l_{[t]}^{\Delta\Gamma(2)}(x)$). For this reason portfolio managers often take steps to construct portfolios with relatively high convexity. Roughly speaking, this is done by spreading out the cash flows as much as possible (see Notes and Comments).

9.1.4 Factor Models for Bond Portfolios

For large portfolios of fixed-income instruments, such as the overall fixed-income position of a major bank, modelling changes in the yield for every cash flow maturity date becomes impractical. Moreover, the statistical task of estimating a distribution for $\boldsymbol{X}_{t+1}$ is difficult because the yields are highly dependent for different times to maturity. A pragmatic approach is therefore to build a factor model for yields that captures the main features of the evolution of the yield curve. Three-factor models of the yield curve in which the factors typically represent *level*, *slope* and *curvature* are often used in practice.

In this section we describe two different approaches to approximating the loss operator for a bond portfolio, and we show how statistical analysis techniques from the area of factor modelling are used to calibrate the approximating functions.

The approach based on the Nelson and Siegel (1987) model. The Nelson–Siegel model is usually formulated in terms of instantaneous *forward interest rates*, which are defined from prices by

$$f(t, T) = -\frac{\partial}{\partial T} \ln p(t, T).$$

These can be interpreted as representing the rates that are offered at time t for borrowing at future times T. Yields are related to forward rates by

$$y(t, T) = \frac{1}{T-t} \int_t^T f(t, u)\, \mathrm{d}u. \tag{9.18}$$

In the Nelson–Siegel approach the forward curve is modelled by

$$f(\tau_t, T) = Z_{t,1} + Z_{t,2} \exp(-\eta_t(T-\tau_t)) + Z_{t,3}\eta_t(T-\tau_t)\exp(-\eta_t(T-\tau_t)),$$

where the factors are $(Z_{t,1}, Z_{t,2}, Z_{t,3})$ and η_t is a positive rate parameter, which is chosen to give the best fit to forward rate data. The relationship (9.18) between the forward and yield curves implies that

$$y(\tau_t, T) = Z_{t,1} + k_1(T-\tau_t, \eta_t)Z_{t,2} + k_2(T-\tau_t, \eta_t)Z_{t,3}, \tag{9.19}$$

where the functions k_1 and k_2 are given by

$$k_1(s, \eta) = \frac{1-\mathrm{e}^{-\eta s}}{\eta s}, \qquad k_2(s, \eta) = k_1(s, \eta) - \mathrm{e}^{-\eta s}.$$

These functions are illustrated in Figure 9.1. We now give an economic interpretation of the factors.

Clearly, $\lim_{s\to\infty} k_1(s, \eta) = \lim_{s\to\infty} k_2(s, \eta) = 0$, while $\lim_{s\to 0} k_1(s, \eta) = 1$ and $\lim_{s\to 0} k_2(s, \eta) = 0$. It follows that $\lim_{T\to\infty} y(\tau_t, T) = Z_{t,1}$, so that the first factor is usually interpreted as a long-term level factor. $Z_{t,2}$ is interpreted as a slope factor because the difference between short-term yield and long-term yield satisfies $\lim_{T\to\tau_t} y(\tau_t, T) - \lim_{T\to\infty} y(\tau_t, T) = Z_{t,2}$; $Z_{t,3}$ has an interpretation as a curvature factor.

Using the factor model (9.19), the mapping (9.11) for the bond portfolio becomes

$$V_t = g(\tau_t, \boldsymbol{Z}_t) = \sum_{i=1}^{d} \lambda_i \exp(-(T_i - \tau_t)\boldsymbol{k}_{t,i}'\boldsymbol{Z}_t).$$

where $\boldsymbol{k}_{t,i} = (1, k_1(T_i-\tau_t, \eta_t), k_2(T_i-\tau_t, \eta_t))'$. It is then straightforward to derive the loss operator $l_{[t]}(\boldsymbol{x})$ or its linear version $l^{\Delta}_{[t]}(\boldsymbol{x})$, which, in contrast to (9.13), are functions on $\mathbb{R}^3$ rather than $\mathbb{R}^d$ (d is the number of bonds in the portfolio).

To use this method to evaluate the linear loss operator at time t, in practice we require realized values $\boldsymbol{z}_t$ for the risk factors $\boldsymbol{Z}_t$. However, we have to overcome the

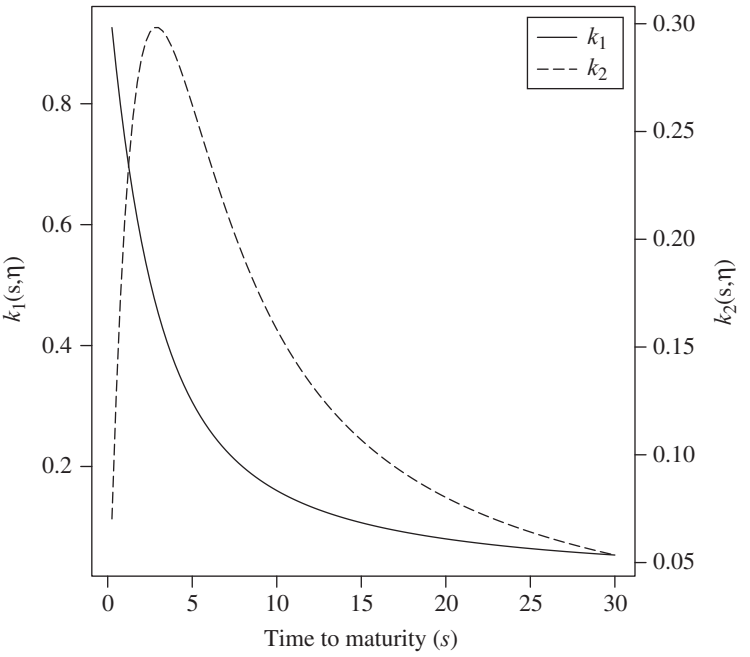

Figure 9.1. The Nelson–Siegel functions $k_1(s, \eta)$ and $k_2(s, \eta)$ for an η value of 0.623 (see also Example 9.2).

fact that the Nelson–Siegel factors $\boldsymbol{Z}_t$ are not directly observed at time t. Instead they have to be estimated from observable yield curve data.

Let us suppose that at time t we have the data vector

$$\boldsymbol{Y}_t = (y(\tau_t, \tau_t + s_1), \ldots, y(\tau_t, \tau_t + s_m))'$$

giving the yields for m different times to maturity, $s_1, \ldots, s_m$, where m is large. This is assumed to follow the factor model $\boldsymbol{Y}_t = B_t \boldsymbol{Z}_t + \boldsymbol{\varepsilon}_t$, where $B_t \in \mathbb{R}^{m\times 3}$ is the matrix with ith row consisting of $(1, k_1(s_i, \eta_t), k_2(s_i, \eta_t))$ and $\boldsymbol{\varepsilon}_t \in \mathbb{R}^m$ is an error vector. This model fits into the framework of the general factor model in (6.50).

For a given value of η_t the estimation of $\boldsymbol{Z}_t$ can be carried out as a cross-sectional regression using weighted least squares. To estimate η_t, a more complicated optimization is carried out (see Notes and Comments). We now show how the method works for real market yield data.

Example 9.2 (Nelson–Siegel factor model of yield curve). The data are daily Canadian zero-coupon bond yields for 120 different quarterly maturities ranging from 0.25 years to 30 years. They have been generated using pricing data for Government of Canada bonds and treasury bills. We model the yield curve on 8 August 2011. The estimated values are $z_{t,1} = 3.82$, $z_{t,2} = -2.75$, $z_{t,3} = -5.22$ and $\hat{\eta}_t = 0.623$. The curves $k_1(s, \eta)$ and $k_2(s, \eta)$ are therefore as shown in Figure 9.1.

Example 9.2 illustrates the estimation of the Nelson–Siegel factor model at a single time point t. We note that, to make statistical inferences about bond-portfolio

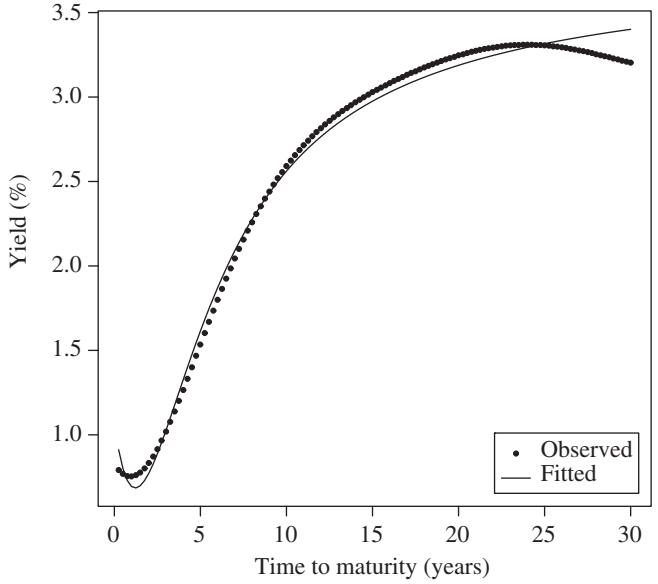

Figure 9.2. Canadian yield curve data (points) and fitted Nelson–Siegel curve for 8 August 2011 (see Example 9.2 for further details).

losses based on historical data, we would typically construct historical risk-factor time series $(\boldsymbol{Z}_u)$ from yield data $(\boldsymbol{Y}_u)$ by estimating cross-sectional regression models at all times u in some set $\{t-n+1, \ldots, t\}$. The estimated risk-factor time series $(\boldsymbol{Z}_u)$ and the corresponding risk-factor changes form the data for the statistical methods that are the subject of Section 9.2.

The approach based on principal component analysis. Another approach to building a factor model of the term structure involves the use of principal component analysis (PCA). We refer to Section 6.4 for an introduction to this method. The key difference to the Nelson–Siegel approach is that here the dimension reduction via factor modelling is applied at the level of the changes in yields rather than the yields themselves.

Let $\boldsymbol{R}_{t+1}$ denote the vector of yield changes for the bonds in the portfolio, so that $R_{t+1,i} = y(\tau_{t+1}, T_i) - y(\tau_t, T_i)$, $1 \leqslant i \leqslant d$. We recall from (6.62) that PCA can be used to construct approximate factor models of the form

$$\boldsymbol{R}_{t+1} = \boldsymbol{\mu} + \Gamma_1 \boldsymbol{X}_{t+1} + \boldsymbol{\varepsilon}_{t+1}, \tag{9.20}$$

where $\boldsymbol{X}_{t+1}$ is a p-dimensional vector of principal components ($p \ll d$), $\Gamma_1 \in \mathbb{R}^{d\times p}$ contains the corresponding loading matrix, $\boldsymbol{\mu}$ is the mean vector of $\boldsymbol{R}_{t+1}$ and $\boldsymbol{\varepsilon}_{t+1}$ is an error vector. The columns of the matrix Γ_1 consist of the first p eigenvectors (ordered by decreasing eigenvalue) of the covariance matrix of $\boldsymbol{R}_{t+1}$. The principal components $\boldsymbol{X}_{t+1}$ will form the risk-factor changes in our portfolio analysis, hence our choice of notation.

Typically, the error term is neglected and $\boldsymbol{\mu} \approx \boldsymbol{0}$, so that we make the approximation $\boldsymbol{R}_{t+1} \approx \Gamma_1 \boldsymbol{X}_{t+1}$. In the case of the linear loss operator for the bond portfolio

in (9.13), we use the approximation

$$l^{\Delta}_{[t]}(\boldsymbol{x}) = -\sum_{i=1}^{d} \lambda_i \, p(\tau_t, T_i)(y(\tau_t, T_i)\Delta t - (T_i - \tau_t)(\Gamma_1 \boldsymbol{x})_i), \tag{9.21}$$

so that a function of a p-dimensional argument $\boldsymbol{x}$ is substituted for a function of a d-dimensional argument.

To work with this function we require an estimate for the matrix Γ_1. This can be obtained from historical time-series data on yield changes by estimating sample principal components, as explained in Example 9.3.

Example 9.3 (PCA factor model of yield changes). We again analyse Canadian bond yield data as in Example 9.2. To estimate the Γ_1 matrix of principal component loadings we require longitudinal (time-series) data rather than the cross-sectional data that were used in the previous example.

We will assume for simplicity that the times to maturity $T_1 - \tau_t, \ldots, T_d - \tau_t$ of the bonds in the portfolio correspond exactly to the times to maturity $s_1, \ldots, s_d$ available in the historical data set (if not we would make an appropriate selection of the data) and that the risk-management horizon Δt is one day.

In the Canadian data set we have 2488 days of data spanning the period from 2 January 2002 to 30 December 2011 (ten full trading years); recall that each day gives rise to a data vector $\boldsymbol{Y}_u = (y(\tau_u, \tau_u + s_1), \ldots, y(\tau_u, \tau_u + s_d))'$ of yields for the different maturities. In line with (9.20) we analyse the daily returns (first differences) of these data $\boldsymbol{R}_u = \boldsymbol{Y}_u - \boldsymbol{Y}_{u-1}$ using PCA under the assumption that they form a stationary time series. (Note that a small error is incurred by analysing daily yield changes for yields with fixed times to maturity rather than fixed maturity date, but this will be neglected for the purposes of illustration.)

When we compute the variances of the sample principal components (using the same technique as for Figure 6.5), we find that the first component explains 87.0% of the variance of the data, the first two components explain 95.9%, and the first three components explain 97.5%. We choose to work with the first three principal components, meaning that we set $p = 3$. The matrix Γ_1 is estimated by G_1, a matrix whose columns are the first three eigenvectors of the sample covariance matrix. We recall from (6.63) that the complete eigenvector matrix for the sample covariance matrix is denoted by G.

The first three eigenvectors are shown graphically in Figure 9.3 and lend themselves to a standard interpretation. The first principal component has negative loadings for all maturities; the second has negative loadings up to ten years and positive loadings thereafter; the third has positive loadings for very short maturities (less than 2.5 years) and very long maturities (greater than 15 years) but negative loadings otherwise. This suggests that the first principal component can be thought of as inducing a change in the level of all yields, the second induces a change of slope and the third induces a change in the curvature of the yield curve.

We note that to make statistical inferences about bond-portfolio losses based on historical data we require a historical time series of risk-factor changes $(\boldsymbol{X}_u)$ for times

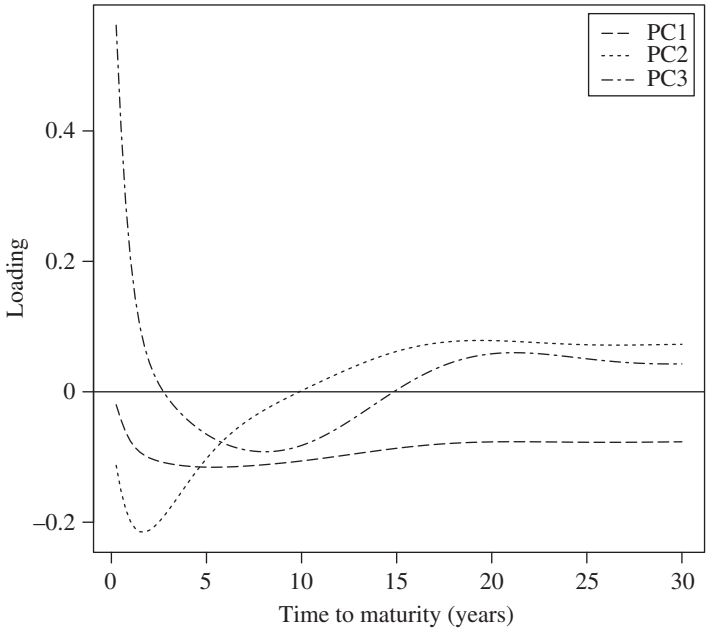

Figure 9.3. First three principal component loading vectors plotted against time to maturity. The data are daily changes in yield for Canadian zero-coupon bonds in the ten-year period 2002–11. The horizontal line at zero shows when loadings are positive or negative. See Example 9.3 for further details.

u in some set $\{t-n+1,\dots,t\}$. These data are not directly observed but are instead extracted from the time series of sample principal components $(G^{-1}(\boldsymbol{R}_u - \bar{\boldsymbol{R}}))$, where $\bar{\boldsymbol{R}}$ is the sample mean vector. The risk-factor change data $(\boldsymbol{X}_u)$ are taken to be the first p component series; these form the data for the statistical methods that are the subject of Section 9.2.

Notes and Comments

The mapping framework introduced in this section is similar to the approach pioneered by the RiskMetrics Group (see the RiskMetrics Technical Document (JPMorgan 1996) and Mina and Xiao (2001)). The mapping of positions is also discussed in Dowd (1998), Jorion (2007) and in Volume III of the Market Risk Analysis series: Alexander (2009). The latter series of four volumes is relevant to much of the material of this chapter.

The use of first- and second-order approximations to the portfolio value (the so-called delta–gamma approximation) may be found in Duffie and Pan (1997) and Rouvinez (1997) (see also Duffie and Pan 2001). Formulas for the Greeks in the Black–Scholes model may be found in a number of textbooks, including Haug (1998) and Wilmot (2000). Leoni (2014) is a very readable introduction.

Many standard finance textbooks treat interest-rate risk and bonds. For a detailed discussion of duration and its use in the management of interest-rate risk, the books

by Jarrow and Turnbull (1999) and Hull (1997) are good starting points. The construction of fixed-income portfolios with higher convexity (so-called barbell portfolios) is discussed in Tuckman and Serrat (2011).

More advanced mathematical textbooks on interest-rate modelling include Brigo and Mercurio (2006), Carmona and Tehranchi (2006) and Filipović (2009)

There are many different approaches to modelling the yield curve. We have concentrated on the parametric factor model proposed by Nelson and Siegel (1987) and further developed in Siegel and Nelson (1988) and Svensson (1994). One theoretical deficiency of the Nelson–Siegel model is that it is not consistent with no-arbitrage pricing theory (Filipović 1999). In more recent work, Christensen, Diebold and Rudebusch (2011) have proposed a model that approximates the Nelson–Siegel model but also fits into the class of three-factor, arbitrage-free affine models.

Useful information about estimating Nelson–Siegel models in practice can be found in Ferstl and Hayden (2011); we have used their R package `termstrc` to carry out the analysis in Example 9.2. Diebold and Li (2006) have developed an approach to forecasting the yield curve in which they fit cross-sectional Nelson–Siegel factor models to multivariate time series of yields for different times to maturity and then use vector autoregression (VAR) models to forecast the Nelson–Siegel factors and hence the entire yield curve. They report satisfactory results when the value of η is held constant over all cross-sectional regressions. The alternative factor-model approach based on PCA is discussed in Hull (1997) and Alexander (2001) (see examples in Section 6.2 of the latter in particular).

RiskMetrics takes a different approach to the problem of the dimension of bond portfolios. A few benchmark yields are taken for each country and a procedure is used to approximately map cash flows at days between benchmark points to the two nearest benchmark points; we refer to Section 6.2 of the RiskMetrics technical document (JPMorgan 1996) for details.

9.2 Market Risk Measurement

In this section we discuss methods used in the financial industry to estimate the loss distribution and associated risk measures for portfolios subject to market risk. In the formal framework of Section 9.1 this amounts to the problem of estimating the distribution of $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$, or a linear or quadratic approximation thereof, where $\boldsymbol{X}_{t+1}$ is the vector of risk-factor changes from time t to time $t+1$ and $l_{[t]}$ is the known loss operator function at time t.

The problem comprises two tasks: on the one hand we have the statistical problem of estimating the distribution of $\boldsymbol{X}_{t+1}$; on the other hand we have the computational or numerical problem of evaluating the distribution of $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$. To accomplish the first task, we first have to consider carefully the nature of the distribution we wish to estimate, in particular, whether we focus on the conditional or unconditional distribution of risk-factor changes.

9.2.1 Conditional and Unconditional Loss Distributions

Generally, in market-risk measurement, it is natural to compute conditional measures of risk based on the most recent available information about financial markets. In this case, the task is to estimate $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$, the conditional distribution of risk-factor changes, given $\mathcal{F}_t$, the sigma algebra representing the available information at time t. In most cases, $\mathcal{F}_t$ is given by $\mathcal{F}_t = \sigma(\{\boldsymbol{X}_s : s \leqslant t\})$, the sigma algebra generated by past and present risk-factor changes up to and including time t. The *conditional loss distribution* is the distribution of the loss operator $l_{[t]}(\cdot)$ under $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$, that is, the distribution with df $F_{L_{t+1}|\mathcal{F}_t}(l) = P(l_{[t]}(\boldsymbol{X}_{t+1}) \leqslant l \mid \mathcal{F}_t)$.

While the conditional approach is very natural in market-risk measurement, it may not always yield a prudent assessment of risk. If we are in the middle of a quiet period on financial markets, we may underestimate the possibility of extreme market losses, which can result in an overoptimistic view of a firm's capital adequacy. For this reason it can be informative to compute *unconditional* loss distributions based on assumptions of stationary behaviour over longer time windows that (ideally) contain previous episodes of market volatility.

In the unconditional approach we make the assumption that the process of risk-factor changes $(\boldsymbol{X}_s)_{s\leqslant t}$ forms a stationary multivariate time series. We recall the definition of a stationary univariate time series from Section 4.1.1; the multivariate definition is given in Section 14.1. We estimate the stationary distribution $F_{\boldsymbol{X}}$ of the time series and then evaluate the unconditional loss distribution of $l_{[t]}(\boldsymbol{X})$, where $\boldsymbol{X}$ represents a generic random vector in $\mathbb{R}^d$ with df $F_{\boldsymbol{X}}$. The unconditional loss distribution is thus the distribution of the loss operator $l_{[t]}(\cdot)$ under $F_{\boldsymbol{X}}$.

If the risk-factor changes form an independent and identically distributed (iid) series, we obviously have $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t} = F_{\boldsymbol{X}}$, so that the conditional and unconditional approaches coincide. However, in Section 3.1.1 we have argued that many types of risk-factor-change data show volatility clustering, which is inconsistent with iid behaviour. In the stationary models that are used to account for such behaviour, $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$ is not generally equal to the stationary distribution $F_{\boldsymbol{X}}$. An important example is provided by the popular models from the GARCH family. In Section 4.2.1 we observed that a simple stationary ARCH(1) model with a conditional normal distribution has a leptokurtic stationary distribution, i.e. a non-normal distribution with heavier tails. This is also true of more complicated univariate and multivariate GARCH models.

Since the financial crisis of 2007–9, regulators have called for regular estimates of VaR to be supplemented by *stressed VaR* estimates (Basel Committee on Banking Supervision 2013a). Firms are required to estimate VaR using historical data from stress periods in the financial markets (such as 2008). Stressed VaR calculations have more to do with the choice of historical data than with the choice of conditional or unconditional distribution. In principle, a stressed VaR estimate can be computed using either approach. The key point is that historical time-series data from stress periods are substituted for the up-to-date time-series data from which regular VaR estimates are made.

9.2.2 Variance–Covariance Method

This method was originally pioneered by JPMorgan's RiskMetrics group in the early 1990s. Although used by a minority of banks today, it remains an important contribution to the development of the methodology for market-risk measurement. As mentioned in Section 2.2.3, it is an example of an analytical method in which the linearized loss distribution has a known form and estimates of VaR and expected shortfall can be computed with simple formulas.

In the variance–covariance method we assume that the conditional distribution of risk-factor changes $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$ is a multivariate normal distribution with mean vector $\boldsymbol{\mu}_{t+1}$ and covariance matrix Σ_{t+1}. In other words, we assume that, given $\mathcal{F}_t$, $\boldsymbol{X}_{t+1} \sim N_d(\boldsymbol{\mu}_{t+1}, \Sigma_{t+1})$, where $\boldsymbol{\mu}_{t+1}$ and Σ_{t+1} are $\mathcal{F}_t$-measurable.

The estimation of $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$ can be carried out in a number of ways. We can fit a (multivariate) time-series model to historical data $\boldsymbol{X}_{t-n+1}, \ldots, \boldsymbol{X}_t$ and use the fitted model to derive estimates of $\boldsymbol{\mu}_{t+1}$ and Σ_{t+1}. In Section 4.2.5 we explained the procedure for univariate GARCH or ARMA–GARCH models (see Examples 4.25 and 4.26 in particular). The same idea carries over to the multivariate GARCH models of Section 14.2 (see Section 14.2.6 in particular).

Alternatively, and more straightforwardly, the model-free exponentially weighted moving-average (EWMA) procedure suggested by the RiskMetrics group can be used. The univariate version of this technique was presented in Section 4.2.5 and the multivariate version is a simple extension of the idea. Let us suppose that we work in the context of a multivariate model with conditional mean $\boldsymbol{\mu}_t = E(\boldsymbol{X}_t \mid \mathcal{F}_{t-1}) = \mathbf{0}$. The conditional covariance matrix Σ_{t+1} is estimated recursively by

$$\hat{\Sigma}_{t+1} = \theta \boldsymbol{X}_t \boldsymbol{X}_t' + (1-\theta)\hat{\Sigma}_t, \tag{9.22}$$

where θ is a small positive number (typically of the order $\theta \approx 0.04$). The estimator (9.22) takes the form of a weighted sum of the estimate of Σ_t calculated at time $t-1$ and a term $\boldsymbol{X}_t \boldsymbol{X}_t'$ that satisfies $E(\boldsymbol{X}_t \boldsymbol{X}_t' \mid \mathcal{F}_{t-1}) = \Sigma_t$. The interpretation is that the estimate at time t is obtained by perturbing the estimate at time $t-1$ by a term that reacts to the "latest information" about the joint variability of risk-factor changes.

For n large we can calculate that

$$\hat{\Sigma}_{t+1} \approx \theta \sum_{i=0}^{n-1} (1-\theta)^i \boldsymbol{X}_{t-i} \boldsymbol{X}_{t-i}'.$$

This means that, after the EWMA procedure has been running for a while, the influence of starting values for the conditional covariance matrix is negligible and estimates are effectively weighted sums of the matrices $\boldsymbol{X}_t \boldsymbol{X}_t'$ where the weights decay exponentially. These estimates are usually quite close to estimates derived by formal multivariate GARCH modelling. The method can be refined by relaxing the assumption that the conditional mean satisfies $\boldsymbol{\mu}_t = \mathbf{0}$ and including an estimate of $\boldsymbol{\mu}_t$ obtained by exponential smoothing to get the updating equation

$$\hat{\Sigma}_{t+1} = \theta(\boldsymbol{X}_t - \hat{\boldsymbol{\mu}}_t)(\boldsymbol{X}_t - \hat{\boldsymbol{\mu}}_t)' + (1-\theta)\hat{\Sigma}_t.$$

The second critical assumption in the variance–covariance method is that the linear loss operator (9.7) for the portfolio in question is a sufficiently accurate approximation of the actual loss operator (9.5). The linear loss operator is a function of the form

$$l^{\Delta}_{[t]}(\boldsymbol{x}) = -(c_t + \boldsymbol{b}_t'\boldsymbol{x}) \tag{9.23}$$

for some constant c_t and constant vector $\boldsymbol{b}_t$, which are known to us at time t. We have seen a number of examples including

- the stock portfolio of Example 2.1, where the loss operator takes the form $l^{\Delta}_{[t]}(\boldsymbol{x}) = -v_t \boldsymbol{w}_t'\boldsymbol{x}$ and $\boldsymbol{w}_t$ is the vector of portfolio weights at time t;
- the European call option of Example 2.2; and
- the zero-coupon bond portfolio with linear loss operator given by (9.13).

An important property of the multivariate normal is that a linear function (9.23) of a normal vector must have a univariate normal distribution, as discussed in Section 6.1.3. From (6.13) we infer that, conditional on $\mathcal{F}_t$,

$$L^{\Delta}_{t+1} = l^{\Delta}_{[t]}(\boldsymbol{X}_{t+1}) \sim N(-c_t - \boldsymbol{b}_t'\boldsymbol{\mu}_{t+1}, \boldsymbol{b}_t'\Sigma_{t+1}\boldsymbol{b}_t). \tag{9.24}$$

VaR and expected shortfall may be easily calculated for the normal loss distribution in (9.24). For VaR we use formula (2.18) in Example 2.11. For expected shortfall we use formula (2.24) in Example 2.14.

Weaknesses of the method and extensions. The variance–covariance method offers a simple analytical solution to the risk-measurement problem but this convenience is achieved at the cost of two crude simplifying assumptions. First, linearization may not always offer a good approximation of the relationship between the true loss distribution and the risk-factor changes, particularly for derivative portfolios and longer time intervals. Second, the assumption of normality is unlikely to be realistic for the conditional distribution of the risk-factor changes for shorter-interval data such as daily data and weekly data. Another way of putting this is to say that the *innovation distribution* in a suitable time-series model of such data is generally heavier tailed than normal (see Example 4.24).

The convenience of the variance–covariance method relies on the fact that a linear combination of a multivariate Gaussian vector has a univariate Gaussian distribution. However, we have seen in Chapter 6 that there are other multivariate distribution families that are *closed under linear operations*, and variance–covariance methods can also be developed for these. Examples include multivariate t distributions and multivariate *generalized hyperbolic* distributions (see, in particular, Proposition 6.13 and (6.45)).

For example, suppose we model risk-factor changes in such a way that the conditional distribution is a multivariate t distribution; in other words, assume that $\boldsymbol{X}_{t+1} \mid \mathcal{F}_t \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$, where this notation was explained in Section 6.4 (see Example 6.7). Then, conditional on $\mathcal{F}_t$, we get from (9.23) that

$$L^{\Delta}_{t+1} = l^{\Delta}_{[t]}(\boldsymbol{X}_{t+1}) \sim t(\nu, -c_t - \boldsymbol{b}_t'\boldsymbol{\mu}, \boldsymbol{b}_t'\tilde{\Sigma}\boldsymbol{b}_t), \tag{9.25}$$

and risk measures can be calculated using (2.19) and (2.25) in Example 2.14.

9.2.3 Historical Simulation

Historical simulation is by far the most popular method used by banks for the trading book; Pérignon and Smith (2010) report that 73% of US and international commercial banks that disclose their methodology use this method. Moreover, many of these firms use historical simulation in a simple unconditional manner, as will be explained in this section. In Section 9.2.4 we discuss different approaches to adapting historical simulation to give conditional measures of risk that take account of changing market volatility.

Instead of estimating the distribution of $l_{[t]}(\boldsymbol{X}_{t+1})$ under some explicit parametric model for $\boldsymbol{X}_{t+1}$, the historical-simulation method can be thought of as estimating the distribution of the loss operator under the *empirical distribution* of data $\boldsymbol{X}_{t-n+1}, \ldots, \boldsymbol{X}_t$. The method can be concisely described using the loss-operator notation; we construct a univariate data set by applying the operator to each of our historical observations of the risk-factor change vector to get a set of historically simulated losses:

$$\{\tilde{L}_s = l_{[t]}(\boldsymbol{X}_s) \colon s = t - n + 1, \ldots, t\}. \tag{9.26}$$

The values $\tilde{L}_s$ show what would happen to the current portfolio if the risk-factor changes on day s were to recur. We make inferences about the loss distribution and risk measures using these historically simulated loss data.

If we assume for a moment that the risk-factor changes are iid with df $F_{\boldsymbol{X}}$ and write $F_n(l)$ for the empirical df of the data $\tilde{L}_{t-n+1}, \ldots, \tilde{L}_t$, then we may use the strong law of large numbers to show that, as $n \to \infty$,

$$\begin{aligned} F_n(l) = \frac{1}{n} \sum_{s=t-n+1}^{t} I_{\{\tilde{L}_s \leqslant l\}} = \frac{1}{n} \sum_{s=t-n+1}^{t} I_{\{l_{[t]}(\boldsymbol{X}_s) \leqslant l\}} \\ \to P(l_{[t]}(\boldsymbol{X}) \leqslant l) = F_L(l), \end{aligned}$$

where $\boldsymbol{X}$ is a generic vector of risk-factor changes with distribution $F_{\boldsymbol{X}}$ and where $L = l_{[t]}(\boldsymbol{X})$. Thus $F_n(l)$ is a consistent estimator of the df of $l_{[t]}(\boldsymbol{X})$ under $F_{\boldsymbol{X}}$.

The same conclusion will also apply for many strictly stationary time-series models, such as GARCH processes, under a suitable adaptation of the strong law of large numbers. Since the empirical df of the historically simulated loss data estimates the distribution of $l_{[t]}(\boldsymbol{X})$ under $F_{\boldsymbol{X}}$, historical simulation in its basic form is an unconditional method.

In practice, there are various ways we can use the historically simulated loss data. It is common to estimate VaR using the method of *empirical quantile estimation*, whereby theoretical quantiles of the loss distribution are estimated by *sample quantiles* of the data. As an alternative, the EVT-based methods of Section 5.2 can also be used to derive parametric estimates of the tail of the loss distribution. Further discussion of these topics is deferred to Section 9.2.6.

Strengths and weaknesses of the method. The historical-simulation method has obvious attractions: it is easy to implement and reduces the risk-measure estimation

problem to a one-dimensional problem; no statistical estimation of the multivariate distribution of $\boldsymbol{X}$ is necessary, and no assumptions about the dependence structure of risk-factor changes are made; we usually work with the original loss operator and not a linearized approximation.

However, as we have observed, it is an unconditional method and is therefore unsuited to giving dynamic, conditional measures of risk that capture the volatile nature of risk in the trading book. Advocates of historical simulation often find a virtue in the fact that it gives more stable, less volatile estimates of risk than a conditional method. But it is more prudent to separate the statistical or econometric problem of accurately estimating risk measures from the regulatory problem of imposing more stable capital requirements; to some extent, stability is behind the sixty-day smoothing used in (2.20). For this reason we look at dynamic extensions of historical simulation in Section 9.2.4.

Another issue is that the success of the approach is dependent on our ability to collect sufficient quantities of relevant, synchronized data for all risk factors. Whenever there are gaps in the risk-factor history, or whenever new risk factors are introduced into the modelling, there may be problems filling the gaps and completing the historical record. These problems will tend to reduce the effective value of n and mean that empirical estimates of VaR and expected shortfall have very poor accuracy. Ideally we want n to be fairly large, since the method is an unconditional method and we want a number of extreme scenarios in the historical record to provide more informative estimates of the tail of the loss distribution.

The method has been described as being like "driving a car while looking through the rear view mirror", a deficiency that is shared to an extent by all purely statistical procedures. It is for this reason that the stressed VaR calculations mentioned in Section 9.2.1 were introduced by regulators.

Finally, although the historical-simulation method can be easily described, it may prove difficult to implement efficiently for large portfolios of derivative instruments. Computing the historically simulated losses in (9.26) involves what practitioners refer to as *full revaluation* of the portfolio under each of the historical scenarios $\boldsymbol{X}_s$, and this may be computationally costly. To get round the problem of full revaluation in producing the simulated losses in (9.26), we can consider substituting the quadratic loss operator $l_{[t]}^{\Delta\Gamma}$ for the loss operator $l_{[t]}$ and working with second-order approximations to the losses. This means that only the risk factor sensitivities are required, and it is these that are often routinely calculated for hedging purposes anyway.

9.2.4 Dynamic Historical Simulation

In this section we present two approaches to incorporating volatility forecasting into historical simulation: a univariate approach based on univariate volatility prediction using the kind of models presented in Chapter 4, and a multivariate approach.

A univariate approach to dynamic historical simulation. For a given loss operator $l_{[t]}$ at time t, we recall the construction of the historical simulation data $\{\tilde{L}_s = l_{[t]}(\boldsymbol{X}_s) \colon s = t-n+1, \ldots, t\}$ in (9.26). We assume that these are realizations from

a stationary univariate stochastic process $(\tilde{L}_s)$ obtained by applying the function $l_{[t]}\colon \mathbb{R}^d \to \mathbb{R}$ to a stationary multivariate process of risk-factor changes $(\boldsymbol{X}_s)$, and we also assume that $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$ is the next random variable in this process.

Moreover, we assume that the stationary process $(\tilde{L}_s)$ satisfies, for all s, equations of the form $\tilde{L}_s = \mu_s + \sigma_s Z_s$, where μ_s is an $\mathcal{F}_{s-1}$-measurable conditional mean term, σ_s is an $\mathcal{F}_{s-1}$-measurable volatility, and the (Z_s) are SWN(0, 1) innovations with df F_Z. An example of a model satisfying these assumptions would be an ARMA process with GARCH errors as defined in Section 4.2.3. As shown in Section 4.2.5, we can derive simple formulas for the VaR and expected shortfall of the conditional loss distribution $F_{L_{t+1}|\mathcal{F}_t}$ under these assumptions.

Writing VaR_α^t for the α-quantile of $F_{L_{t+1}|\mathcal{F}_t}$ and ES_α^t for the corresponding expected shortfall, we obtain

$$\mathrm{VaR}_\alpha^t = \mu_{t+1} + \sigma_{t+1} q_\alpha(Z), \qquad \mathrm{ES}_\alpha^t = \mu_{t+1} + \sigma_{t+1}\,\mathrm{ES}_\alpha(Z), \tag{9.27}$$

where Z is a generic rv with the df F_Z.

To estimate the risk measures in (9.27), we require estimates of μ_{t+1} and σ_{t+1} and estimates of the quantile and expected shortfall of the innovation df F_Z. In a model with Gaussian innovations the latter need not be estimated and are simply $q_\alpha(Z) = \Phi^{-1}(\alpha)$ and $\mathrm{ES}_\alpha(Z) = \phi(\Phi^{-1}(\alpha))/(1-\alpha)$, where the latter formula was derived in Example 2.14. In a model with non-Gaussian innovations, $q_\alpha(Z)$ and $\mathrm{ES}_\alpha(Z)$ depend on any further parameters of the innovation distribution. For example, we might assume (scaled) t innovations; in this case, the quantile and expected shortfall of a standard univariate t distribution (the latter given in (2.25)) would have to be scaled by the factor $\sqrt{(\nu-2)/\nu}$ to take account of the fact that the innovation distribution is assumed to have variance 1.

We now give a number of possible estimation strategies. In all cases the data are the historical simulation data $\tilde{L}_{t-n+1}, \ldots, \tilde{L}_t$.

(1) Fit an ARMA–GARCH model with an appropriate innovation distribution to the data by the ML method and use the prediction methodology discussed in Section 4.2.5 to estimate σ_{t+1} and μ_{t+1}. Any further parameters of the innovation distribution can be estimated simultaneously in the model fitting.

For example, suppose we use an AR(1)–GARCH(1,1) model, which, according to Definition 4.22, takes the form

$$\begin{aligned}
\tilde{L}_s &= \mu_s + \sigma_s Z_s,\\
\mu_s &= \mu + \phi_1(\tilde{L}_{s-1} - \mu_{s-1}),\\
\sigma_s^2 &= \alpha_0 + \alpha_1(\tilde{L}_{s-1} - \mu_{s-1})^2 + \beta_1\sigma_{s-1}^2
\end{aligned}$$

at any time s. The conditional mean μ_{t+1} and standard deviation σ_{t+1} are then estimated recursively by

$$\begin{aligned}
\hat{\mu}_{t+1} &= \hat{\mu} + \hat{\phi}_1(\tilde{L}_t - \hat{\mu}_t),\\
\hat{\sigma}_{t+1} &= \sqrt{\hat{\alpha}_0 + \hat{\alpha}_1(\tilde{L}_t - \hat{\mu}_t)^2 + \hat{\beta}_1\hat{\sigma}_t^2},
\end{aligned}$$

where ML estimates of the parameters of the AR(1)–GARCH(1,1) model are denoted using hats.

(2) Fit an ARMA–GARCH model by QML (see Section 4.2.4) and use prediction methodology as in strategy (1) to estimate σ_{t+1} and μ_{t+1}. In a separate second step use the model residuals to find estimates of $q_\alpha(Z)$ and $\mathrm{ES}_\alpha(Z)$. As for the basic historical-simulation method, this can be achieved using simple empirical estimates of quantiles and expected shortfalls or semi-parametric estimators based on EVT (see Section 9.2.6).

(3) Use the univariate EWMA procedure (see Section 4.2.5) to estimate

$$\sigma_{t-n+1}, \ldots, \sigma_t, \sigma_{t+1}.$$

The conditional mean terms $\mu_{t-n+1}, \ldots, \mu_t, \mu_{t+1}$ could also be estimated by exponential smoothing but it is easier to set them equal to zero, as they are likely to be very small. Standardize each of the historical simulation losses $\tilde{L}_{t-n+1}, \ldots, \tilde{L}_t$ by dividing by the EWMA volatility estimates $\hat{\sigma}_{t-n+1}, \ldots, \hat{\sigma}_t$. This yields a set of residuals, from which the innovation distribution F_Z can be estimated as in strategy (2).

These procedures often work well in practice but there can be some loss of information involved with applying volatility modelling at the level of the historically simulated data rather than at the level of the risk-factor changes themselves. We now present a second method that incorporates volatility at the level of the individual risk factors. While the method is more computationally intensive, it can result in more accurate estimates of risk measures.

A multivariate approach to dynamic historical simulation. In this method we work with risk-factor change data $\boldsymbol{X}_{t-n+1}, \ldots, \boldsymbol{X}_t$ and assume that the data vectors are realizations from a multivariate time-series process $(\boldsymbol{X}_s)$ that satisfies equations of the form

$$\boldsymbol{X}_s = \boldsymbol{\mu}_s + \Delta_s \boldsymbol{Z}_s, \quad \Delta_s = \mathrm{diag}(\sigma_{s,1}, \ldots, \sigma_{s,d}),$$

where $(\boldsymbol{\mu}_s)$ is a process of vectors and (Δ_s) a process of diagonal matrices such that $\mu_{s,1}, \ldots, \mu_{s,d}, \sigma_{s,1}, \ldots, \sigma_{s,d}$ are all $\mathcal{F}_{s-1}$-measurable and $(\boldsymbol{Z}_s) \sim \mathrm{SWN}(\boldsymbol{0}, P)$ for some correlation matrix P (in other words, the $\boldsymbol{Z}_s$ are iid random vectors whose covariance matrix is the correlation matrix P). Under these assumptions, $E(X_{s,k} \mid \mathcal{F}_{s-1}) = \mu_{s,k}$ and $\mathrm{var}(X_{s,k} \mid \mathcal{F}_{s-1}) = \sigma^2_{s,k}$, so the vector $\boldsymbol{\mu}_s$ contains the conditional means and the matrix Δ_s contains the volatilities of the component series at time s. An example of a model that fits into this framework is the CCC–GARCH (constant conditional correlation) process (see Definition 14.11).

In this context we may use multivariate dynamic historical simulation. The key idea of the method is to apply historical simulation to the unobserved innovations $(\boldsymbol{Z}_s)$ rather than the observed data $(\boldsymbol{X}_s)$ (as in standard historical simulation). The first step is to compute estimates $\{\hat{\boldsymbol{\mu}}_s : s = t-n+1, \ldots, t\}$ and $\{\hat{\Delta}_s : s = t-n+1, \ldots, t\}$ of the conditional mean vectors and volatility matrices.

This can be achieved by fitting univariate time-series models of ARMA–GARCH type to each of the component series in turn; alternatively, we can use the univariate EWMA approach for each series. In either case we also use prediction methodology to obtain estimates of $\hat{\Delta}_{t+1}$, the volatility matrix in the next time period, and (if desired) $\hat{\boldsymbol{\mu}}_{t+1}$, the conditional mean vector.

In the second step we construct residuals

$$\{\hat{\boldsymbol{Z}}_s = \hat{\Delta}_s^{-1}(\boldsymbol{X}_s - \hat{\boldsymbol{\mu}}_s) \colon s = t-n+1, \ldots, t\}$$

and treat these as "observations" of the unobserved innovations. To make statistical inferences about the distribution of $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1}) = l_{[t]}(\boldsymbol{\mu}_{t+1} + \Delta_{t+1}\boldsymbol{Z}_{t+1})$ given $\mathcal{F}_t$ we construct the data set

$$\{\tilde{L}_s = l_{[t]}(\hat{\boldsymbol{\mu}}_{t+1} + \hat{\Delta}_{t+1}\hat{\boldsymbol{Z}}_s) \colon s = t-n+1, \ldots, t\}. \tag{9.28}$$

To estimate VaR (or expected shortfall) we can apply simple empirical estimators or EVT-based methods directly to these data (see Section 9.2.6 for more details).

9.2.5 Monte Carlo

The Monte Carlo method is a rather general name for any approach to risk measurement that involves the simulation of an explicit parametric model for risk-factor changes. Many banks report that they use a Monte Carlo method to compute measures of market risk in the trading book (Pérignon and Smith 2010). However, the Monte Carlo method only offers a solution to the problem of evaluating the distribution of $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$ under a given model for $\boldsymbol{X}_{t+1}$. It does not solve the statistical problem of finding a suitable model for $\boldsymbol{X}_{t+1}$.

In the market-risk context let us assume that we have estimated a time-series model for historical risk-factor change data $\boldsymbol{X}_{t-n+1}, \ldots, \boldsymbol{X}_t$ and that this is a model from which we can readily simulate. We use the model to generate m independent realizations $\tilde{\boldsymbol{X}}_{t+1}^{(1)}, \ldots, \tilde{\boldsymbol{X}}_{t+1}^{(m)}$ from the estimated conditional distribution of risk-factor changes $\hat{F}_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$.

In a similar fashion to the historical-simulation method, we apply the loss operator to these simulated vectors to obtain simulated realizations $\{\tilde{L}_{t+1}^{(i)} = l_{[t]}(\tilde{\boldsymbol{X}}_{t+1}^{(i)}) \colon i = 1, \ldots, m\}$ from the estimated conditional loss distribution $\hat{F}_{L_{t+1}|\mathcal{F}_t}$. As for the historical-simulation method, the simulated loss data from the Monte Carlo method are used to estimate VaR and expected shortfall, e.g. by using simple empirical estimators or EVT-based methods (see Section 9.2.6 for more details).

Note that the use of Monte Carlo means that we are free to choose the number of replications m ourselves, within the obvious constraints of computation time. Generally, m can be chosen to be much larger than n (the number of data) so we obtain more accuracy in empirical VaR and expected shortfall estimates than is possible in the case of historical simulation.

Weaknesses of the method. As we have already remarked, the method does not solve the problem of finding a multivariate model for $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$ and any results that are obtained will only be as good as the model that is used.

For large portfolios the computational cost of the Monte Carlo approach can be considerable, as every simulation ideally requires the full revaluation of the portfolio to compute the loss operator. This is particularly problematic if the portfolio contains many derivatives that cannot be priced in closed form. The problem of computational cost is even more relevant to the Monte Carlo method than it is to the historical-simulation method, because we typically choose larger numbers of scenarios m for risk-factor changes in the Monte Carlo method. If second-order sensitivities are available, the loss operator can be replaced by the quadratic loss operator $l_{[t]}^{\Delta\Gamma}$ to reduce the computational cost. Moreover, variance-reduction techniques for evaluating tail probabilities and quantiles, such as importance sampling, can also be of help (see Notes and Comments).

9.2.6 Estimating Risk Measures

In both the historical simulation and Monte Carlo methods we estimate risk measures using simulated loss data. In this section we discuss different methods for estimating VaR and expected shortfall from a data sample. Let us suppose that we have data $L_1, \dots, L_n$ from an underlying loss distribution F_L and the aim is to estimate $\mathrm{VaR}_\alpha = q_\alpha(F_L) = F_L^{\leftarrow}(\alpha)$ or $\mathrm{ES}_\alpha = (1-\alpha)^{-1}\int_\alpha^1 q_\theta(F_L)\,\mathrm{d}\theta$.

L-estimators. These estimators take the form of linear combinations of sample order statistics, and the "L" in their name refers to *linear*. In Chapter 5 we defined the upper-order statistics $L_{1,n} \geqslant \cdots \geqslant L_{n,n}$, as is standard in extreme value theory. Many of the results concerning L-estimators are given in terms of lower-order statistics $L_{(1)} \leqslant \cdots \leqslant L_{(n)}$. Note that we can easily move between the two conventions by observing that $L_{k,n} = L_{(n-k+1)}$ for $k = 1, \dots, n$.

The simplest L-estimator of VaR is the sample quantile obtained by inverting the empirical distribution function $F_n(x) = n^{-1}\sum_{i=1}^n 1_{\{L_i \leqslant x\}}$ of the data $L_1, \dots, L_n$. It may be easily verified that the inverse of the empirical df is given by

$$F_n^{\leftarrow}(\alpha) = L_{(k)} \quad \text{for } \frac{k-1}{n} < \alpha \leqslant \frac{k}{n}.$$

We may write this more compactly as $F_n^{\leftarrow}(\alpha) = L_{(\lceil n\alpha \rceil)}$, where $\lceil x \rceil = \min\{k \in \mathbb{Z}\colon k \geqslant x\}$. This is the *ceiling* function that gives the smallest integer not less than x.

In working with order statistics we often use both the ceiling function and the *floor* function, $\lfloor x \rfloor = \max\{k \in \mathbb{Z}\colon k \leqslant x\}$, the largest integer not greater than x. It is easy to see that they are related by $\lfloor -x \rfloor = -\lceil x \rceil$. This fact, together with the relation $L_{k,n} = L_{(n-k+1)}$, allows us to write the sample quantile in terms of upper-order statistics. We have that $L_{(\lceil n\alpha \rceil)} = L_{k,n}$, where $k = n - \lceil n\alpha \rceil + 1 = \lfloor n(1-\alpha) \rfloor + 1$, giving

$$\widehat{\mathrm{VaR}}_\alpha = L_{k,n}, \quad k = \lfloor n(1-\alpha) \rfloor + 1. \tag{9.29}$$

For example, if $n = 1000$ and $\alpha = 0.995$, the estimator would be $L_{6,1000}$, the sixth largest value. For the same data and $\alpha = 0.9945$ the estimator is also $L_{6,1000}$.

Inverting the empirical distribution function yields a sample quantile function that is discontinuous in α. To obtain a continuous function in α there are a number of

alternative definitions of sample quantiles that interpolate linearly between adjacent order statistics. For example, the default method in the statistical package R estimates the α-quantile to be

$$\widehat{\mathrm{VaR}}_\alpha = \lambda_{\alpha,k,n} L_{k+1,n} + (1 - \lambda_{\alpha,k,n}) L_{k,n}, \quad k = \lceil (n-1)(1-\alpha) \rceil, \tag{9.30}$$

where the weights are given by $\lambda_{\alpha,k,n} = (n-k) - (n-1)\alpha$. If $n = 1000$ and $\alpha = 0.995$, then $k = 5$ and the estimator is $0.995 L_{6,1000} + 0.005 L_{5,1000}$. If we want to estimate the 0.9945 quantile from the same data, then $k = 6$ and the estimator becomes $0.4945 L_{7,1000} + 0.5055 L_{6,1000}$.

The estimators (9.29) and (9.30), being based on only one or two order statistics, are subject to a large variance, particularly for quantiles in the tail of the distribution and for small sample sizes.

To obtain an L-estimator of expected shortfall we recall from Section 8.2.1 the general form of the distortion risk measures, of which expected shortfall is a special case. Distortion risk measures are given by

$$\varrho(L) = \int_0^1 F_L^{\leftarrow}(u) \, \mathrm{d}D(u)$$

for convex distortion functions D on $[0, 1]$; the distortion function for expected shortfall is $D_\alpha(u) = (1-\alpha)^{-1}(u-\alpha)^+$. L-estimators for distortion risk measures may be derived by inserting the inverse of the empirical df as an estimator of $F_L^{\leftarrow}$ to obtain

$$\hat{\varrho}(L) = \int_0^1 F_n^{\leftarrow}(u) \, \mathrm{d}D(u) = \sum_{k=1}^n L_{(k)} \left(D_\alpha\left(\frac{k}{n}\right) - D_\alpha\left(\frac{k-1}{n}\right) \right).$$

In the special case of expected shortfall the estimator is

$$\begin{aligned}
\widehat{\mathrm{ES}}_\alpha &= \frac{1}{n(1-\alpha)} \sum_{k=1}^n L_{(k)} ((k - n\alpha)^+ - ((k-1) - n\alpha)^+) \\
&= \frac{1}{n(1-\alpha)} \left(\left(\sum_{k=\lceil n\alpha \rceil + 1}^n L_{(k)} \right) + (\lceil n\alpha \rceil - n\alpha) L_{(\lceil n\alpha \rceil)} \right) \\
&= \frac{1}{n(1-\alpha)} \left(\left(\sum_{k=1}^{\lfloor n(1-\alpha) \rfloor} L_{k,n} \right) + (\lceil n\alpha \rceil - n\alpha) L_{\lfloor n(1-\alpha) \rfloor + 1, n} \right).
\end{aligned}$$

The final term involving $L_{\lfloor n(1-\alpha) \rfloor + 1, n}$ may sometimes be omitted for a simpler estimator.

EVT-based estimators. Simple empirical estimates of the VaR and, especially, the expected shortfall are likely to be inaccurate when n is of modest size (say only a few years of daily data). This is a problem for historical simulation in particular. A possible solution is to use the techniques of *extreme value theory* (EVT) to provide

estimates of the tail of the loss distribution that are as faithful as possible to the most extreme data and that use parametric forms that are supported by theory. In Section 5.2.3 we presented a standard EVT method based on the generalized Pareto distribution that is useful in this context.

To use this method to estimate VaR_α and ES_α we can set a high threshold $u = L_{k+1,n}$ at the $(k+1)$-upper-order statistic and fit a GPD distribution to excess losses over u. We thereby obtain ML estimates $\hat{\xi}$ and $\hat{\beta}$ based on k exceedances of the threshold. To form a quantile estimator, the value k must satisfy $k/n > 1-\alpha$; moreover, k should be sufficiently large to give reasonably accurate estimates of the GPD parameters.

We then form the risk-measure estimates

$$\widehat{\mathrm{VaR}}_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{1-\alpha}{k/n}\right)^{-\hat{\xi}} - 1\right),$$
$$\widehat{\mathrm{ES}}_\alpha = \frac{\widehat{\mathrm{VaR}}_\alpha}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{1-\hat{\xi}}.$$

For more guidance on the choice of threshold, see Section 5.2.2; for a comparison of the EVT quantile estimates with simple empirical quantile estimates, see Section 5.2.5.

9.2.7 *Losses over Several Periods and Scaling*

In the banking context the methods we have described in previous sections are generally applied to daily risk-factor change data, and risk measures are routinely calculated for a one-day horizon. However, for regulatory capital purposes there is a requirement to calculate a 99% VaR estimate for a period of ten trading days (two weeks).

An obvious approach to this calculation is to model historical risk-factor changes over ten-day intervals using exactly the same methodology that has been discussed in this chapter. However, for a fixed amount n of historical time-series data, this results in a dramatic reduction in the precision of the statistical estimates of model parameters. For example, if we have $n = 1000$ days (just under four years) of historical data, this would give only 100 non-overlapping observations of ten-day risk-factor changes. To obtain similar accuracy to an analysis of the daily returns we would have to collect $n = 10\,000$ daily data (around thirty-eight years). It is possible to artificially preserve the value of n by the formation of overlapping risk-factor returns (a construction that is described in Section 3.1). However, this introduces new serial dependencies into the data, which complicates statistical modelling and does not lead to an obvious gain in statistical accuracy.

For these reasons most banks use a simple scaling rule, known as the *square-root-of-time rule*, to move between estimates of one-day VaR and estimates of ten-day VaR. We now look at the (limited) theoretical support for this rule and discuss an alternative Monte Carlo approach.

Scaling. For $h \in \mathbb{Z}$ and $h \geqslant 1$ suppose we denote the loss from time t over the next h periods by $L_{t+h}^{(h)}$. Arguing as in (9.3) and (9.4) we have

$$\begin{aligned} L_{t+h}^{(h)} &= -(V_{t+h} - V_t) \\ &= -(g(\tau_{t+h}, \boldsymbol{Z}_{t+h}) - g(\tau_t, \boldsymbol{Z}_t)) \\ &= -(g(\tau_{t+h}, \boldsymbol{Z}_t + \boldsymbol{X}_{t+1} + \cdots + \boldsymbol{X}_{t+h}) - g(\tau_t, \boldsymbol{Z}_t)) \\ &=: l_{[t]}^{(h)}\bigg(\sum_{i=1}^{h} \boldsymbol{X}_{t+i}\bigg), \end{aligned}$$

where $l_{[t]}^{(h)}$ represents a loss operator at time t for the h-period loss. The general question of interest is how risk measures applied to the conditional distribution of $L_{t+h}^{(h)}$ given $\mathcal{F}_t$ scale with h, and this has no simple answer except in special cases.

Note that the h-period loss operator differs from the one-period loss operator in situations where the mapping depends explicitly on time (such as derivative portfolios). For simplicity let us consider the case in which the mapping does not depend on calendar time, so that $l_{[t]}^{(h)}(\boldsymbol{x}) = l_{[t]}(\boldsymbol{x})$. The linearized form of this operator is of the form $l_{[t]}^{\Delta}(\boldsymbol{x}) = \boldsymbol{b}_t'\boldsymbol{x}$ for some vector $\boldsymbol{b}_t$ that is known at time t. We look at the simpler problem of scaling for risk measures applied to the linearized loss distribution:

$$L_{t+h}^{(h)\Delta} = l_{[t]}^{\Delta}\bigg(\sum_{i=1}^{h} \boldsymbol{X}_{t+i}\bigg) = \sum_{i=1}^{h} \boldsymbol{b}_t' \boldsymbol{X}_{t+i}. \tag{9.31}$$

The following example gives a justification for the square-root-of-time rule.

Example 9.4 (square-root-of-time scaling). Suppose the risk-factor change vectors are iid with distribution $N_d(\boldsymbol{0}, \Sigma)$. Then $\sum_{i=1}^{h} \boldsymbol{X}_{t+i} \sim N_d(\boldsymbol{0}, h\Sigma)$ and the distribution of $L_{t+h}^{(h)\Delta}$ in (9.31) satisfies $L_{t+h}^{(h)\Delta} \sim N(0, h\boldsymbol{b}_t'\Sigma\boldsymbol{b}_t)$. It then follows easily from (2.18) and (2.24) that both quantiles and expected shortfalls for this distribution scale according to the square root of time ($\sqrt{h}$). For example, writing $\mathrm{ES}_\alpha^{(h)}$ for the expected shortfall, we have

$$\mathrm{ES}_\alpha^{(h)} = \sqrt{h}\sigma \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha},$$

where $\sigma^2 = \boldsymbol{b}_t'\Sigma\boldsymbol{b}_t$. Clearly, $\mathrm{ES}_\alpha^{(h)} = \sqrt{h}\,\mathrm{ES}_\alpha^{(1)}$ and, with similar notation, $\mathrm{VaR}_\alpha^{(h)} = \sqrt{h}\,\mathrm{VaR}_\alpha^{(1)}$.

Although this scaling rule is quite commonly used in practice, empirical risk-factor change data generally support neither a Gaussian distributional assumption nor an iid assumption (see Section 3.1). Moreover, for the kinds of dynamic time-series model (such as GARCH) that are appropriate, very little is known about the scaling of risk measures for the conditional loss distribution of the h-period loss $L_{t+h}^{(h)}$ (or its linearized form).

Monte Carlo approach. It is possible to use a Monte Carlo approach to the problem of determining risk measures for the h-period conditional loss distribution. Suppose we have a time-series model for the risk-factor changes $(\boldsymbol{X}_s)_{s \leqslant t}$. We simulate future

paths of the process $\tilde{\boldsymbol{X}}_{t+1}^{(i)}, \ldots, \tilde{\boldsymbol{X}}_{t+h}^{(i)}$ for $i = 1, \ldots, m$, where m is a predetermined large number of replications. We then apply the h-period loss operator to these simulated data to obtain Monte Carlo simulated losses:

$$\{\tilde{L}_{t+h}^{(h)(i)} = l_{[t]}^{(h)}(\tilde{\boldsymbol{X}}_{t+1}^{(i)} + \cdots + \tilde{\boldsymbol{X}}_{t+h}^{(i)}) \colon i = 1, \ldots, m\}.$$

These are used to make statistical inferences about the loss distribution and associated risk measures, as described in Section 9.2.5. We can also use the Monte Carlo approach to examine the performance of square-root-of-time scaling and to experiment with alternative power laws (see Notes and Comments).

Notes and Comments

Standard methods for market risk are described in detail in Jorion (2007) and Crouhy, Galai and Mark (2001). For the variance–covariance approach using EWMA, see Mina and Xiao (2001). A useful overview of the popularity of different approaches in practice is given by Pérignon and Smith (2010). The multivariate approach to dynamic historical simulation is described by Hull and White (1998) and Barone-Adesi, Bourgoin and Giannopoulos (1998).

The book by Glasserman (2003) is an excellent general introduction to Monte Carlo simulation techniques in finance. Glasserman, Heidelberger and Shahabuddin (1999) present efficient numerical techniques (based on delta–gamma approximations and advanced simulation techniques) for estimating VaR for derivative portfolios in the presence of heavy-tailed risk factors.

For a reference on different definitions of empirical quantile estimates and their properties, see Hyndman and Fan (1996). Tsukahara (2009) describes L-estimators of distortion risk measures, which apply to the case of expected shortfall. The use of EVT to provide dynamic estimates of risk measures was introduced by McNeil and Frey (2000), who also highlight the differences between conditional and unconditional approaches. For risk-measure estimation applying EVT to a regime-switching model, see Chavez-Demoulin, Embrechts and Sardy (2014).

A useful summary of scaling results for market-risk measures may be found in Kaufmann (2004) (see also Brummelhuis and Kaufmann 2007; Embrechts, Kaufmann and Patie 2005). In these papers the message emerges that, for unconditional VaR scaling over longer time horizons, the square-root-of-time rule often works well. On the other hand, for conditional VaR scaling over short time horizons, McNeil and Frey (2000) use the Monte Carlo approach to present evidence against square-root-of-time scaling. For further comments on these and further scaling issues, see Diebold et al. (1998) and Daníelsson and de Vries (1997c).

For a practically oriented text on market-risk management see Daníelsson (2011).

9.3 Backtesting

Backtesting is the practice of evaluating risk measurement procedures by comparing out-of-sample estimates of risk measures with actual realized losses and gains. Backtesting allows us to address the question of whether a given estimation procedure

produces credible risk-measure estimates. In Section 9.2 we considered standard methods for estimating risk measures at a time t for the distribution of losses in the next period. At the end of the next period we have the opportunity to compare the risk-measure estimate with the actual realized loss. When this procedure is repeated over many time periods we can monitor the performance of methods and compare their relative performance.

In Section 9.3.1 we discuss the backtesting of VaR estimates, and in Section 9.3.2 we discuss the backtesting of expected shortfall. Section 9.3.3 examines the use of elicitability theory to construct natural scoring statistics for comparing the backtest results for different VaR estimation methods. An empirical example of backtesting is described in Section 9.3.4, and in Section 9.3.5 we briefly consider backtests of the whole estimated loss distribution.

9.3.1 Violation-Based Tests for VaR

At any time point t let VaR_α^t denote the α-quantile of the conditional loss distribution $F_{L_{t+1}|\mathcal{F}_t}$. We will refer to the event $\{L_{t+1} > \text{VaR}_\alpha^t\}$ as a VaR *violation* or *exception* and define the event indicator variable by $I_{t+1} = I_{\{L_{t+1} > \text{VaR}_\alpha^t\}}$. Assuming a continuous loss distribution, we have, by definition of the quantile, that

$$E(I_{t+1} \mid \mathcal{F}_t) = P(L_{t+1} > \text{VaR}_\alpha^t \mid \mathcal{F}_t) = 1 - \alpha, \tag{9.32}$$

so that I_{t+1} is a Bernoulli variable with event probability $1 - \alpha$. Moreover, the following lemma shows that the sequence of VaR violation indicators (I_t) forms a Bernoulli trials process, i.e. a process of iid Bernoulli random variables with event probability $1 - \alpha$.

Lemma 9.5. *Let $(Y_t)_{t\in\mathbb{Z}}$ be a sequence of Bernoulli indicator variables adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ and satisfying $E(Y_{t+1} \mid \mathcal{F}_t) = p > 0$ for all t. Then (Y_t) is a process of iid Bernoulli variables.*

Proof. The process $(Y_t - p)_{t\in\mathbb{Z}}$ has the martingale-difference property (see Definition 4.6). Moreover, $\text{var}(Y_t - p) = E(E((Y_t - p)^2 \mid \mathcal{F}_{t-1})) = p(1 - p)$ for all t. As $(Y_t - p)$ is a martingale-difference sequence with a finite variance, it is a white noise process (see Section 4.1.1). Hence (Y_t) is a white noise processes of uncorrelated variables. If two Bernoulli variables Y_t and Y_s are uncorrelated, it follows that

$$\begin{aligned} 0 = \text{cov}(Y_t, Y_s) &= E(Y_t Y_s) - E(Y_t)E(Y_s) \\ &= P(Y_t = 1, Y_s = 1) - P(Y_t = 1)P(Y_s = 1), \end{aligned}$$

which shows that they are also independent. □

There are two important consequences of the independent Bernoulli behaviour for violations. First, if we sum the violation indicators over a number of different times, we obtain binomially distributed random variables. For example, $M = \sum_{t=1}^m I_{t+1} \sim B(m, 1 - \alpha)$. Second, the spacings between consecutive violations are independent and geometrically distributed. Suppose that the event

$\{L_{t+1} > \mathrm{VaR}_\alpha^t\}$ occurs for times $t \in \{T_1, \ldots, T_M\}$, and let $T_0 = 0$. Then the spacings $S_j = T_j - T_{j-1}$ will be independent geometrically distributed random variables with mean $1/(1-\alpha)$, so that $P(S_j = k) = \alpha^{k-1}(1-\alpha)$ for $k \in \mathbb{N}$. Both of these properties can be tested on empirical data.

Suppose that we now estimate VaR_α^t based on information available up to time t, and we denote our estimate by $\widehat{\mathrm{VaR}}_\alpha^t$. The empirical violation indicator variable

$$\hat{I}_{t+1} = I_{\{L_{t+1} > \widehat{\mathrm{VaR}}_\alpha^t\}}$$

represents a one-step, out-of-sample comparison, in which we compare the actual realized value L_{t+1} with our VaR estimate made at time t.

Under the null hypothesis that our estimation method is accurate, in the sense that $E(\hat{I}_{t+1} \mid \mathcal{F}_t) = 1 - \alpha$ at the time points $t = 1, \ldots, m$, the sequence of empirical indicator variables $(\hat{I}_{t+1})_{1 \leqslant t \leqslant m}$ will then form a realization from a Bernoulli trials process with event probability $1 - \alpha$. For example, the quantity $\sum_{t=1}^m \hat{I}_{t+1}$ should behave like a realization from a $B(m, 1-\alpha)$ distribution, and this hypothesis can be easily addressed with a binomial test. There are a number of varieties of binomial test; in a two-sided score test we compute the statistic

$$Z_m = \frac{\sum_{t=1}^m \hat{I}_{t+1} - m(1-\alpha)}{\sqrt{m\alpha(1-\alpha)}} \tag{9.33}$$

and reject the hypothesis of Bernoulli behaviour at the 5% level if $|Z_m| > \Phi^{-1}(0.975)$. Rejection would suggest either systematic underestimation or overestimation of VaR (see Notes and Comments for further references concerning binomial tests).

To check the independence of violations we can construct a test of the geometric hypothesis. Since violations should be rare events with probability $(1-\alpha) \leqslant 0.05$, it proves easier to use the fact that a discrete-time Bernoulli process for rare events can be approximated by a continuous-time Poisson process and that the discrete geometric distribution for the event spacings can be approximated by a continuous exponential distribution.

To be precise let us suppose that the time interval $[t, t+1]$ in discrete time has length Δt in the chosen unit of continuous time. For example, if $[t, t+1]$ represents a trading day, then $\Delta t = 1$ if time is measured in days and $\Delta t = 1/250$ if time is measured in years. If the Bernoulli rare event probability is $1 - \alpha$, then the approximating Poisson process has rate $\lambda = (1-\alpha)\Delta t$ and the approximating exponential distribution has parameter λ and mean $1/\lambda$.

The exponential hypothesis can be tested using a Q–Q plot of the spacings data against the quantiles of a standard exponential reference distribution, similar to the situation discussed in Section 5.3.2. Alternatively, Christoffersen and Pelletier (2004) have proposed a likelihood ratio test of the hypothesis of exponential spacings against a more general Weibull alternative. In the exponential model the so-called hazard function of an event is constant, but the Weibull distribution can model an event clustering phenomenon whereby the hazard function is initially high after an event takes place and then decreases (see Section 10.4.1 for more discussion of hazard rate models).

In Section 9.2.7 we also discussed VaR estimates for the h-period loss distribution. To use the tests described above on h-period estimates we would have to base our backtests on non-overlapping periods. For example, if we calculated two-week VaRs, we could make a comparison of the VaR estimate and the realized loss every two weeks, which would clearly lead to a relatively small amount of violation data with which to monitor the performance of the model. It is also possible to look at overlapping periods, e.g. by recording the violation indicator value every day for the loss incurred over the previous two weeks. However, this would create a series of dependent Bernoulli trials for which formal inference is difficult.

9.3.2 Violation-Based Tests for Expected Shortfall

It is also possible to use information about the magnitudes of VaR violations to backtest estimates of expected shortfall. Let ES_α^t denote the expected shortfall of the conditional loss distribution $F_{L_{t+1}|\mathcal{F}_t}$, and define a *violation residual* by

$$K_{t+1} = \left(\frac{L_{t+1} - \mathrm{ES}_\alpha^t}{\mathrm{ES}_\alpha^t - \mu_{t+1}}\right) I_{\{L_{t+1} > \mathrm{VaR}_\alpha^t\}}, \tag{9.34}$$

where $\mu_{t+1} = E(L_{t+1} \mid \mathcal{F}_t)$. In the event that there is a VaR violation $\{L_{t+1} > \mathrm{VaR}_\alpha^t\}$, the violation residual K_{t+1} compares the actual size of the violation L_{t+1} with its expected size conditional on information up to time t, given by ES_α^t; if there is no VaR violation, the violation residual is 0. The reason for scaling the residual by $(\mathrm{ES}_\alpha^t - \mu_{t+1})$ will become apparent below.

It follows from Lemma 2.13 that, for a continuous loss distribution, the identity

$$E(K_{t+1} \mid \mathcal{F}_t) = 0$$

is satisfied so that the series of violation residuals (K_t) forms a martingale-difference series. Under stronger assumptions we can use this as the basis for a backtest of expected shortfall estimates. Let us assume that the underlying process generating the losses (L_t) satisfies, for all t, equations of the form $L_t = \mu_t + \sigma_t Z_t$, where μ_t is an $\mathcal{F}_{t-1}$-measurable conditional mean term, σ_t is an $\mathcal{F}_{t-1}$-measurable volatility and the (Z_t) are SWN(0, 1) innovations. This assumption would be satisfied, for example, by an ARMA process with GARCH errors, which mimics many of the essential features of financial return data. Under this assumption we have that $\mathrm{ES}_\alpha^t = \mu_{t+1} + \sigma_{t+1}\,\mathrm{ES}_\alpha(Z)$, where $\mathrm{ES}_\alpha(Z)$ denotes the expected shortfall of the innovation distribution. We can then calculate that

$$K_{t+1} = \left(\frac{Z_{t+1} - \mathrm{ES}_\alpha(Z)}{\mathrm{ES}_\alpha(Z)}\right) I_{\{Z_{t+1} > q_\alpha(Z)\}},$$

so that the sequence of violation residuals (K_t) forms a process of iid variables with mean 0 and an atom of probability mass of size α at zero.

Empirical violation residuals are formed, in the obvious way, by calculating

$$\hat{K}_{t+1} = \left(\frac{L_{t+1} - \widehat{\mathrm{ES}}_\alpha^t}{\widehat{\mathrm{ES}}_\alpha^t - \hat{\mu}_{t+1}}\right) \hat{I}_{t+1}, \tag{9.35}$$

where $\widehat{\mathrm{ES}}_\alpha^t$ denotes the estimated value of ES_α^t at time t, $\hat{I}_{t+1}$ is the VaR violation indicator defined in Section 9.3.1, and $\hat{\mu}_{t+1}$ is an estimate of the conditional mean. In practice, the conditional mean μ_{t+1} is not always estimated (particularly in EWMA-based methods), and when it is estimated it is often close to zero; for this reason we might simplify the calculations by setting $\hat{\mu}_{t+1} = 0$.

We expect the empirical violation residuals to behave like realizations of iid variables from a distribution with mean 0. We can test the hypothesis that the non-zero violation residuals have a mean of zero. The simplest approach is to use a t-test, and this is the option we choose in Section 9.3.4. It is also possible to use a bootstrap test that makes no assumption about the underlying distribution of the violation residuals (see Notes and Comments).

9.3.3 Elicitability and Comparison of Risk Measure Estimates

We now present a more recent approach to backtesting that is useful for comparing sets of risk-measure estimates derived using different methodologies. This approach is founded on the observation that the problem of estimating financial risk measures for the next time period is a special case of the general statistical problem of estimating statistics of a predictive or forecasting distribution. It is therefore natural to use ideas from the forecasting literature in backtesting.

In forecasting it is common to make predictions based on the idea of minimizing a scoring function or prediction error function. For example, if we wish to minimize the squared prediction error, it is well known that we should use the mean of the predictive distribution as our forecast; if we wish to minimize the absolute prediction error, we use the median of the predictive distribution.

The mean and median of the predictive distribution are known as *elicitable* statistical functionals of the distribution because they provide optimal forecasts under particular choices of scoring function. When a statistic is elicitable, there are natural ways of comparing different sets of estimates of that statistic using empirical scores.

For example, let (X_t) denote a time series and suppose we use two procedures A and B to estimate the conditional mean $\mu_{t+1} = E(X_{t+1} \mid \mathcal{F}_t)$ at different time points, based on data up to time t, resulting in estimates $\hat{\mu}_{t+1}^{(j)}$, $j \in \{A, B\}$, $t = 1, \ldots, m$. The conditional mean μ_{t+1} is known to be the optimal prediction of X_{t+1} under a squared error scoring function. We therefore compare our estimates with the actual realized values X_{t+1} by computing squared differences. The superior estimation procedure j will tend to be the one that gives the lowest value of the total squared prediction error $\sum_{t=1}^m (X_{t+1} - \hat{\mu}_{t+1}^{(j)})^2$.

We now give a more formal treatment of basic concepts from elicitability theory and show in particular how the ideas relate to the problem of estimating value-at-risk.

Elicitability theory. A law-invariant risk measure ϱ defined on a space of random variables $\mathcal{M}$ can also be viewed as a statistical functional T defined on a space of distribution functions $\mathcal{X}$. If $L \in \mathcal{M}$ has df $F_L \in \mathcal{X}$, then ϱ and T are linked by $\varrho(L) = T(F_L)$; for example, $\mathrm{VaR}_\alpha(L) = F_L^{\leftarrow}(\alpha)$, so the functional in this case is the generalized inverse at α.

The theory of elicitability is usually presented as a theory of statistical functionals of distribution functions. Our account is based on the presentation in Bellini and Bignozzi (2013). Note that we follow Bellini and Bignozzi in restricting our attention to real-valued functionals; this is more natural for the application to financial risk measures but differs from the more common presentation in the statistical forecasting literature where set-valued functionals are allowed (see, for example, Gneiting 2011).

Elicitable statistical functionals are functionals that minimize expected scores where the expected scores are calculated using *scoring functions*. These quantify the discrepancy between a forecast and a realized value from the distribution. The formal definitions of scoring functions and elicitable functionals are as follows.

Definition 9.6. A scoring function is a function $S\colon \mathbb{R}\times\mathbb{R}\to[0,\infty)$ satisfying, for any $y, l\in\mathbb{R}$:

(i) $S(y,l)\geqslant 0$ and $S(y,l)=0$ if and only if $y=l$;

(ii) $S(y,l)$ is increasing for $y>l$ and decreasing for $y<l$;

(iii) $S(y,l)$ is continuous in y.

Definition 9.7. A real-valued statistical functional T defined on a space of distribution functions $\mathcal{X}$ is said to be elicitable on $\mathcal{X}_T\subseteq\mathcal{X}$ if there exists a scoring function S such that, for every $F\in\mathcal{X}_T$,

(1) $\int_{\mathbb{R}} S(y,l)\,\mathrm{d}F(l)<\infty, \forall y\in\mathbb{R}$,

(2) $T(F)=\arg\min_{y\in\mathbb{R}}\int_{\mathbb{R}} S(y,l)\,\mathrm{d}F(l)$.

In this case the scoring function S is said to be *strictly consistent* for T.

In the context of risk measures, if L is an rv with loss distribution function F_L, then an elicitable risk measure minimizes

$$E(S(y,L))=\int_{\mathbb{R}} S(y,l)\,\mathrm{d}F_L(l) \tag{9.36}$$

with respect to y for every $F_L\in\mathcal{X}_T$, where $\mathcal{X}_T$ is the set of dfs for which the integral in (9.36) is defined.

For example, let $\mathcal{X}$ be the set of dfs of integrable random variables. The mean $E(L)=\int_{\mathbb{R}} l\,\mathrm{d}F_L(l)$ is elicitable on the space $\mathcal{X}_T$ of distribution functions with finite variance. This is clear because it minimizes (9.36) for the strictly consistent scoring function $S(y,l)=(y-l)^2$, as may be easily verified.

Application to the VaR and expectile risk measures. We now consider the VaR and expectile risk measures. The former is elicitable for strictly increasing distribution functions and the latter is elicitable in general, subject of course to the moment conditions imposed by Definition 9.7. We summarize this information and give strictly consistent scoring functions in the following two propositions.

Proposition 9.8. *For any $0 < \alpha < 1$ the statistical functional $T(F_L) = F_L^{\leftarrow}(\alpha)$ is elicitable on the set of strictly increasing distribution functions with finite mean. The scoring function*

$$S_\alpha^q(y, l) = |1_{\{l \leqslant y\}} - \alpha|\,|l - y| \tag{9.37}$$

is strictly consistent for T.

Proof. For L with df F_L, the expected score $E(S_\alpha^q(y, L))$ is a continuous function that is differentiable at all the points of continuity of F_L. The derivative is

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}y} E(S_\alpha^q(y, L)) &= \frac{\mathrm{d}}{\mathrm{d}y} \int_{-\infty}^{\infty} |1_{\{y \geqslant x\}} - \alpha|\,|y - x|\,\mathrm{d}F_L(x) \\
&= \frac{\mathrm{d}}{\mathrm{d}y} \int_{-\infty}^{y} (1 - \alpha)(y - x)\,\mathrm{d}F_L(x) + \frac{\mathrm{d}}{\mathrm{d}y} \int_{y}^{\infty} \alpha(x - y)\,\mathrm{d}F_L(x) \\
&= (1 - \alpha) \int_{-\infty}^{y} \mathrm{d}F_L(x) - \alpha \int_{y}^{\infty} \mathrm{d}F_L(x) \\
&= F_L(y) - \alpha.
\end{aligned}$$

There are two cases to consider. If there exists a point y such that $F_L(y) = \alpha$, then $y = F_L^{\leftarrow}(\alpha)$ and y clearly minimizes $E(S_\alpha^q(y, L))$. If, on the other hand, the set $\{y\colon F_L(y) = \alpha\}$ is empty, then it must be the case that there is a point y at which the distribution function F_L jumps and $F_L(x) - \alpha < 0$ for $x < y$ and $F_L(x) - \alpha > 0$ for $x > y$. It follows again that $y = F_L^{\leftarrow}(\alpha)$ and that y is the unique minimizer of $E(S_\alpha^q(y, L))$. □

Proposition 9.9. *For any $0 < \alpha < 1$ the statistical functional T corresponding to the expectile risk measure e_α is elicitable for all loss distributions F_L with finite variance. The scoring function*

$$S_\alpha^e(y, l) = |1_{\{l \leqslant y\}} - \alpha|(l - y)^2 \tag{9.38}$$

is strictly consistent for T.

Proof. This follows easily from (8.31) and Definition 8.21 in Section 8.2.2. □

Characterizations of elicitable risk measures. The expectile is thus both an elicitable and a coherent risk measure (provided $\alpha \geqslant 0.5$); it is in fact the only risk measure to have both these properties, as shown by Ziegel (2015). Bellini and Bignozzi (2013) have provided an elegant result that characterizes all the elicitable risk measures that have the extra properties required for convexity, coherence and comonotone additivity.

Theorem 9.10 (Bellini and Bignozzi (2013)). *Let $T(F_L)$ be the statistical functional corresponding to a law-invariant risk measure that is both monotonic and translation invariant. Then*

(a) $T(F_L)$ is convex and elicitable if and only if it is a risk measure based on a (convex) loss function (see Example 8.8),

(b) *$T(F_L)$ is coherent and elicitable if and only if it is an expectile e_α with $\alpha \geqslant 0.5$, and*

(c) *$T(F_L)$ is coherent, comonotone additive and elicitable if and only if it coincides with the expected loss.*

In particular, we note that the expected shortfall risk measure cannot be elicitable according to this theorem. There are other ways of demonstrating this more directly (see, for example, Gneiting 2011).

Computing empirical scores. As indicated in the introduction to this section, if we want to compare the performance of different methods of estimating elicitable functionals, we can use the strictly consistent scoring functions suggested by elicitability theory. From Proposition 9.8 we know that VaR_α^t, the quantile of $F_{L_{t+1}|\mathcal{F}_t}$, minimizes $E(S_\alpha^q(y, L_{t+1}) \mid \mathcal{F}_t)$.

Suppose, as in Section 9.3.1, that we compute estimates $\widehat{\mathrm{VaR}}_\alpha^t$ of VaR_α^t on days $t = 1, \ldots, m$, based on information up to time t, and backtest each estimate on day $t + 1$. A natural score is given by

$$\sum_{t=1}^{m} S_\alpha^q(\widehat{\mathrm{VaR}}_\alpha^t, L_{t+1}).$$

If we compute this quantity for different estimation methods, then the methods that give the most accurate estimates of the conditional quantiles VaR_α^t will tend to give the smallest scores.

Unfortunately, since expected shortfall is not an elicitable functional, there is no natural empirical score for comparing sets of estimates of expected shortfall.

9.3.4 Empirical Comparison of Methods Using Backtesting Concepts

In this section we apply various VaR estimation methods to the portfolio of a hypothetical investor in international equity indices and backtest the resulting VaR estimates. The methods we compare belong to the general categories of variance–covariance and historical-simulation methods and are a mix of unconditional and conditional approaches.

The investor is assumed to have domestic currency sterling (GBP) and to invest in the FTSE100 Index, the S&P 500 Index and the SMI (Swiss Market Index). The investor thus has currency exposure to US dollars (USD) and Swiss francs (CHF), and the value of the portfolio is influenced by five risk factors (three log index values and two log exchange rates). The corresponding risk-factor time series for the period 2000–2012 are shown in Figure 9.4.

On any day t we standardize the total portfolio value V_t in sterling to be 1 and assume that the portfolio weights (the proportions of this total value invested in each of the FTSE 100, the S&P 500 and the SMI) are 30%, 40% and 30%, respectively. Using similar reasoning to that in Example 2.1, it may be verified that the loss operator is

$$l_{[t]}(\boldsymbol{x}) = 1 - (0.3\mathrm{e}^{x_1} + 0.4\mathrm{e}^{x_2+x_4} + 0.3\mathrm{e}^{x_3+x_5}),$$

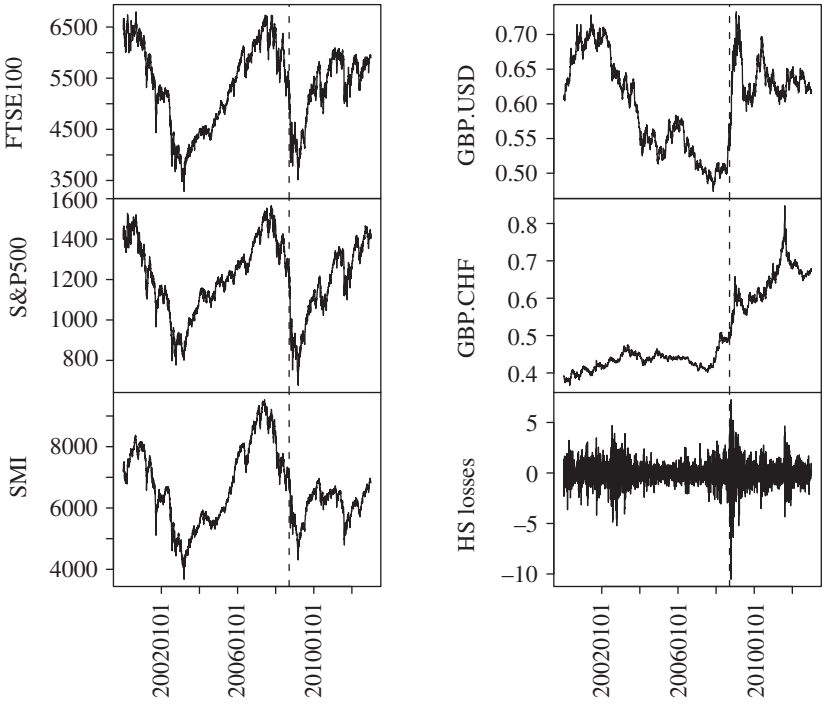

Figure 9.4. Time series of risk-factor values. These are (a) the FTSE 100, (b) the S&P 500, (c) the SMI, (d) the GBP/USD exchange rate and (e) the GBP/CHF exchange rate for the period 2000–2012. The final picture shows the corresponding historical simulation data (9.26) for the portfolio of Section 9.3.4. The vertical dashed line marks the date on which Lehman Brothers filed for bankruptcy.

and its linearized version is

$$l^{\Delta}_{[t]}(\boldsymbol{x}) = -(0.3x_1 + 0.4(x_2 + x_4) + 0.3(x_3 + x_5)),$$

where x_1, x_2 and x_3 represent log-returns on the three indices, and x_4 and x_5 are log-returns on the GBP/USD and GBP/CHF exchange rates.

Our objective is to calculate VaR estimates at the 95% and 99% levels for all trading days in the period 2005–12. Where local public holidays take place in individual markets (e.g. the Fourth of July in the US), we record artificial zero returns for the market in question, thus preserving around 258 days of risk-factor return data in each year. We use the last 1000 days of historical data $\boldsymbol{X}_{t-999}, \ldots, \boldsymbol{X}_t$ to make all VaR estimates for day $t + 1$ using the following methods.

VC. The variance–covariance method assuming multivariate Gaussian risk-factor changes and using the multivariate EWMA method to estimate the conditional covariance matrix of risk-factor changes as described in Section 9.2.2.

HS. The standard unconditional historical-simulation method as described in Section 9.2.3.

HS-GARCH. The univariate dynamic approach to historical simulation in which a GARCH(1, 1) model with a constant conditional mean term and Gaussian innovations is fitted to the historically simulated losses to estimate the volatility of the next day's loss (see Section 9.2.4).

HS-GARCH-*t*. A similar method to HS-GARCH but Student t innovations are assumed in the GARCH model.

HS-GARCH-EVT. A similar method to HS-GARCH and HS-GARCH-*t* but the model parameters are estimated by QML and EVT is applied to the model residuals; see strategy (2) in the univariate approach to dynamic historical simulation described in Section 9.2.4 and see also Section 5.2.6.

HS-MGARCH. The multivariate dynamic approach to historical simulation in which GARCH(1, 1) models with constant conditional mean terms are fitted to each time series of risk-factor changes to estimate volatilities (see Section 9.2.4).

HS-MGARCH-EVT. A similar method to HS-MGARCH but EVT estimators rather than simple empirical estimators are applied to the data constructed in (9.28) to calculate risk-measure estimates.

This collection of methods is of course far from complete and is merely meant as an indication of the kinds of strategies that are possible. In particular, we have confined our interest to rather simple GARCH models and not added, for example, asymmetric innovation distributions, leverage effects (see Section 4.2.3) or regime-switching models, which can often further improve the performance of such methods.

Table 9.1 contains the VaR violation counts for estimates of the 95% and 99% VaR for each of the methods. The violation counts have been broken down by year and the final column shows the total number of violations over the eight-year period. In each cell a binomial test has been carried out using the score statistic (9.33), and test results that are significant at the 5% level are indicated by italics.

At the 95% level the HS-MGARCH and HS-MGARCH-EVT methods clearly give the best overall results over the entire period; the former yields exactly the expected number of violations and the latter yields just one more; the third-best method, in terms of closeness of the number of violations to the expected number, is the HS-GARCH-EVT method. At the 99% level the HS-MGARCH, HS-MGARCH-EVT and HS-GARCH-EVT are again the best methods, although it is difficult to pick a favourite in terms of violation counts only. While HS-MGARCH gives insignificant results in every year period, the HS-MGARCH-EVT and HS-GARCH-EVT methods come closer to the expected number of overall violations; in fact, the HS-MGARCH method gives too few violations overall.

The volatile years 2007 and 2008 are problematic for most methods, with every method yielding more violations than expected. However, the VC method gives an insignificant result in both years at the 95% level, and the HS-MGARCH method gives an insignificant result in both years at the 99% level.

We now compare the results of the VaR backtests using elicitability theory. The results are contained in the first two columns of Table 9.2. According to this metric, the HS-MGARCH-EVT method gives the best VaR estimates at the 95% level, while

Table 9.1. Numbers of violations of the 95% and 99% VaR estimate calculated using various methods, as described in Section 9.3.4. Figures in italics show significant discrepancies between observed and expected violation counts according to a binomial score test at the 5% level.

Year	2005	2006	2007	2008	2009	2010	2011	2012	All
Trading days	258	257	258	259	258	259	258	258	2065
Results for 95% VaR									
Expected no. of violations	13	13	13	13	13	13	13	13	103
VC	8	16	17	19	13	15	14	14	116
HS	*0*	*6*	*28*	*49*	19	*6*	10	*1*	119
HS-GARCH	9	13	*22*	*22*	13	14	9	15	117
HS-GARCH-*t*	9	14	*23*	*22*	14	15	10	15	122
HS-GARCH-EVT	*5*	13	*22*	*21*	13	13	9	13	109
HS-MGARCH	*5*	14	*21*	19	12	9	11	12	103
HS-MGARCH-EVT	*5*	14	*22*	18	13	10	10	12	104
Results for 99% VaR									
Expected no. of violations	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	21
VC	2	*8*	*8*	*8*	2	4	5	*6*	*43*
HS	0	0	*10*	*22*	2	0	2	0	*36*
HS-GARCH	2	*8*	*8*	*10*	5	4	3	3	*43*
HS-GARCH-*t*	2	*8*	*6*	*8*	1	4	2	1	*32*
HS-GARCH-EVT	0	*6*	4	*7*	1	1	2	1	22
HS-MGARCH	0	4	4	5	0	1	2	1	17
HS-MGARCH-EVT	0	4	5	*6*	0	1	2	1	19

the HS-MGARCH method gives the best VaR results at the 99% level. At the 95% level the second lowest score is given by the HS-MGARCH method. At the 99% level the second lowest score is given by the HS-MGARCH-EVT method and the third lowest score is given by the HS-GARCH-EVT method.

It is also noticeable that the standard HS method gives very poor scores at both levels. As an unconditional method it is not well suited to giving estimates of quantiles of the conditional loss distribution; this is also evident from the violation counts in Table 9.1. In Figure 9.5 we show the second half of the year 2008: a period at the height of the 2007–9 credit crisis and one that includes the date on which Lehman Brothers filed for bankruptcy (15 September 2008). The plot shows actual losses as bars with risk-measure estimates for the HS and HS-MGARCH methods superimposed; violations are indicated by circles (HS) and crosses (HS-MGARCH).

Throughout the volatile year 2008, the standard historical-simulation method performs very poorly: there are forty-nine violations of the 95% VaR estimate and twenty-two violations of the 99% VaR estimate. The HS-MGARCH method, being a conditional method, is able to respond to the changes in volatility better and consequently gives nineteen and five violations. In the plot of the second half of the

Table 9.2. Scores based on elicitability theory for the VaR backtests (to four significant figures) and p-values (to two decimal places) for expected shortfall tests as described in Section 9.3.4; figures in italics indicate failure of the expected shortfall test.

	VaR score comparison		Violation residual test			
	95% VaR ($\times 10^6$)	99% VaR ($\times 10^6$)	95% ES	(n)	99% ES	(n)
VC	1081	308.1	*0.00*	116	0.05	43
HS	1399	466.4	*0.02*	119	0.25	36
HS-GARCH	1072	306.8	*0.00*	117	0.05	43
HS-GARCH-t	1074	299.6	0.12	122	0.68	32
HS-GARCH-EVT	1074	295.7	0.59	109	0.65	22
HS-MGARCH	1064	287.8	0.99	103	0.55	17
HS-MGARCH-EVT	1063	289.7	0.83	104	0.94	19

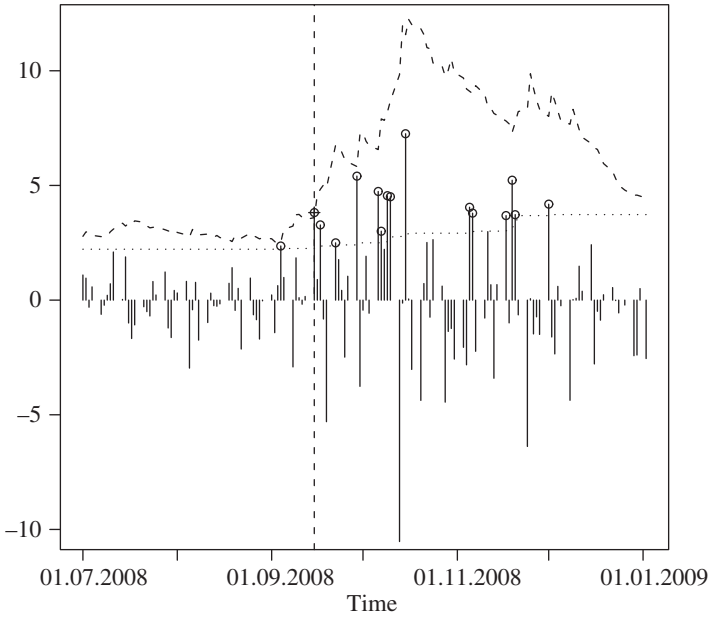

Figure 9.5. Daily losses for the second half of 2008 together with 99% VaR estimates and corresponding violations for the HS and HS-MGARCH methods. The HS VaR estimates are indicated by a dotted line and the corresponding violations are indicated by circles. The HS-MGARCH estimates are given by a dashed line; the only violation for this method in the time period occurred on 15 September 2008 (the day on which Lehman Brothers filed for bankruptcy) and is marked by a dashed vertical line and a crossed circle. For more information see Section 9.3.4.

year we see sixteen of the twenty-two violations of the 99% VaR estimate for the HS method and one of the five violations for the HS-MGARCH method.

In Figure 9.6 we address the hypothesis that VaR violations should form a Bernoulli trials process with geometrically distributed (or approximately exponentially distributed) spacings between violations. The figure shows a Q–Q plot of the

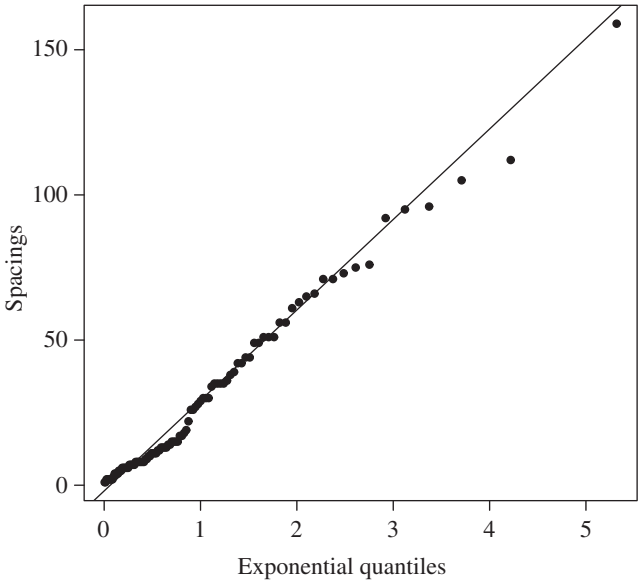

Figure 9.6. Q–Q plot of the spacings (in days) between violations of the estimates of the 99% VaR obtained by the HS-MGARCH method against the corresponding quantiles of a standard exponential distribution (see Section 9.3.4).

spacings (in days) between violations of the estimates of the 99% VaR obtained by the HS-MGARCH method against the corresponding quantiles of a standard exponential distribution; the exponential hypothesis is plausible on the basis of this picture.

In Table 9.2 we also give results for the backtest of expected shortfall based on violation residuals described in Section 9.3.2. For estimates of the 95% expected shortfall, three of the seven methods fail the zero-mean test for the violation residuals; the methods that do not fail are HS-GARCH-t, HS-GARCH-EVT, HS-MGARCH and HS-MGARCH-EVT. For estimates of the 99% expected shortfall, results appear to be even better; five of the methods give insignificant results (HS, HS-GARCH-t, HS-GARCH-EVT, HS-MGARCH, HS-MGARCH-EVT) while the other two give a marginally significant result ($p = 0.05$). However, at the 99% level we have only a small number of violation residuals to test, as indicated in the table by the column marked (n).

On the basis of all results, the HS-MGARCH and HS-MGARCH-EVT methods give the best overall performance in the example we have considered, with the HS-GARCH-EVT method also performing well.

9.3.5 Backtesting the Predictive Distribution

As well as backtesting VaR and expected shortfall we can also devise tests that assess the overall quality of the estimated conditional loss distributions from which the risk-measure estimates are derived. Of course, our primary interest focuses on the measures of tail risk, but it is still useful to backtest our estimates of the

whole predictive distribution to obtain additional confirmation of the risk-measure estimation procedure.

Suppose our objective at every time t is to estimate the conditional loss distribution $F_{L_{t+1}|\mathcal{F}_t}$ and let $U_{t+1} = F_{L_{t+1}|\mathcal{F}_t}(L_{t+1})$. For a given loss process (L_t), consider the formation of a process (U_t) by applying this transformation. As in Section 9.3.2, let us assume that the underlying process generating the losses (L_t) satisfies, for all t, equations of the form $L_t = \mu_t + \sigma_t Z_t$, where μ_t is an $\mathcal{F}_{t-1}$-measurable conditional mean term, σ_t is an $\mathcal{F}_{t-1}$-measurable volatility and the (Z_t) are SWN(0, 1) innovations.

Under this assumption it follows easily from the fact that

$$F_{L_{t+1}|\mathcal{F}_t}(l) = P(\mu_{t+1} + \sigma_{t+1} Z_{t+1} \leqslant l \mid \mathcal{F}_t) = F_Z((l - \mu_{t+1})/\sigma_{t+1})$$

that $U_{t+1} = G_Z(Z_{t+1})$, so (U_t) is a strict white noise process. Moreover, if G_Z is continuous, then the stationary or unconditional distribution of (U_t) must be standard uniform (see Proposition 7.2).

In actual applications we estimate $F_{L_{t+1}|\mathcal{F}_t}$ from data up to time t and we backtest our estimates by forming $\hat{U}_{t+1} := \hat{F}_{L_{t+1}|\mathcal{F}_t}(L_{t+1})$ on day $t + 1$. Suppose we estimate the predictive distribution on days $t = 0, \ldots, n - 1$ and form backtesting data $\hat{U}_1, \ldots, \hat{U}_n$; we expect these to behave like a sample of iid uniform data. The distributional assumption can be assessed by standard goodness-of-fit tests like the chi-squared test or the Kolmogorov–Smirnov test (see Section 15.1.2 for references). It is also possible to form the data $\Phi^{-1}(\hat{U}_1), \ldots, \Phi^{-1}(\hat{U}_n)$, where Φ is the standard normal df; these should behave like iid standard normal data (see again Proposition 7.2) and this can be tested as in Section 3.1.2. The strict white noise assumption can be tested using the approach described in Section 4.1.3.

Notes and Comments

The binomial test for numbers of VaR violations and the geometric test for the times between violations can be found in Kupiec (1995); in both cases, a likelihood ratio test is recommended. For the binomial test, alternatives are the Wald test and the score test (see, for example, Casella and Berger 2002, pp. 493–495). The score test in particular seems to give a test at about the right level for VaR probabilities $\alpha = 0.99$ and $\alpha = 0.95$ in samples of size $m = 250$ or $m = 500$ (i.e. a test with the right Type 1 error of falsely rejecting the null hypothesis of binomial behaviour). Further papers on testing VaR violations for independent Bernoulli behaviour include Christoffersen, Hahn and Inoue (2001) and Christoffersen and Pelletier (2004); the latter paper develops a number of tests of exponential behaviour for the durations between violations and finds that the likelihood ratio test against a Weibull alternative is generally most powerful for detecting clustering of violations.

There has been a large growth in papers on backtesting, and a good overview of regulatory implications is given in Embrechts et al. (2014). Our backtesting material is partly taken from McNeil and Frey (2000), where examples of the binomial test for violation counts and the test of expected shortfall using exceedance residuals can be found. In that paper a bootstrap test of the exceedance residuals is proposed as an

alternative to the simple t-test used in this chapter; see Efron and Tibshirani (1994, p. 224) for a description of the bootstrap hypothesis test. Kerkhof and Melenberg (2004) describe an econometric framework for backtesting risk-based regulatory capital. Two interesting papers on backtesting expected shortfall are Acerbi and Szekely (2014) and Costanzino and Curran (2014).

The relevance of elicitability theory to backtesting is discussed by Gneiting (2011), who also provides a proof that expected shortfall is not elicitable based on the work of Osband (1985). Further relevant papers on elicitability are Davis (2014), Ziegel (2015) and Bellini and Bignozzi (2013).

The idea of testing the estimate of the predictive distribution may be found in Berkowitz (2001, 2002). See also Berkowitz and O'Brien (2002) for a more general article on testing the accuracy of the VaR models of commercial banks. Finally, for a critical insider view on the use of VaR technology on Wall Street in the early days, see the relevant chapters in Brown (2012).

10
Credit Risk

Credit risk is the risk of a loss arising from the failure of a counterparty to honour its contractual obligations. This subsumes both default risk (the risk of losses due to the default of a borrower or a trading partner) and downgrade risk (the risk of losses caused by a deterioration in the credit quality of a counterparty that translates into a downgrading in some rating system). Credit risk is omnipresent in the portfolio of a typical financial institution. To begin with, the lending and corporate bond portfolios are obviously affected by credit risk. Perhaps less obviously, credit risk accompanies any over-the-counter (OTC, i.e. non-exchange-guaranteed) derivative transaction such as a swap, because the default of one of the parties involved may substantially affect the actual pay-off of the transaction. Moreover, there is a specialized market for credit derivatives, such as credit default swaps, in which financial institutions are active players. Credit risk therefore relates to the core activities of most banks. It is also highly relevant to insurance companies, who are exposed to substantial credit risk in their investment portfolios and counterparty default risk in their reinsurance treaties.

The management of credit risk at financial institutions involves a range of tasks. To begin with, an enterprise needs to determine the capital it should hold to absorb losses due to credit risk, for both regulatory and economic capital purposes. It also needs to manage the credit risk on its balance sheet. This involves ensuring that portfolios of credit-risky instruments are well diversified and that portfolios are optimized according to risk–return considerations. The risk profile of the portfolio can also be improved by hedging risk concentrations with credit derivatives or by transferring risk to investors through securitization. Moreover, institutions need to manage their portfolio of traded credit derivatives. This involves the tasks of pricing, hedging and managing collateral for such trades. Finally, financial institutions need to control the counterparty credit risk in their trades and contracts with other institutions. In fact, in the aftermath of the 2007–9 financial crisis, counterparty risk management became one of the most important issues for financial institutions.

With these tasks in mind we have split our treatment of credit risk into four chapters. In the present chapter we establish the foundations for the analysis of credit risk. We introduce the most common credit-risky instruments (Section 10.1), discuss various measures of credit quality (Section 10.2) and present models for the credit risk of a single firm (Sections 10.3–10.6). Moreover, we study basic single-name credit derivatives such as credit default swaps.

Chapters 11 and 12 are concerned with portfolio models, and the crucial issue of dependence between defaults comes to the fore. Chapter 11 treats one-period models with a view to capital adequacy and credit risk management issues for portfolios of largely non-traded assets. Chapter 12 deals with properties of portfolio credit derivatives such as collateralized debt obligations; moreover, we discuss the valuation of these products in standard copula models. Finally, Chapter 17 is concerned with more advanced fully dynamic models of portfolio credit risk. This chapter is also the natural place for a detailed discussion of counterparty credit risk because a proper analysis requires dynamic multivariate credit risk models.

Credit risk models can be divided into *structural* or *firm-value models* on the one hand and *reduced-form models* on the other. Broadly speaking, in a structural model default occurs when a stochastic variable (or, in dynamic models, a stochastic process), generally representing an asset value, falls below a threshold, generally representing liabilities. In reduced-form models the precise mechanism leading to default is left unspecified and the default time of a firm is modelled as a non-negative rv, whose distribution typically depends on economic covariables. In this chapter we treat structural models in Section 10.3, simple reduced-form models with deterministic hazard rates in Section 10.4 and more advanced reduced-form models in Sections 10.5 and 10.6.

10.1 Credit-Risky Instruments

In this section we give an overview of the universe of credit-risky instruments, starting with the simplest examples of loans and bonds. We include discussion of the counterparty credit risk in OTC derivatives trades and we also describe some of the more common modern credit derivative products. In what follows we often use the generic term *obligor* for the borrower, bond issuer, trading partner or counterparty to whom there is a credit exposure. The name stems from the fact that in all cases the obligor has a contractual obligation to make certain payments under certain conditions.

10.1.1 Loans

Loans are the oldest credit-risky "instruments" and come in a myriad of forms. It is common to categorize them according to the type of obligor into *retail* loans (to individuals and small or medium-sized companies), *corporate* loans (to larger companies), *interbank* loans and *sovereign* loans (to governments). In each of these categories there are likely to be a number of different lending products. For example, retail customers may borrow money from a bank using mortgages against property, credit cards and overdrafts.

The common feature of most loans is that a sum of money, known as the *principal*, is advanced to the borrower for a particular term in exchange for a series of defined interest payments, which may be at fixed or floating interest rates. At the end of the term the borrower is required to pay back the principal.

A useful distinction to make is between *secured* and *unsecured* lending. If a loan is secured, the borrower has pledged an asset as collateral for the loan. A prime

example is a mortgage, where the collateral is a property. In the event that the borrower is unable to fulfill its obligation to make interest payments or repay the principal, a situation that is termed *default*, the lender may take possession of the asset. In this way the loss in the event of default may be partly mitigated and money may be recovered by selling the asset. In an unsecured loan the lender has no such claim on a collateral asset and recoveries in the event of default may be a smaller fraction of the so-called *exposure*, which is the value of the outstanding principal and interest payments.

Unlike bonds, which are publicly traded securities, loans are private agreements between the borrower and the lender. Hence there is a wide variety of different loan contracts with different legal features. This makes loans difficult to value under fair-value principles. Book value is commonly used, and where fair-value approaches are applied these mostly fall under the heading of level 3 valuation (see Section 2.2.2).

10.1.2 Bonds

Bonds are publicly traded securities issued by companies and governments that allow the issuer to raise funding on financial markets. Bonds issued by companies are called *corporate bonds* and bonds issued by governments are known as *treasuries*, *sovereign bonds* or, particularly in the UK, *gilts* (gilt-edged securities).

The structure of the pay-offs is akin to that of a loan. The security commits the bond issuer (borrower) to make a series of interest payments to the bond buyer (lender) and pay back the principal at a fixed maturity. The interest payments, or coupons, may be *fixed* at the issuance of the bond (so-called fixed-coupon bonds). Alternatively, there are also bonds where the interest payments vary with market rates (so-called *floating-rate notes*). The reference for the floating rate is often LIBOR (the London Interbank Offered Rate). There are also *convertible bonds*, which allow the purchaser to convert them into shares of the issuing company at predetermined time points. These typically offer lower rates than conventional corporate bonds because the investor is being offered the option to participate in the future growth of the company.

A bondholder is subject to a number of risks, particularly interest-rate risk, spread risk and default risk. As for loans, default risk is the risk that promised coupon and principal payments are not made. Historically, government bonds issued by developed countries have been considered to be default free; for obvious reasons, after the European debt crisis of 2010–12, this notion was called into question.

Spread risk is a form of market risk that refers to changes in *credit spreads*. The credit spread of a defaultable bond measures the difference in the yield of the bond and the yield of an equivalent default-free bond (see Section 10.3.2 for a formal definition of credit spreads). An increase in the spread of a bond means that the market value of the bond falls, which is generally interpreted as indicating that the financial markets perceive an increased default risk for the bond.

10.1.3 Derivative Contracts Subject to Counterparty Risk

A significant proportion of all derivative transactions is carried out over the counter, and there is no central clearing counterparty such as an organized exchange that guarantees the fulfilment of the contractual obligations. These trades are therefore subject to the risk that one of the contracting parties defaults during the transaction, thus affecting the cash flows that are actually received by the other party. This risk, known as *counterparty credit risk*, received a lot of attention during the financial crisis of 2007–9, as some of the institutions heavily involved in derivative transactions experienced worsening credit quality or—in the case of Lehman Brothers—even a default event. Counterparty credit risk management is now a key issue for all financial institutions and is the focus of many new regulatory developments.

In order to illustrate the challenges in measuring and managing counterparty credit risk, we consider the example of an interest swap. This is a contract where two parties A and B agree to exchange a series of interest payments on a given nominal amount of money for a given period. For concreteness assume that A receives payments at a fixed interest rate and makes floating payments at a rate equal to the three-month LIBOR.

Suppose now that A defaults at time τ_A before the maturity of the contract, so that the contract is settled at that date. The consequences will depend on the value of the remaining interest payments at that point in time. If interest rates have risen relative to their value at inception of the contract, the fixed interest payments have decreased in value so that the value of the swap contract has increased for B. Since A is no longer able to fulfill its obligations, its default constitutes a loss for B; the exact size of the loss will depend on the term structure of interest rates at the default time τ_A. On the other hand, if interest rates have fallen relative to their value at $t = 0$, the fixed payments have increased in value so that the swap has a negative value for B. At settlement, B will still have to pay the value of the contract into the bankruptcy pool, so that there is no upside for B in A's default. If B defaults first, the situation is reversed: falling rates lead to a counterparty-risk-related loss for A. This simple example illustrates two important points: the size of the counterparty credit exposure is not known a priori, and it is not even clear who has the credit exposure.

The management of counterparty risk raises a number of issues. First, counterparty risk has to be taken into account in pricing and valuation. This has led to various forms of credit value adjustment (CVA). Second, counterparty risk needs to be controlled using risk-mitigation techniques such as netting and collateralization. Under a netting agreement, the value of all derivatives transactions between A and B is computed and only the aggregated value is subject to counterparty risk; since offsetting transactions cancel each other out, this has the potential to reduce counterparty risk substantially. Under a collateralization agreement, the parties exchange collateral (cash and securities) that serve as a pledge for the receiver. The value of the collateral is adjusted dynamically to reflect changes in the value of the underlying transactions.

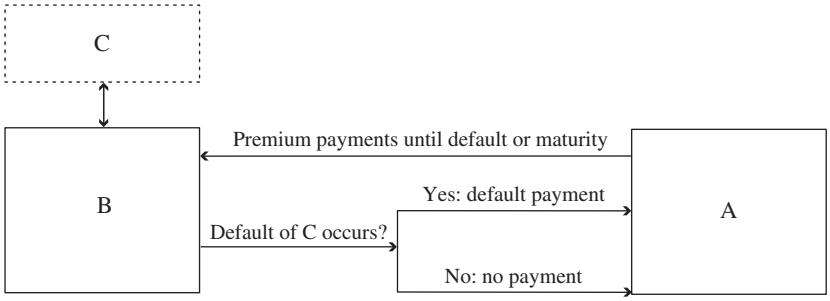

Figure 10.1. The basic structure of a CDS. Firm C is the reference entity, firm A is the protection buyer, and firm B is the protection seller.

The proper assessment of counterparty risk requires a joint modelling of the default times of the two counterparties (A and B) and of the price dynamics of the underlying derivative contract. For that reason we defer the detailed discussion of this topic to Chapter 17.

10.1.4 Credit Default Swaps and Related Credit Derivatives

Credit derivatives are securities that are primarily used for the hedging and trading of credit risk. In contrast to the products considered so far, the promised pay-off of a credit derivative is related to credit events affecting one or more firms. Major participants in the market for credit derivatives are banks, insurance companies and investment funds. Retail banks are typically net buyers of protection against credit events; other investors such as hedge funds and investment banks often act as both sellers and buyers of credit protection.

Credit default swaps. Credit default swaps (CDSs) are the workhorses of the credit derivatives market, and the market for CDSs written on larger corporations is fairly liquid; some numbers on the size of the market are given in Notes and Comments. The basic structure of a CDS is depicted in Figure 10.1. A CDS is a contract between two parties, the *protection buyer* and the *protection seller*. The pay-offs are related to the default of a reference entity (a financial firm or sovereign issuing bonds).

If the reference entity experiences a default event before the maturity date T of the contract, the protection seller makes a default payment to the protection buyer, which mimics the loss due to the default of a bond issued by the reference entity (the reference asset); this part of a CDS is called the *default payment leg*. In this way the protection buyer has acquired financial protection against the loss on the reference asset he would incur in case of a default. As compensation, the protection buyer makes periodic premium payments (typically quarterly or semiannually) to the protection seller (the *premium payment leg*); after the default of the reference entity, premium payments stop. There is no initial payment. The premium payments are quoted in the form of an annualized percentage x^* of the notional value of the reference asset; x^* is termed the (fair or market-quoted) *CDS spread*. For a mathematical description of the payments, see Section 10.4.4.

There are a number of technical and legal issues in the specification of a CDS. In particular, the parties have to agree on the precise definition of a default event and on a procedure to determine the size of the default payment in case a default event of the reference entity occurs. Due to the efforts of bodies such as the International Swaps and Derivatives Association (ISDA), some standardization of these issues has taken place.

Investors enter into CDS contracts for various reasons. To begin with, bond investors with a large credit exposure to the reference entity may buy CDS protection to insure themselves against losses due to the default of a bond. This may be easier than reducing the original bond position because CDS markets are often more liquid than bond markets. Moreover, CDS positions are quickly settled.

CDS contracts are also held for speculative reasons. In particular, so-called *naked* CDS positions, where the protection buyer does not own the bond, are often assumed by investors who are speculating on the widening of the credit spread of the reference entity. These positions are similar to short-selling bonds issued by the reference entity. Note that, in contrast to insurance, there is no requirement for the protection buyer to have *insurable interest*, that is, to actually own a bond issued by the reference entity. The speculative motive for holding CDSs is at least as important as the insurance motive.

There has been some debate about the risks of the CDS market, particularly with respect to the large volume of naked positions and whether or not these should be limited. By taking naked CDS positions speculators can depress the prices of the bonds issued by the reference entity so that default becomes a self-fulfilling prophecy. The debate about the pros and cons of limiting naked CDS positions is akin to the debate about the pros and cons of limiting short selling on equity markets.

A CDS is traded over the counter and is not guaranteed by a clearing house. A CDS position can therefore be subject to a substantial amount of counterparty risk, particularly if a trade is backed by insufficient collateral. A case in point arose during the credit crisis when AIG, which had sold many protection positions, had to be bailed out by the US government to prevent the systemic consequences of allowing it to default on its CDS contracts. There is concern that CDS markets have created a new form of dependency across financial institutions so that the default of one large (systemically important) institution could create a cascade of defaults across the financial sector due to counterparty risk.

On the other hand, CDSs are useful risk-management tools. Because of the liquidity of CDS markets, CDSs are the natural underlying security for many more complex credit derivatives. Models for pricing portfolio-related credit derivatives are usually calibrated to quoted CDS spreads. With improved collateral management in CDS markets it has been argued that the potential for CDS markets to create large-scale default contagion has been substantially reduced (see Notes and Comments).

Credit-linked notes. A credit-linked note is a combination of a credit derivative and a coupon bond that is sold as a fixed package. The coupon payments (and sometimes also the repayment of the principal) are reduced if a third party (the reference entity)

experiences a default event during the lifetime of the contract, so the buyer of a credit-linked note is providing credit protection to the issuer of the note.

Credit-linked notes are issued essentially for two reasons. First, from a legal point of view, a credit-linked note is treated as a fixed-income investment, so that investors who are unable to enter into a transaction involving credit derivatives directly (such as life insurance companies) may nonetheless sell credit protection by buying credit-linked notes. Second, an investor buying a credit-linked note pays the price up front, so that the protection buyer (the issuer of the credit-linked note) is protected against losses caused by the default of the protection seller.

10.1.5 PD, LGD and EAD

Regardless of whether we make a loan, buy a defaultable bond, engage in an OTC derivatives transaction, or act as protection seller in a CDS, the risk of a credit loss is affected by three, generally related, quantities: the *exposure at default* (EAD), the *probability of default* (PD) and the *loss given default* (LGD) or, equivalently, the size of the recovery in the event of default. They are key inputs to the Basel formula in the internal-ratings-based (IRB) approach to determining capital requirements for credit-risky portfolios, so it is important to consider them.

Exposure at default. If we make a loan or buy a bond, our exposure is relatively easy to determine, since it is mainly the principal that is at stake. However, there is some additional uncertainty about the value of the interest payments that could be lost. A further source of exposure uncertainty is due to the widespread use of credit lines. Essentially, a credit line is a ceiling up to which a corporate client can borrow money at given terms, and it is up to the borrower to decide which part of the credit line he actually wants to use. For OTC derivatives, the counterparty risk exposure is even more difficult to quantify, since it is a stochastic variable depending on the unknown time at which a counterparty defaults and the evolution of the value of the derivative up to that point; a case in point is the example of an interest rate swap discussed in Subsection 10.1.3.

In practice, the concept that is used to describe exposure is exposure at default or EAD, which recognizes that the exposure for many instruments will depend on the exact default time. In counterparty credit risk the use of collateral can also reduce the exposure and thus mitigate losses.

Probability of default. When measuring the risk of losses over a fixed time horizon, e.g. one year, we are particularly concerned with estimating the probability that obligors default by the time horizon, a quantity known to practitioners as the probability of default, or PD. The PD is related to the credit quality of an obligor, and Sections 10.2 and 10.3 discuss some of the models that are used to quantify default risk. For instruments where the loss is dependent on the exact timing of default, e.g. OTC derivatives with counterparty risk, the risk of default is described by the whole distribution of possible default times and not just the probability of default by a fixed horizon.

Loss given default. In the event of default, it is unlikely that the entire exposure is lost. For example, when a mortgage holder defaults on a residential mortgage, and there is no realistic possibility of restructuring the debt, the lender can sell the property (the collateral asset) and the proceeds from the sale will make good some of the lost principal. Similarly, when a bond issuer goes into administration, the bondholders join the group of creditors who will be partly recompensed for their losses by the sale of the firm's assets.

Practitioners use the term loss given default, or LGD, to describe the proportion of the exposure that is actually lost in the event of default, or its converse, the recovery, to describe the amount of the exposure that can be recovered through debt restructuring and asset sales.

Dependence of these quantities. It is important to realize that EAD, PD and LGD are dependent quantities. While it is common to attempt to model them in terms of independent random variables, it is unrealistic to do so. For example, in a period of financial distress, when PDs are high, the asset values of firms are depressed and firms are defaulting, recoveries are likely to be correspondingly low, so that there is positive dependence between PDs and LGDs. This will be discussed further in 11.2.3.

Notes and Comments

For further reading on loans and loan pricing we refer to Benzschawel, Dagraca and Fok (2010). For an overview of bonds see Sharpe, Alexander and Bailey (1999).

To get an idea of the size of the CDS market, note that the nominal value (gross notional amount) of the market stood at approximately $60 trillion by the end of 2007, before coming down to a still considerable amount of approximately $25 trillion by the end of 2012. In 2013 the net notional amount was of the order of $2 trillion. For comparison, by the end of 2012 world GDP stood at roughly $80 trillion. To give an example of the size of the speculative market in CDSs, Cont (2010) reports that "when it filed for bankruptcy on September 14, 2008, Lehman Brothers had $155 billion of outstanding debt, but more than $400 billion notional value of CDS contracts had been written with Lehman as reference entity". A good discussion of the role of such credit derivatives in the credit crisis is given in Stulz (2010). The effect of improved collateral management for CDSs on the risk of large-scale contagion in CDS markets is addressed in Brunnermeier, Clerc and Scheicher (2013).

In this brief introduction we have discussed a few essential features of credit derivatives but have omitted the rather involved regulatory, legal and accounting issues related to these instruments. Readers interested in these topics are referred to the paper collections edited by Gregory (2003) and Perraudin (2004), in which pricing issues are also discussed. An excellent treatment of credit derivatives at textbook level is Schönbucher (2003). For a discussion of credit derivatives from the viewpoint of financial engineering we refer to Neftci (2008).

10.2 Measuring Credit Quality

There are various ways of quantifying the credit quality or, equivalently, the default risk of obligors but, broadly speaking, these approaches may be divided into two philosophies. On the one hand, credit quality can be described by a credit rating or credit score that is based on empirical data describing the borrowing and repayment history of the obligor, or of similar obligors. On the other hand, for obligors whose equity is traded on financial markets, prices can be used to infer the market's view of the credit quality of the obligor. This section is devoted to the first philosophy, and market-implied measures of credit quality are treated in the context of structural models in Section 10.3.

Credit ratings and credit scores fulfill a similar function—they both allow us to order obligors according to their credit risk and map that risk to an estimate of default probability. Credit ratings tend to be expressed on an ordered categorical scale, whereas credit scores are often expressed in terms of points on a metric scale. The task of rating obligors, particularly large corporates or sovereigns, is often outsourced to a rating agency such as Moody's or Standard & Poor's (S&P); proprietary rating systems internal to a financial institution can also be used. In the S&P rating system there are seven pre-default rating categories, labelled AAA, AA, A, BBB, BB, B, CCC, with AAA being the highest rating and CCC the lowest rating; Moody's uses nine pre-default rating categories and these are labelled Aaa, Aa, A, Baa, Ba, B, Caa, Ca, C. A finer alpha-numeric system is also used by both agencies.

Credit scores are traditionally used for retail customers and are based on so-called scorecards that banks develop through extensive statistical analyses of historical data. The basic idea is that default risk is modelled as a function of demographic, behavioural and financial covariates that describe the obligor. Using techniques such as logistic regression these covariates are weighted and combined into a score.

10.2.1 Credit Rating Migration

In the credit-migration approach each firm is assigned to a credit-rating category at any given time point. The probability of moving from one credit rating to another over a given risk horizon (typically one year) is then specified. Transition probabilities are typically presented in the form of a matrix; an example from Moody's is presented in Table 10.1. Transition matrices are estimated from historical default data, and standard statistical methods used for this purpose are discussed in Section 10.2.2

In the credit-migration approach we assume that the current credit rating completely determines the default probability, so that this probability can be read from the transition matrix. For instance, if we use the transition matrix presented in Table 10.1, we obtain a one-year default probability for an A-rated company of 0.06%, whereas the default probability of a Caa-rated company is 13.3%. In practice, a correction to the figures in Table 10.1 would probably be undertaken to account for rating withdrawals: that is, transitions to the WR state. The simplest correction would be to divide the first nine probabilities in each row of the table by one minus the final probability in that row; this implicitly assumes that the act of rating withdrawal

Table 10.1. Probabilities of migrating from one rating quality to another within one year. "WR" represents the proportion of firms that were no longer rated at the end of the year, for various reasons including takeover by another company. Source: Ou (2013, Exhibit 26).

Initial rating	Rating at year-end (%)									
	Aaa	Aa	A	Baa	Ba	B	Caa	Ca–C	Default	WR
Aaa	87.20	8.20	0.63	0.00	0.03	0.00	0.00	0.00	0.00	3.93
Aa	0.91	84.57	8.43	0.49	0.06	0.02	0.01	0.00	0.02	5.48
A	0.06	2.48	86.07	5.47	0.57	0.11	0.03	0.00	0.06	5.13
Baa	0.039	0.17	4.11	84.84	4.05	7.55	1.63	0.02	0.17	5.65
Ba	0.01	0.05	0.35	5.52	75.75	7.22	0.58	0.07	1.06	9.39
B	0.01	0.03	0.11	0.32	4.58	73.53	5.81	0.59	3.85	11.16
Caa	0.01	0.02	0.02	0.12	0.38	8.70	61.71	3.72	13.34	12.00
Ca–C	0.00	0.00	0.00	0.00	0.40	2.03	9.38	35.46	37.93	14.80

Table 10.2. Average cumulative default rates (%). Source: Ou (2013, Exhibit 33).

Initial rating	Term						
	1	2	3	4	5	10	15
Aaa	0.00	0.01	0.01	0.04	0.11	0.50	0.93
Aa	0.02	0.07	0.14	0.26	0.38	0.92	1.75
A	0.06	0.20	0.41	0.63	0.87	2.48	4.26
Baa	0.18	0.50	0.89	1.37	1.88	4.70	8.62
Ba	1.11	3.08	5.42	7.93	10.18	19.70	29.17
B	4.05	9.60	15.22	20.13	24.61	41.94	52.22
Caa–C	16.45	27.87	36.91	44.13	50.37	69.48	79.18

contains no information about the likelihood of upgrade, downgrade or default of an obligor.

Rating agencies also produce cumulative default probabilities over larger time horizons. In Table 10.2 we reproduce Moody's cumulative default probabilities for companies with a given current credit rating. For instance, according to this table the probability that a company whose current credit rating is Baa defaults within the next four years is 1.37%. These cumulative default probabilities have been estimated directly from default data. Alternative estimates of multi-year default probabilities can be inferred from one-year transition matrices, as explained in more detail in the next section.

Remark 10.1 (accounting for business cycles). It is a well-established empirical fact that default rates tend to vary with the state of the economy, being high during recessions and low during periods of economic expansion (see Figure 10.2 for an illustration). Transition rates as estimated by S&P or Moody's, on the other hand, are historical averages over longer time horizons covering several business cycles. For instance, the transition rates in Table 10.1 have been estimated from rating-migration data over the period 1970–2012. Moreover, rating agencies focus on the

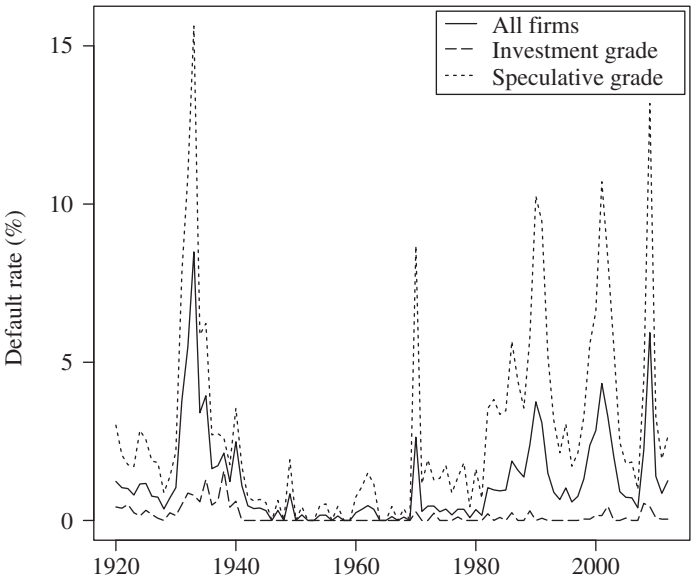

Figure 10.2. Moody's annual default rates from 1920 to 2012. Source for data: Ou (2013, Exhibit 30).

average credit quality "through the business cycle" when attributing a credit rating to a particular firm. The default probabilities from the credit-migration approach are therefore estimates for the average default probability, independent of the current economic environment. In some situations we are interested in "point-in-time" estimates of default probabilities reflecting the current macroeconomic environment, such as in the pricing of a short-term loan. In these situations adjustments to the long-term average default probabilities from the credit-migration approach can be made; for instance, we could use equity prices as an additional source of information, as is done in the public-firm EDF (expected default frequency) model discussed in Section 10.3.3.

10.2.2 Rating Transitions as a Markov Chain

Let (R_t) denote a discrete-time stochastic process defined at times $t = 0, 1, \ldots$ that takes values in $S = \{0, 1, \ldots, n\}$. The set S defines rating states of increasing creditworthiness, with 0 representing default. (R_t) models the evolution of an obligor's rating over time.

We will assume that (R_t) is a *Markov chain*. This means that conditional transition probabilities satisfy the Markov property

$$P(R_t = k \mid R_0 = r_0, R_1 = r_1, \ldots, R_{t-1} = j) = P(R_t = k \mid R_{t-1} = j)$$

for all $t \geqslant 1$ and all $j, r_0, r_1, \ldots, r_{t-2}, k \in S$. In words, the conditional probabilities of rating transitions, given an obligor's rating history, depend only on the previous rating $R_{t-1} = j$ at the last time point and not on the more distant history of how the obligor arrived at a rating state j at time $t - 1$.

The Markov assumption for rating migrations has been criticized; there is evidence for both *momentum* and *stickiness* in empirical rating histories (see Lando and Skodeberg 2002). Momentum is the phenomenon by which obligors who have been recently downgraded to a particular rating are more likely to experience further downgrades than obligors who have had the same rating for a long time. Stickiness is the converse phenomenon by which rating agencies are initially hesitant to downgrade obligors until the evidence for credit deterioration is overwhelming. But despite these issues, the Markov chain assumption is very widely made, because it leads to tractable models with a well-understood theory and to natural estimators for transition probabilities.

The Markov chain is *stationary* if

$$P(R_t = k \mid R_{t-1} = j) = P(R_1 = k \mid R_0 = j)$$

for all $t \geqslant 1$ and all rating states j and k. In this case we can define the transition matrix to be the $(n+1) \times (n+1)$ matrix $\boldsymbol{P} = (p_{jk})$ with elements $p_{jk} = P(R_t = k \mid R_{t-1} = j)$ for any $t \geqslant 1$. Simple conditional probability arguments can be used to derive the *Chapman–Kolmogorov equations*, which say that for any $t \geqslant 2$, and any $j, k \in S$,

$$\begin{aligned} P(R_t = k \mid R_{t-2} = j) &= \sum_{l \in S} P(R_t = k \mid R_{t-1} = l) P(R_{t-1} = l \mid R_{t-2} = j) \\ &= \sum_{l \in S} p_{lk} p_{jl}. \end{aligned}$$

An implication of this is that the matrix of transition probabilities over two time steps is given by $\boldsymbol{P}^2 = \boldsymbol{P} \times \boldsymbol{P}$. Similarly, the matrix of transition probabilities over T time periods is $\boldsymbol{P}^T$. It is, however, not clear how we would compute a matrix of transition probabilities for a fraction of a time period. In fact, this requires the notion of a Markov chain in *continuous time*, which is discussed below.

We now turn to the problem of estimating $\boldsymbol{P}$. Suppose we observe, or are given information about, the ratings of companies at the time points $0, 1, \ldots, T$. This information usually relates to a fluctuating population or cohort of companies, with only a few having complete rating histories throughout $[0, T]$: new companies may be added to the cohort at any time; some companies may default and leave the cohort; others may have their rating withdrawn. In the latter case we will assume that the withdrawal of rating occurs independently of the default or rating-migration risk of the company (which may not be true).

For $t = 0, \ldots, T-1$ and $j \in S \setminus \{0\}$, let N_{tj} denote the number of companies that are rated j at time t and for which a rating is available at time $t+1$; let N_{tjk} denote the subset of those companies that are rated k at time $t+1$. Under the discrete-time, homogeneous Markovian assumption, independent multinomial experiments effectively take place at each time t. In each experiment the N_{tj} companies rated j can be thought of as being randomly allocated to the ratings $k \in S$ according to probabilities p_{jk} that satisfy $\sum_{k=0}^{n} p_{jk} = 1$.

In this framework the likelihood is given by

$$L((p_{jk}); (N_{tj}), (N_{tjk})) = \prod_{t=0}^{T-1} \left(\prod_{j=1}^{n} \left(N_{tj}! \prod_{k=0}^{n} \frac{p_{jk}^{N_{tjk}}}{N_{tjk}!} \right) \right),$$

and if this is maximized subject to the constraints that $\sum_{k=0}^{n} p_{jk} = 1$ for $j = 1, \ldots, n$, we obtain the maximum likelihood estimator

$$\hat{p}_{jk} = \frac{\sum_{t=0}^{T-1} N_{tjk}}{\sum_{t=0}^{T-1} N_{tj}}. \tag{10.1}$$

There are a number of drawbacks to modelling rating transitions as a discrete-time Markov chain. In practice, rating changes tend to take place on irregularly spaced dates. While such data can be approximated by a regularly spaced time series (or panel) of, say, yearly, quarterly or monthly ratings, there will be a loss of information in doing so. The discrete-time model described above would ignore any information about intermediate transitions taking place between two times t and $t + 1$. For example, if an obligor is downgraded from A to BBB to BB over the course of the period $[t, t+1]$, this obligor will simply be recorded as migrating from A to BB and the information about transitions from A to BBB and BBB to BB will not be recorded. Moreover, the estimation procedure for a discrete-time chain tends to result in sparse estimates of transition matrices with quite a lot of zero entries. For example, if no transitions between AAA and default within a single time period are observed, then the probability of such a transition will be estimated to be zero. However, in reality such a transition is possible, if unlikely, and so its estimated probability of occurrence should not be zero.

It is thus more satisfactory to model rating transitions as a phenomenon in continuous time. In this case, transition probabilities are not modelled directly but are instead given in terms of transition rates. Intuitively, the relationship between transition rates and transition probabilities can be described as follows.

Assume that over any small time step of duration δt the probability of a transition from rating j to k is given approximately by $\lambda_{jk}\delta t$ for some constant $\lambda_{jk} > 0$, which is the *transition rate* between rating j and rating k. The probability of staying at rating j is given by $1 - \sum_{k \neq j} \lambda_{jk}\delta t$. If we define a matrix Λ to have off-diagonal entries λ_{jk} and diagonal entries $-\sum_{k \neq j} \lambda_{jk}$, we can summarize the implied transition probabilities for the small time step δt in the matrix $(I_{n+1} + \Lambda\delta t)$. We now consider transitions in the period $[0, t]$ and denote the corresponding matrix of transition probabilities by $\boldsymbol{P}(t)$. If we divide the time period into N small time steps of size $\delta t = t/N$ for N large, the matrix of transition probabilities can be approximated by

$$\boldsymbol{P}(t) \approx \left(I_{n+1} + \frac{\Lambda t}{N} \right)^N,$$

which converges, as $N \to \infty$, to the so-called matrix exponential of Λt:

$$\boldsymbol{P}(t) = \mathrm{e}^{\Lambda t}.$$

This formulation gives us a method of computing transition probabilities for any time horizon t in terms of the matrix Λ, the so-called *generator matrix*.

A Markov chain in continuous time with generator matrix Λ can be constructed in the following way. An obligor remains in rating state j for an exponentially distributed amount of time with parameter

$$\lambda_{jj} = \sum_{k \neq j} \lambda_{jk},$$

i.e. minus the diagonal element of the generator matrix. When a transition takes place the new rating is determined by a multinomial experiment in which the probability of a transition from state j to state k is given by $\lambda_{jk}/\lambda_{jj}$.

This construction also leads to natural estimators for the matrix Λ. Since λ_{jk} is the instantaneous rate of migrating from j to k, we can estimate it by

$$\hat{\lambda}_{jk} = \frac{N_{jk}(T)}{\int_0^T Y_j(t)\,\mathrm{d}t}, \tag{10.2}$$

where $N_{jk}(T)$ is the total number of observed transitions from j to k over the time period $[0, T]$ and $Y_i(t)$ is the number of obligors with rating j at time t; the denominator therefore represents the total time spent in state j by all the companies in the data set. Note that this is the continuous-time analogue of the maximum likelihood estimator in (10.1); it is not surprising, therefore, that (10.2) can be shown to be the maximum likelihood estimator for the transition intensities of a homogenous continuous-time Markov chain.

Notes and Comments

There is a large literature on credit scoring, and useful starter references are Thomas (2009) and Hand and Henley (1997). In addition to the well-known commercial rating agencies there are now open rating systems. One example is the Credit Research Initiative at the Risk Management Institute of the National University of Singapore (see www.rmicri.org).

An alternative discussion of models based on rating migration is given in Chapters 7 and 8 of Crouhy, Galai and Mark (2001). Statistical approaches to the estimation of rating-transition matrices are discussed in Hu, Kiesel and Perraudin (2002) and Lando and Skodeberg (2002). The latter paper also shows that there is momentum in rating-transition data, which contradicts the assumption that rating transitions form a Markov chain. An example of an industry model based on credit ratings is CreditMetrics: see RiskMetrics Group (1997).

The literature on the statistical properties of rating transitions is surveyed extensively in Chapter 4 of Duffie and Singleton (2003). The maximum likelihood estimator of the infinitesimal generator of a continuous-time Markov chain is formally derived in Albert (1962). For further information on Markov chains we refer to the standard textbook by Norris (1997).

10.3 Structural Models of Default

In structural or firm-value models of default one postulates a mechanism for the default of a firm in terms of the relationship between its assets and liabilities. Typically, default occurs whenever a stochastic variable (or in dynamic models a stochastic process) generally representing an asset value falls below a threshold representing liabilities. The kind of thinking embodied in these models has been very influential in the analysis of credit risk and in the development of industry solutions, so that this is a natural starting point for a discussion of credit risk models. We begin with a detailed analysis of the seminal model of Merton (1974) (in Sections 10.3.1 and 10.3.2). Industry implementations of structural models are discussed in Section 10.3.3.

From now on we denote a generic stochastic process in continuous time by (X_t); the value of the process at time $t \geqslant 0$ is given by the rv X_t.

10.3.1 The Merton Model

The model proposed in Merton (1974) is the prototype of all firm-value models. Consider a firm whose asset value follows some stochastic process (V_t). The firm finances itself by *equity* (i.e. by issuing shares) and by *debt*. In Merton's model, debt consists of zero-coupon bonds with common maturity T; the nominal value of debt at maturity is given by the constant B. Moreover, it is assumed that the firm cannot pay out dividends or issue new debt.

The values at time t of equity and debt are denoted by S_t and B_t. Default occurs if the firm misses a payment to its debtholders, which in the Merton model can occur only at the maturity T of the bonds. At T we have to distinguish between two cases.

(i) $V_T > B$: the value of the firm's assets exceeds the nominal value of the liabilities. In that case the debtholders (the owners of the zero-coupon bonds) receive B, the shareholders receive the residual value $S_T = V_T - B$, and there is no default.

(ii) $V_T \leqslant B$: the value of the firm's assets is less than its liabilities and the firm cannot meet its financial obligations. In that case shareholders have no interest in providing new equity capital, as these funds would go immediately to the bondholders. They therefore let the firm go into default. Control over the firm's assets is passed on to the bondholders, who liquidate the firm and distribute the proceeds among themselves. Shareholders pay and receive nothing, so that we have $B_T = V_T$, $S_T = 0$.

Summarizing, we have the relationships

$$S_T = \max(V_T - B, 0) = (V_T - B)^+, \tag{10.3}$$

$$B_T = \min(V_T, B) = B - (B - V_T)^+. \tag{10.4}$$

Equation (10.3) implies that the value of the firm's equity at time T equals the payoff of a European call option on V_T, while (10.4) implies that the value of the firm's

debt at maturity equals the nominal value of the liabilities minus the pay-off of a European put option on V_T with exercise price equal to B.

This model is of course a stylized description of default. In reality, the structure of a company's debt is much more complex, so that default can occur on many different dates. Moreover, under modern bankruptcy code, default does not automatically imply bankruptcy, i.e. liquidation of a firm. Nonetheless, Merton's model is a useful starting point for modelling credit risk and for pricing securities subject to default.

Remark 10.2. The option interpretation of equity and debt is useful in explaining potential conflicts of interest between the shareholders and debtholders of a company. It is well known that, all other things being equal, the value of an option increases if the volatility of the underlying security is increased. Shareholders therefore have an interest in the firm taking on risky projects. Bondholders, on the other hand, have a short position in a put option on the firm's assets and would therefore like to see the volatility of the asset value reduced.

In the Merton model it is assumed that under the real-world or physical probability measure P the process (V_t) follows a diffusion model (known as the Black–Scholes model or geometric Brownian motion) of the form

$$\mathrm{d}V_t = \mu_V V_t \,\mathrm{d}t + \sigma_V V_t \,\mathrm{d}W_t \tag{10.5}$$

for constants $\mu_V \in \mathbb{R}$ (the drift of the asset value process), $\sigma_V > 0$ (the volatility of the asset value process), and a standard Brownian motion (W_t). Equation (10.5) can be solved explicitly, and it can be shown that

$$V_T = V_0 \exp((\mu_V - \tfrac{1}{2}\sigma_V^2)T + \sigma_V W_T).$$

Since $W_T \sim N(0, T)$, it follows that $\ln V_T \sim N(\ln V_0 + (\mu_V - \frac{1}{2}\sigma_V^2)T, \sigma_V^2 T)$. Under the dynamics (10.5), the default probability of the firm is readily computed. We have

$$P(V_T \leqslant B) = P(\ln V_T \leqslant \ln B) = \Phi\left(\frac{\ln(B/V_0) - (\mu_V - \frac{1}{2}\sigma_V^2)T}{\sigma_V\sqrt{T}}\right). \tag{10.6}$$

It may be deduced from (10.6) that the default probability is increasing in B, decreasing in V_0 and μ_V and, for $V_0 > B$, increasing in σ_V. All these properties are perfectly in line with economic intuition.

Figure 10.3 shows two simulated trajectories for the asset value process (V_t) for values $V_0 = 1$, $\mu_V = 0.03$ and $\sigma_V = 0.25$. Assuming that $B = 0.85$ and $T = 1$, one path is a default path, terminating at a value $V_T < B$, while the other is a non-default path.

10.3.2 Pricing in Merton's Model

In the context of Merton's model one can price securities whose pay-off depends on the value V_T of the firm's assets at T. Prime examples are the firm's debt, or, equivalently, the zero-coupon bonds issued by the firm, and the firm's equity. In our analysis of pricing in the context of the Merton model we make use of a few basic

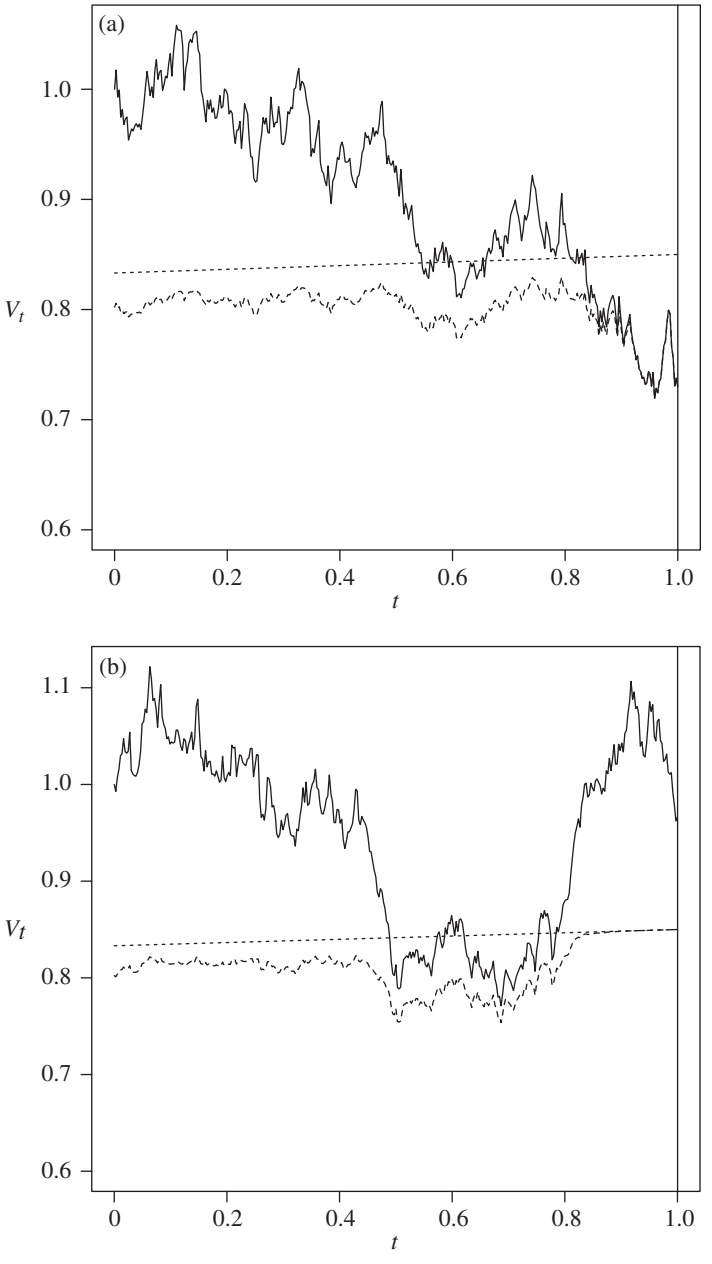

Figure 10.3. Illustration of (a) a default path and (b) a non-default path in Merton's model. The solid lines show simulated one-year trajectories for the asset value process (V_t) starting at $V_0 = 1$ with parameters $\mu_V = 0.03$ and $\sigma_V = 0.25$. Assuming that the debt has face value $B = 0.85$ and maturity $T = 1$ and that the interest rate is $r = 0.02$, the dotted curve shows the value of default-free debt $(Bp_0(t, T))$ while the dashed line shows the evolution of the company's debt B_t according to formula (10.12). The difference between the asset value V_t and the debt B_t is the value of equity S_t.

concepts from financial mathematics and stochastic calculus; references to useful texts in financial mathematics are given in Notes and Comments.

We make the following assumptions.

Assumption 10.3.

(i) *The risk-free interest rate is deterministic and equal to* $r \geqslant 0$.

(ii) *The firm's asset-value process* (V_t) *is independent of the way the firm is financed, and in particular it is independent of the debt level* B.

(iii) *The asset value* (V_t) *can be traded on a frictionless market, and the asset-value dynamics are given by the geometric Brownian motion (10.5).*

These assumptions merit some comments. First, the independence of (V_t) from the financial structure of the firm is questionable, because a very high debt level, and hence a high default probability, may adversely affect the ability of a firm to generate business, hence affecting the value of its assets. This is a special case of the indirect bankruptcy costs discussed in Section 1.4.2. Second, while there are many firms with traded equity, the value of the assets of a firm is usually neither completely observable nor traded. We come back to this issue in Section 10.3.3. For an example where (iii) holds, think of an investment company or trust that invests in liquidly traded securities and uses debt financing to leverage its position. In that case V_t corresponds to the value of the investment portfolio at time t, and this portfolio consists of traded securities by assumption.

Pricing of equity and debt. Consider a claim on the asset value of the firm with maturity T and pay-off $h(V_T)$, such as the firm's equity and debt in (10.3) and (10.4). Under Assumption 10.3, the fair value $f(t, V_t)$ of this claim at time $t \leqslant T$ can be computed using the risk-neutral pricing rule as the expectation of the discounted pay-off under the risk-neutral measure Q, that is,

$$f(t, V_t) = E^Q(\mathrm{e}^{-r(T-t)}h(V_T) \mid \mathcal{F}_t). \tag{10.7}$$

According to (10.3), the firm's equity corresponds to a European call on (V_t) with exercise price B and maturity T. The risk-neutral value of equity obtained from (10.7) is therefore given simply by the Black–Scholes price C^{BS} of a European call. This yields

$$S_t = C^{\mathrm{BS}}(t, V_t; r, \sigma_V, B, T) := V_t\Phi(d_{t,1}) - B\mathrm{e}^{-r(T-t)}\Phi(d_{t,2}), \tag{10.8}$$

where the arguments are given by

$$d_{t,1} = \frac{\ln V_t - \ln B + (r + \frac{1}{2}\sigma_V^2)(T-t)}{\sigma_V\sqrt{T-t}}, \qquad d_{t,2} = d_{t,1} - \sigma_V\sqrt{T-t}. \tag{10.9}$$

Next we turn to the valuation of the risky debt issued by the firm. Since we assumed a constant interest rate r, the price at $t \leqslant T$ of a default-free zero-coupon bond with maturity T and a face value of one equals $p_0(t, T) = \mathrm{e}^{-r(T-t)}$. According to (10.4) we have

$$B_t = Bp_0(t, T) - P^{\mathrm{BS}}(t, V_t; r, \sigma_V, B, T), \tag{10.10}$$

where $P^{\mathrm{BS}}(t, V; r, \sigma_V, B, T)$ denotes the Black–Scholes price of a European put with strike B, maturity T on (V_t) for given interest rate r, and volatility σ_V. It is well known that

$$P^{\mathrm{BS}}(t, V_t; r, \sigma_V, B, T) = B\mathrm{e}^{-r(T-t)}\Phi(-d_{t,2}) - V_t\Phi(-d_{t,1}), \tag{10.11}$$

with $d_{t,1}$ and $d_{t,2}$ as in (10.9). Combining (10.10) and (10.11) we get

$$B_t = p_0(t, T)B\Phi(d_{t,2}) + V_t\Phi(-d_{t,1}). \tag{10.12}$$

Lines showing the evolution of B_t as a function of the evolution of V_t under the assumption that $r = 0.02$ have been added to Figure 10.3. The difference between the asset value V_t and the debt B_t is the value of equity S_t; note how the value of equity is essentially negligible for $t > 0.8$ in the default path.

Volatility of the firm's equity. It is interesting to compute the volatility of the equity of the firm under Assumption 10.3. To this end we define the quantity

$$\nu(t, V_t) := \frac{V_t C_V^{\mathrm{BS}}(t, V_t)}{C^{\mathrm{BS}}(t, V_t)}.$$

In the context of option pricing this is known as the *elasticity* of a European call with respect to the price of the underlying security. In our context it measures the percentage change in the value of equity per percentage change in the value of the underlying assets.

If we apply Itô's formula to $S_t = C^{\mathrm{BS}}(t, V_t; r, \sigma_V, B, T)$ we obtain

$$\mathrm{d}S_t = \sigma_V C_V^{\mathrm{BS}}(t, V_t)V_t\,\mathrm{d}W_t + (C_t^{\mathrm{BS}}(t, V_t) + \mu_V C_V^{\mathrm{BS}}(t, V_t)V_t + \tfrac{1}{2}\sigma_V^2 V_t^2 C_{VV}^{\mathrm{BS}})\,\mathrm{d}t.$$

Using the definition of the elasticity ν, we may rewrite the $\mathrm{d}W_t$ term in the form

$$\sigma_V C_V^{\mathrm{BS}}(t, V_t)V_t\,\mathrm{d}W_t = \sigma_V \nu(t, V_t)C^{\mathrm{BS}}(t, V_t)\,\mathrm{d}W_t,$$

from which we conclude that the volatility of the firm's equity at time t is a function $\sigma_S(t, V_t)$ of time and of the current asset value V_t that takes the form

$$\sigma_S(t, V_t) = \nu(t, V_t)\sigma_V. \tag{10.13}$$

The volatility of the firm's equity is therefore greater than σ_V, since the elasticity of a European call is always greater than one.

Risk-neutral and physical default probabilities. Next we compare physical and risk-neutral default probabilities in Merton's model. It is a basic result from financial mathematics that under the risk-neutral measure Q the process (V_t) satisfies the stochastic differential equation (SDE) $\mathrm{d}V_t = rV_t\,\mathrm{d}t + \sigma_V V_t\,\mathrm{d}\tilde{W}_t$ for a standard Q-Brownian motion $\tilde{W}$. Note how the drift μ_V in (10.5) has been replaced by the risk-free interest rate r. The risk-neutral default probability is therefore given by the formula (10.6), evaluated with $\mu_V = r$:

$$q = Q(V_T \leqslant B) = \Phi\left(\frac{\ln B - \ln V_0 - (r - \frac{1}{2}\sigma_V^2)T}{\sigma_V\sqrt{T}}\right).$$

Comparing this with the physical default probability $p = P(V_T \leqslant B)$ as given in (10.6) we obtain the relationship

$$q = \Phi\left(\Phi^{-1}(p) + \frac{\mu_V - r}{\sigma_V}\sqrt{T}\right). \tag{10.14}$$

The correction term $(\mu_V - r)/\sigma_V$ equals the *Sharpe ratio* of V (a popular measure of the risk premium earned by the firm). The transition formula (10.14) is sometimes applied in practice to go from physical to risk-neutral default probabilities. Note, however, that (10.14) is supported by theoretical arguments only in the narrow context of the Merton model.

Credit spread. We may use (10.12) to infer the credit spread $c(t, T)$ implied by Merton's model. The credit spread measures the difference between the continuously compounded yield to maturity of a defaultable zero-coupon bond $p_1(t, T)$ and that of a default-free zero-coupon bond $p_0(t, T)$. It is defined by

$$c(t, T) = \frac{-1}{T - t}(\ln p_1(t, T) - \ln p_0(T, t)) = \frac{-1}{T - t} \ln \frac{p_1(t, T)}{p_0(t, T)}. \tag{10.15}$$

Throughout the book we use the convention that a zero-coupon bond has a nominal value equal to 1. In line with this convention we assume that the pay-off at T of a zero-coupon bond issued by the firm is given by $(1/B)B_T$, so that the price of such a bond at time $t \leqslant T$ is given by $p_1(t, T) = (1/B)B_t$. We therefore obtain

$$c(t, T) = \frac{-1}{T - t} \ln \left(\Phi(d_{t,2}) + \frac{V_t}{Bp_0(t, T)}\Phi(-d_{t,1})\right). \tag{10.16}$$

Since $d_{t,1}$ can be rewritten as

$$d_{t,1} = \frac{-\ln(Bp_0(t, T)/V_t) + \frac{1}{2}\sigma_V^2(T - t)}{\sigma_V\sqrt{T - t}},$$

and similarly for $d_{t,2}$, we conclude that, for a fixed time to maturity $T - t$, the spread $c(t, T)$ depends only on the volatility σ_V and on the ratio $d := Bp_0(t, T)/V_t$, which is the ratio of the discounted nominal value of the firm's debt to the value of the firm's assets and is hence a measure of the relative debt level or *leverage* of the firm. As the price of a European put (10.11) is increasing in the volatility, it follows from (10.10) that $c(t, T)$ is increasing in σ_V. In Figure 10.4 we plot the credit spread as a function of σ_V and of the time to maturity $\tau = T - t$.

Extensions. Merton's model is quite simplistic. Over the years this has given rise to a rich literature on firm-value models. We briefly comment on the most important research directions (bibliographic references are given in Notes and Comments). To begin with, the observation that, in reality, firms can default at essentially any time (and not only at a deterministic point in time T) has led to the development of so-called *first-passage-time models*. In this class of models default occurs when the asset-value process crosses a default threshold B for the first time; the threshold is usually interpreted as the average value of the liabilities. Formally, the default time τ is defined by $\tau = \inf\{t \geqslant 0 \colon V_t \leqslant B\}$. Further technical developments

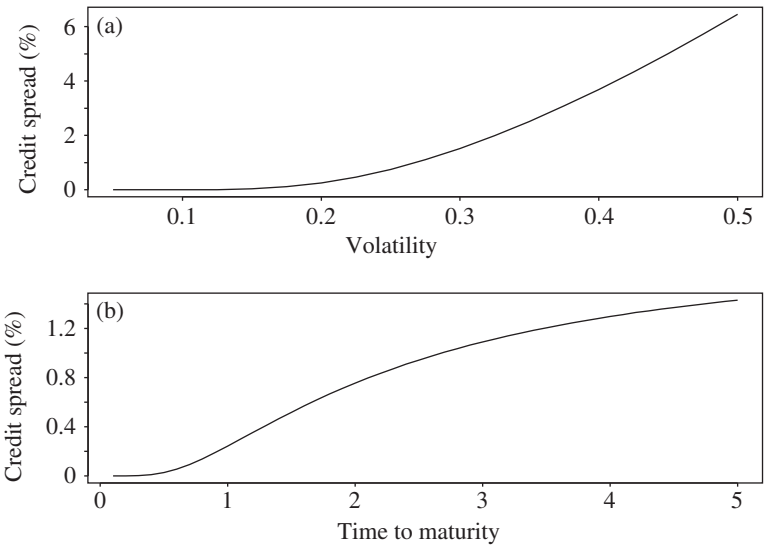

Figure 10.4. Credit spread $c(t, T)$ in per cent as a function of (a) the firm's volatility σ_V and (b) the time to maturity $\tau = T - t$ for fixed leverage measure $d = 0.6$ (in (a) $\tau = 2$ years; in (b) $\sigma_V = 0.25$). Note that, for a time to maturity smaller than approximately three months, the credit spread implied by Merton's model is basically equal to zero. This is not in line with most empirical studies of corporate bond spreads and has given rise to a number of extensions of Merton's model that are listed in Notes and Comments. We will see in Section 10.5.3 that reduced-form models lead to a more reasonable behaviour of short-term credit spreads.

include models with stochastic default-free interest rates and models where the asset-value process (V_t) is given by a diffusion with jumps.

Firm-value models with an *endogenous default threshold* are an interesting economic extension of Merton's model. Here the default boundary B is not fixed a priori by the modeller but is determined endogenously by strategic considerations of the shareholders. Finally, structural models with *incomplete information* on asset value and/or liabilities provide an important link between the structural and reduced-form approaches to credit risk modelling.

10.3.3 Structural Models in Practice: EDF and DD

There are a number of industry models that descend from the Merton model. An important example is the so-called *public-firm EDF model* that is maintained by Moody's Analytics. The acronym EDF stands for *expected default frequency*; this is a specific estimate of the physical default probability of a given firm over a one-year horizon. The methodology proposed by Moody's Analytics builds on earlier work by KMV (a private company named after its founders Kealhofer, McQuown and Vasicek) in the 1990s, and is also known as the KMV model. Our presentation of the public-firm EDF model is based on Crosbie and Bohn (2002) and Sun, Munves and Hamilton (2012). We concentrate on the main ideas, since detailed information about actual implementation and calibration procedures is proprietary and these procedures may change over time.

Overview. Recall that in the classic Merton model the one-year default probability of a given firm is given by the probability that the asset value in one year lies below the threshold B representing the overall liabilities of the firm. Under Assumption 10.3, the one-year default probability is a function of the current asset value V_0, the (annualized) drift μ_V and volatility σ_V of the asset-value process, and the threshold B; using (10.6) with $T = 1$ and recalling that $\Phi(d) = 1 - \Phi(-d)$ we infer that

$$\mathrm{EDF}_{\mathrm{Merton}} = 1 - \Phi\left(\frac{\ln V_0 - \ln B + (\mu_V - \frac{1}{2}\sigma_V^2)}{\sigma_V}\right). \tag{10.17}$$

In the public-firm EDF model a similar structure is assumed for the EDF; however, $1 - \Phi$ is replaced by some empirically estimated function, B is replaced by a new default threshold $\tilde{B}$ representing the structure of the firm's liabilities more closely, and the term $(\mu_V - \frac{1}{2}\sigma_V^2)$ in the numerator is sometimes omitted for expositional ease. Moreover, the current asset value V_0 and the asset volatility σ_V are inferred (or "backed out") from information about the firm's equity value.

Determination of the asset value and the asset volatility. Firm-value-based credit risk models are based on the *market value* V_0 of the firm's assets. This makes sense as the market value is a forward-looking measure that reflects investor expectations about the business prospects and future cash flows of the firm. Unfortunately, in contrast to the assumptions underlying Merton's model, in most cases there is no market for the assets of a firm, so that the asset value is not directly observable. Moreover, the market value can differ greatly from the value of a company as measured by accountancy rules (the so-called book value), so that accounting information and balance sheet data are of limited use in inferring the asset value V_0. For these reasons the public-firm EDF model relies on an indirect approach and infers values of V_t at different times t from the more easily observed values of a firm's equity S_t. This approach simultaneously provides estimates of V_0 and of the asset volatility σ_V. The latter estimate is needed since σ_V has a strong impact on default probabilities; all other things being equal, risky firms with a comparatively high asset volatility σ_V have a higher default probability than firms with a low asset volatility.

We explain the estimation approach in the context of the Merton model. Recall that under Assumption 10.3 we have that

$$S_t = C^{\mathrm{BS}}(t, V_t; r, \sigma_V, B, T). \tag{10.18}$$

Obviously, at a fixed point in time, $t = 0$ say, (10.18) is an equation with two unknowns, V_0 and σ_V. To overcome this difficulty, one may use an iterative procedure. In step (1), (10.18) with some initial estimate $\sigma_V^{(0)}$ is used to infer a time series of asset values $(V_t^{(0)})$ from equity values. Then a new volatility estimate $\sigma_V^{(1)}$ is estimated from this time series; a new time series $(V_t^{(1)})$ is then constructed using (10.18) with $\sigma_V^{(1)}$. This procedure is iterated n times, until the volatility estimates $\sigma_V^{(n-1)}$ and $\sigma_V^{(n)}$ generated in step $(n-1)$ and step (n) are sufficiently close.

In the public-firm EDF model, the capital structure of the firm is modelled in a more sophisticated manner than in Merton's model. There are several classes

of liabilities, such as long- and short-term debt and convertible bonds, the model allows for intermediate cash payouts corresponding to coupons and dividends, and default can occur at any time. Moreover, the default point (the threshold value $\tilde{B}$ such that the company defaults if (V_t) falls below $\tilde{B}$) is determined from a more detailed analysis of the term structure of the firm's debt. The equity value is thus no longer given by (10.18) but by some different function $f(t, V_t, \sigma_V)$, which has to be computed numerically. The general idea of the approach used to estimate V_0 and σ_V is, however, exactly as described above.

Calculation of EDFs. In the Merton model, default occurs if the value of a firm's assets falls below the value of its liabilities. With lognormally distributed asset values, as implied for instance by Assumption 10.3, this leads to default probabilities of the form $\mathrm{EDF}_{\mathrm{Merton}}$ as in (10.17). This relationship between asset value and default probability may be too simplistic to be an accurate description of actual default probabilities. For instance, asset values are not necessarily lognormal but might follow a distribution with heavy tails and there might be payments due at an intermediate point in time causing default at that date.

For these reasons, in the public-firm EDF model a new state variable is introduced in an intermediate step. This is the so-called *distance-to-default* (DD), given by

$$\mathrm{DD} := (\log V_0 - \log \tilde{B})/\sigma_V. \tag{10.19}$$

Here, $\tilde{B}$ represents the default threshold; in some versions of the model $\tilde{B}$ is modelled as the sum of the liabilities payable within one year and half of the longer-term debt. Sometimes practitioners call the distance-to-default the "number of standard deviations a company is away from its default threshold". Note that (10.19) is in fact an approximation of the argument of (10.17), since μ_V and σ_V^2 are usually small.

In the EDF methodology it is assumed that the distance-to-default *ranks* firms in the sense that firms with a higher DD exhibit a higher default probability. The functional relationship between DD and EDF is determined empirically; using a database of historical default events, the proportion of firms with DD in a given small range that default within a year is estimated. This proportion is the empirically estimated EDF. The DD-to-EDF mapping exhibits "heavy tails": for high-quality firms with a large DD the empirically estimated EDF is much higher than $\mathrm{EDF}_{\mathrm{Merton}}$ as given in (10.17). For instance, for a firm with a DD equal to 4 we find that $\mathrm{EDF}_{\mathrm{Merton}} \approx 0.003\%$, whereas the empirically estimated EDF equals 0.4%.

In Table 10.3 we illustrate the computation of the EDF for two different firms, Johnson & Johnson (a well-capitalized firm that operates in the relatively stable health care market) and RadioShack (a firm that is active in the highly volatile consumer electronics business). If we compare the numbers, we see that the EDF for Johnson & Johnson is close to zero whereas the EDF for RadioShack is quite high. This difference reflects the higher leverage of RadioShack and the riskiness of the underlying business, as reflected by the comparatively large asset volatility $\sigma_V = 24\%$. Indeed, on 11 September 2014, the *New York Times* reported that a bankruptcy filing for RadioShack could be near, suggesting that the EDF had good predictive power in this case.

Table 10.3. A summary of the public-firm EDF methodology. The example is taken from Sun, Munves and Hamilton (2012); it is concerned with the situation of Johnson & Johnson (J&J) and RadioShack as of April 2012. All quantities are in US dollars.

Variable	J&J	RadioShack	Notes
Market value of assets V_0	236 bn	1834 m	Determined from time series of equity prices
Asset volatility σ_V	11%	24%	
Default threshold $\tilde{B}$	39 bn	1042 m	Short-term liabilities and half of long-term liabilities
DD	16.4	2.3	Given by $(\log V_0 - \log \tilde{B})/\sigma_V$
EDF (one year)	0.01%	3.58%	Determined using empirical mapping between DD and EDF

10.3.4 Credit-Migration Models Revisited

Recall that in the credit-migration approach each firm is assigned to a credit-rating category at any given time point. There are a finite number of such ratings and they are ordered by credit quality and include the category of default. The probability of moving from one credit rating to another credit rating over the given risk horizon (typically one year) is then specified. In this section we explain how a migration model can be embedded in a firm-value model and thus be treated as a structural model. This will be useful in the discussion of portfolio versions of these models in Chapter 11. Moreover, we compare the public-firm EDF model and credit-migration approaches.

Credit-migration models as firm-value models. We consider a firm that has been assigned to some non-default rating category j at $t = 0$ and for which transition probabilities $p_{jk}, 0 \leqslant k \leqslant n$, over the period $[0, T]$ are available on the basis of that rating. These express the probability that the firm belongs to rating class k at the time horizon T, given that it is in class j at $t = 0$. In particular, $p_{j,0}$ is the default probability of the firm over $[0, T]$.

Suppose that the asset-value process (V_t) of the firm follows the model given in (10.5), so that

$$V_T = V_0 \exp((\mu_V - \tfrac{1}{2}\sigma_V^2)T + \sigma_V W_T) \tag{10.20}$$

is lognormally distributed. We can now choose thresholds

$$0 = \tilde{d}_0 < \tilde{d}_1 < \cdots < \tilde{d}_n < \tilde{d}_{n+1} = \infty \tag{10.21}$$

such that $P(\tilde{d}_k < V_T \leqslant \tilde{d}_{k+1}) = p_{jk}$ for $k \in \{0, \ldots, n\}$. We have therefore translated the transition probabilities into a series of thresholds for an assumed asset-value process. The threshold $\tilde{d}_1$ is the default threshold, which in the Merton model of Section 10.3.1 was interpreted as the value of the firm's liabilities. The higher thresholds are the asset-value levels that mark the boundaries of higher rating categories. The firm-value model in which we have embedded the migration model can be summarized by saying that the firm belongs to rating class k at the time horizon T if and only if $\tilde{d}_k < V_T \leqslant \tilde{d}_{k+1}$.

The migration probabilities in the firm-value model obviously remain invariant under simultaneous strictly increasing transformations of V_T and the thresholds $\tilde{d}_j$. If we define

$$X_T := \frac{\ln V_T - \ln V_0 - (\mu_V - \frac{1}{2}\sigma_V^2)T}{\sigma_V\sqrt{T}}, \tag{10.22}$$

$$d_k := \frac{\ln \tilde{d}_k - \ln V_0 - (\mu_V - \frac{1}{2}\sigma_V^2)T}{\sigma_V\sqrt{T}}, \tag{10.23}$$

then we can also say that the firm belongs to rating class k at the time horizon T if and only if $d_k < X_T \leqslant d_{k+1}$. Observe that X_T is a standardized version of the *asset-value log-return* $\ln V_T - \ln V_0$, and we can easily verify that $X_T = W_T/\sqrt{T}$ so that it has a standard normal distribution. In this case the formulas for the thresholds are easily obtained and are $d_k = \Phi^{-1}(\sum_{l=0}^{k-1} p_{jl})$ for $k = 1, \ldots, n$.

The public-firm EDF model and credit-migration approaches compared. The public-firm EDF model uses market data, most notably the current stock price, as inputs for the EDF computation. The EDF therefore reacts quickly to changes in the economic prospects of a firm, as these are reflected in the firm's share price and hence in the estimated distance-to-default. Moreover, EDFs are quite sensitive to the current macroeconomic environment. The distance-to-default is observed to rise in periods of economic expansion (essentially due to higher share prices reflecting better economic conditions) and to decrease in recession periods. Rating agencies, on the other hand, are typically slow in adjusting their credit ratings, so that the current rating does not always reflect the economic condition of a firm. This is particularly important if the credit quality of a firm deteriorates rapidly, as is typically the case with companies that are close to default. For instance, the investment bank Lehman Brothers had a fairly good rating (Aa or better) when it defaulted in September 2008. EDFs might therefore be better predictors of default probabilities over short time horizons.

On the other hand, the public-firm EDF model is quite sensitive to global over- and underreaction of equity markets. In particular, the bursting of a stock market bubble may lead to drastically increased EDFs, even if the economic outlook for a given corporation has not changed very much. This can lead to huge fluctuations in the amount of risk capital that is required to back a given credit portfolio. From this point of view the relative inertia of ratings-based models could be considered an advantage, as the ensuing risk capital requirements tend to be more stable over time.

Notes and Comments

There are many excellent texts, at varying technical levels, in which the basic mathematical finance results used in Section 10.3.2 can be found. Models in discrete time are discussed in Cox and Rubinstein (1985) and Jarrow and Turnbull (1999); excellent introductions to continuous-time models include Baxter and Rennie (1996), Björk (2004), Bingham and Kiesel (2004) and Shreve (2004b).

Lando (2004) gives a good overview of the rich literature on firm-value models. First-passage-time models have been considered by, among others, Black and Cox (1976) and, in a set-up with stochastic interest rates, Longstaff and Schwartz (1995). The problem of the unrealistically low credit spreads for small maturities $\tau = T - t$, which we pointed out in Figure 10.4, has also led to extensions of Merton's model. Partial remedies within the class of firm-value models include models with jumps in the firm value, as in Zhou (2001), time-varying default thresholds, as in Hull and White (2001), stochastic volatility models for the firm-value process with time-dependent dynamics, as in Overbeck and Schmidt (2005), and incomplete information on firm value or default threshold, as in Duffie and Lando (2001), Frey and Schmidt (2009) and Cetin (2012). Models with endogenous default thresholds have been considered by, among others, Leland (1994), Leland and Toft (1996) and Hilberink and Rogers (2002).

Duffie and Lando (2001) established a relationship between firm-value models and reduced-form models in continuous time. Essentially, they showed that, from the perspective of investors with *incomplete accounting information* (i.e. incomplete information about the assets or liabilities of a firm), a firm-value model becomes a reduced-form model. A less technical discussion of these issues can be found in Jarrow and Protter (2004).

The public-firm EDF model was first described in Crosbie and Bohn (2002); the model variant that is currently in use is described in Dwyer and Qu (2007) and Sun, Munves and Hamilton (2012).

10.4 Bond and CDS Pricing in Hazard Rate Models

Hazard rate models are the most basic reduced-form credit risk models and are therefore a natural starting point for our discussion of this model class. Moreover, hazard rate models are used as an input in the construction of the popular copula models for portfolio credit derivatives. For these reasons this section is devoted to bond and CDS pricing in hazard rate models. We begin by introducing the necessary mathematical background in Section 10.4.1. Since the pricing results that we present in this section rely on the concept of risk-neutral pricing and martingale modelling, we briefly review these notions in Section 10.4.2. The pricing of bonds and CDSs and some of the related empirical evidence is discussed in Sections 10.4.3, 10.4.4 and 10.4.5.

10.4.1 Hazard Rate Models

A hazard rate model is a model in which the distribution of the default time of an obligor is directly specified by a hazard function without modelling the mechanism by which default occurs.

To set up a hazard rate model we consider a probability space $(\Omega, \mathcal{F}, P)$ and a random time τ defined on this space, i.e. an $\mathcal{F}$-measurable rv taking values in $[0, \infty]$. In economic terms, τ can be interpreted as the default time of some company. We denote the df of τ by $F(t) = P(\tau \leqslant t)$ and the tail or survival function by

$\bar{F}(t) = 1 - F(t)$; we assume that $P(\tau = 0) = F(0) = 0$, and that $\bar{F}(t) > 0$ for all $t < \infty$. We define the *jump* or *default indicator process* (Y_t) associated with τ by

$$Y_t = I_{\{\tau \leqslant t\}}, \quad t \geqslant 0. \tag{10.24}$$

Note that (Y_t) is a right-continuous process that jumps from 0 to 1 at the default time τ and that $1 - Y_t = I_{\{\tau > t\}}$ is the *survival indicator* of the firm.

Definition 10.4 (cumulative hazard function and hazard function). The function $\Gamma(t) := -\ln(\bar{F}(t))$ is called the *cumulative hazard function* of the random time τ. If F is absolutely continuous with density f, the function $\gamma(t) := f(t)/(1 - F(t)) = f(t)/\bar{F}(t)$ is called the *hazard function* of τ.

By definition we have $\bar{F}(t) = \mathrm{e}^{-\Gamma(t)}$. If F has density f, we calculate that $\Gamma'(t) = f(t)/\bar{F}(t) = \gamma(t)$, so that we can represent the survival function of τ in terms of the hazard function by

$$\bar{F}(t) = \exp\left(-\int_0^t \gamma(s)\,\mathrm{d}s\right). \tag{10.25}$$

The hazard function $\gamma(t)$ at a fixed time t gives the *hazard rate* at t, which can be interpreted as a measure of the instantaneous risk of default at t, given survival up to time t. In fact, for $h > 0$ we have $P(\tau \leqslant t + h \mid \tau > t) = (F(t+h) - F(t))/\bar{F}(t)$. Hence we obtain

$$\lim_{h \to 0} \frac{1}{h} P(\tau \leqslant t + h \mid \tau > t) = \frac{1}{\bar{F}(t)} \lim_{h \to 0} \frac{F(t+h) - F(t)}{h} = \gamma(t).$$

For illustrative purposes we determine the hazard function for the Weibull distribution. This is a popular distribution for survival times with df $F(t) = 1 - \mathrm{e}^{-\lambda t^\alpha}$ for parameters $\lambda, \alpha > 0$. For $\alpha = 1$ the Weibull distribution reduces to the standard exponential distribution. Differentiation yields

$$f(t) = \lambda\alpha t^{\alpha-1}\mathrm{e}^{-\lambda t^\alpha} \quad \text{and} \quad \gamma(t) = \lambda\alpha t^{\alpha-1}.$$

In particular, γ is decreasing in t if $\alpha < 1$ and increasing if $\alpha > 1$. For $\alpha = 1$ (exponential distribution) the hazard rate is time independent and equal to the parameter λ.

Filtrations and conditional expectations. In financial models, filtrations are used to model the information available to investors at various points in time. Formally, a *filtration* $(\mathcal{F}_t)$ on $(\Omega, \mathcal{F})$ is an increasing family $\{\mathcal{F}_t : t \geqslant 0\}$ of sub-σ-algebras of $\mathcal{F}$: $\mathcal{F}_t \subset \mathcal{F}_s \subset \mathcal{F}$ for $0 \leqslant t \leqslant s < \infty$. The σ-algebra $\mathcal{F}_t$ represents the state of knowledge of an observer at time t, and $A \in \mathcal{F}_t$ is taken to mean that at time t the observer is able to determine if the event A has occurred.

In this section it is assumed that the only quantity that is observable for investors is the default event of the firm under consideration or, equivalently, the default indicator process (Y_t) associated with τ. The appropriate filtration is therefore given by $(\mathcal{H}_t)$ with

$$\mathcal{H}_t = \sigma(\{Y_u : u \leqslant t\}), \tag{10.26}$$

the *default history* up to and including time t. By definition, τ is an $(\mathcal{H}_t)$ stopping time, as $\{\tau \leqslant t\} = \{Y_t = 1\} \in \mathcal{H}_t$ for all $t \geqslant 0$; moreover, $(\mathcal{H}_t)$ is obviously the smallest filtration with this property.

In order to study bond and CDS pricing in hazard rate models we need to compute conditional expectations with respect to the σ-algebra $\mathcal{H}_t$. We begin our analysis of this issue with an auxiliary result on the structure of $\mathcal{H}_t$-measurable rvs. The result formalizes the fact that every $\mathcal{H}_t$-measurable rv can be expressed as a function of events related to the default history at t.

Lemma 10.5. *Every $\mathcal{H}_t$-measurable rv H is of the form $H = h(\tau)I_{\{\tau \leqslant t\}} + cI_{\{\tau > t\}}$ for a measurable function $h\colon [0,t] \to \mathbb{R}$ and some constant $c \in \mathbb{R}$.*

Proof. The σ-algebra $\mathcal{H}_t$ is generated by the events $\{Y_u = 1\} = \{\tau \leqslant u\}$, $u < t$, and $\{Y_t = 0\} = \{\tau > t\}$, and hence by the rvs $(\tau \wedge t) := \min\{\tau, t\}$ and $I_{\{\tau > t\}}$. This implies that any $\mathcal{H}_t$-measurable rv H can be written as $H = g(\tau \wedge t, I_{\{\tau > t\}})$ for some measurable function $g\colon [0,t] \times \{0,1\} \to \mathbb{R}$. The claim follows if we define $h(u) := g(u,0)$, $u \leqslant t$, and $c := g(t,1)$. □

Lemma 10.6. *Let τ be a random time with jump indicator process $Y_t = I_{\{\tau \leqslant t\}}$ and natural filtration $(\mathcal{H}_t)$. Then, for any integrable rv X and any $t \geqslant 0$, we have*

$$E(I_{\{\tau > t\}} X \mid \mathcal{H}_t) = I_{\{\tau > t\}} \frac{E(I_{\{\tau > t\}} X)}{P(\tau > t)}. \tag{10.27}$$

Proof. Since $E(I_{\{\tau > t\}} X \mid \mathcal{H}_t)$ is $\mathcal{H}_t$-measurable and zero on $\{\tau \leqslant t\}$, we obtain from Lemma 10.5 that $E(I_{\{\tau > t\}} X \mid \mathcal{H}_t) = I_{\{\tau > t\}} c$ for some constant c. Taking expectations yields $E(I_{\{\tau > t\}} X) = cP(\tau > t)$ and hence $c = E(I_{\{\tau > t\}} X)/P(\tau > t)$. □

Lemma 10.6 can be used to determine conditional survival probabilities. Fix $t < T$ and consider the quantity $P(\tau > T \mid \mathcal{H}_t)$. Applying (10.27) with $X := I_{\{\tau > T\}}$ yields

$$P(\tau > T \mid \mathcal{H}_t) = E(X \mid \mathcal{H}_t) = E(I_{\{\tau > t\}} X \mid \mathcal{H}_t) = I_{\{\tau > t\}} \frac{\bar{F}(T)}{\bar{F}(t)}. \tag{10.28}$$

If τ admits the hazard function $\gamma(t)$, we get the important formula

$$P(\tau > T \mid \mathcal{H}_t) = I_{\{\tau > t\}} \exp\left(- \int_t^T \gamma(s)\, \mathrm{d}s \right), \quad t < T. \tag{10.29}$$

The next proposition is concerned with stochastic process properties of the jump indicator process of a random time τ.

Proposition 10.7. *Let τ be a random time with absolutely continuous df F and hazard function γ. Then $M_t := Y_t - \int_0^t I_{\{\tau > u\}} \gamma(u)\, \mathrm{d}u$, $t \geqslant 0$, is an $(\mathcal{H}_t)$-martingale: that is, $E(M_s \mid \mathcal{H}_t) = M_t$ for all $0 \leqslant t \leqslant s < \infty$.*

In Section 10.5.1 we extend this result to doubly stochastic random times and discuss its financial and mathematical relevance.

Proof. Let $s > t$. We have to show that $E(M_s - M_t \mid \mathcal{H}_t) = 0$, i.e. that $E(Y_s - Y_t \mid \mathcal{H}_t) = E(\int_t^s \gamma(u) I_{\{u<\tau\}}\,\mathrm{d}u \mid \mathcal{H}_t)$. Using (10.28) we get

$$E(Y_s - Y_t \mid \mathcal{H}_t) = I_{\{\tau>t\}} P(\tau \leqslant s \mid \mathcal{H}_t) = I_{\{\tau>t\}}\left(1 - \frac{\bar{F}(s)}{\bar{F}(t)}\right)$$
$$= I_{\{\tau>t\}} \frac{\bar{F}(t) - \bar{F}(s)}{\bar{F}(t)}.$$

Note that $X := \int_t^s \gamma(u) I_{\{u<\tau\}}\,\mathrm{d}u$ is 0 on $\{\tau \leqslant t\}$, so $X = X I_{\{\tau>t\}}$. Hence we obtain from Lemma 10.6, the Fubini Theorem and the identity $\bar{F}'(t) = -f(t) = -\gamma(t)\bar{F}(t)$ that

$$E(X \mid \mathcal{H}_t) = I_{\{\tau>t\}} \frac{E(X)}{\bar{F}(t)} = I_{\{\tau>t\}} \frac{\int_t^s \gamma(u)\bar{F}(u)\,\mathrm{d}u}{\bar{F}(t)} = I_{\{\tau>t\}} \frac{\bar{F}(t) - \bar{F}(s)}{\bar{F}(t)},$$

and the result follows. □

10.4.2 Risk-Neutral Pricing Revisited

The remainder of Section 10.4 is devoted to an analysis of risk-neutral pricing results for credit products in hazard rate models. Risk-neutral pricing has become so popular that the conceptual underpinnings are often overlooked. A prime case in point is the mechanical use of the Gauss copula to price CDO tranches, a practice that led to well-documented problems during the 2007–9 financial crisis. It therefore seems appropriate to clarify the applicability and the limitations of risk-neutral pricing in the context of credit risk models.

Risk-neutral pricing. We build on the elementary discussion of risk-neutral valuation given in Section 2.2.2. In that section we considered a simple one-period default model for a defaultable zero-coupon bond with maturity T equal to one year and a deterministic recovery rate $1 - \delta$ equal to 60%. Moreover, we assumed that the real-world default probability was $p = 1\%$, the risk-free simple interest rate was $r_{0,1} = 5\%$, and the market price of the bond at $t = 0$ was $p_1(0, 1) = 0.941$.

Risk-neutral pricing is intimately linked to the notion of a risk-neutral measure. In general terms a risk-neutral measure is an artificial probability measure Q, equivalent to the historical measure P, such that the discounted prices of all traded securities are Q-martingales (fair bets). We have seen in Section 2.2.2 that in the simple one-period default model a risk-neutral measure Q is simply given by an artificial default probability q such that

$$p_1(0, 1) = (1.05)^{-1}((1 - q) \cdot 1 + q \cdot 0.6).$$

Obviously, q is uniquely determined by this equation and is given by $q = 0.03$. Note that in this example the risk-neutral default probability q is higher than the real-world default probability p. This reflects risk aversion on the part of investors and is typical for real markets; empirical evidence on the relationship between the physical and historical default probabilities will be presented in Section 10.4.5.

The *risk-neutral pricing rule* states that the price of a derivative security can be computed as the mathematical expectation of the discounted pay-off under a risk-neutral measure Q. In mathematical terms the price at time t of a derivative with pay-off H and maturity $T \geqslant t$ is thus given by

$$V_t^H = E^Q\left(\exp\left(-\int_t^T r_s \, \mathrm{d}s\right) H \;\middle|\; \mathcal{F}_t\right), \tag{10.30}$$

where r_s denotes the continuously compounded default-free short rate of interest at time s and where the σ-algebra $\mathcal{F}_t$ represents the information available to investors at time t (see the discussion of filtrations in Section 10.4.1). Note that in one-period models, (10.30) reduces to the simpler expression $V_0^H = E^Q(H/(1+r_{0,1}))$, where $r_{0,1}$ is the simple interest rate for the period.

There are two theoretical justifications for risk-neutral pricing. One argument is based on absence of arbitrage: according to the first fundamental theorem of asset pricing, a model for security prices is arbitrage free if and only if it admits at least one equivalent martingale measure Q. Hence, if a financial product is to be priced in accordance with no-arbitrage principles, its price must be given by the risk-neutral pricing formula for some risk-neutral measure Q. A second justification relies on hedging: in financial models it is often possible to replicate the pay-off of a financial product by (dynamic) trading in the available assets, and in a frictionless market the cost of carrying out such a hedge is given by the risk-neutral pricing rule.

Hedging and market completeness. Next we take a closer look at the concept of hedging. We work in the simple one-period default model that was introduced in the previous paragraph. Consider an investor, e.g. an investment bank, who plans to sell derivatives on the defaultable zero-coupon bond. For concreteness we consider a *default put option* with maturity date $T = 1$. This contract pays one unit if the bond defaults and zero otherwise; it can be thought of as a simplified version of a CDS. Obviously, the pay-off of the default put is unknown at date $t = 0$ and thus constitutes a risk for the investor. A possible strategy for dealing with this risk is to form a *hedging portfolio* in the defaultable bond and in cash that reduces the risk of selling the put: suppose that at time $t = 0$ we go short 2.5 units of the bond and hold $\frac{50}{21} \approx 2.38$ units of cash. At time $t = 1$ there are two possibilities for the value V_1 of this portfolio.

- Default occurs: in which case $V_1 = (-2.5) \cdot 0.6 + \frac{50}{21} \cdot 1.05 = 1$.
- No default: in which case $V_1 = (-2.5) \cdot 1 + \frac{50}{21} \cdot 1.05 = 0$.

In either case the value V_1 of the hedge portfolio equals the pay-off of the option and we have found a so-called *replicating strategy* for the option. In particular, by forming the replicating strategy the investor completely eliminates the risk from selling the option, and the *law of one price* dictates that the fair price at $t = 0$ of the option should equal the value of the hedge portfolio at $t = 0$ given by $V_0 = (-2.5) \cdot 0.941 + \frac{50}{21} \approx 0.0285$ (otherwise either the buyer or the seller could make some risk-free profit).

To construct the portfolio in this simple one-period, two-state setting we have to consider two linear equations. Denote by ξ_1 and ξ_2 the units of the defaultable bond and the amount of cash in our portfolio. At time $t = 1$ we must have

$$\xi_1 \cdot 0.6 + \xi_2 \cdot 1.05 = 1 \quad \text{(the default case)}, \tag{10.31}$$

$$\xi_1 \cdot 1.0 + \xi_2 \cdot 1.05 = 0 \quad \text{(the no-default case)}, \tag{10.32}$$

which leads to the above values of $\xi_1 = -2.5$ and $\xi_2 = \frac{50}{21}$. In mathematical finance a derivative security is called *attainable* if there is a replicating portfolio strategy in the underlying assets. The above argument shows that in the simple one-period default model with only two states every derivative security is attainable. Such models are termed *complete*.

The fair price of the default put (the initial value V_0 of the replicating portfolio) can alternatively be computed by the risk-neutral pricing rule. Recall that the risk-neutral default probability is given by $q = 0.03$. The risk-neutral pricing rule applied to the default put thus leads to a value of $(1.05)^{-1}(0.97 \cdot 0 + 0.03 \cdot 1) = 0.0285$, which is equal to V_0. This is, of course, not a lucky coincidence; a basic result from mathematical finance states that the fair price of any attainable claim can be computed as the expected value of the discounted pay-off under a risk-neutral measure. Armed with this result, we typically first compute the price (the expected value of the discounted pay-off under a risk-neutral measure) and then determine the replicating strategy. For this reason a lot of research focuses on the problem of computing prices. However, one should bear in mind that the economic justification for the risk-neutral pricing rule stems partially from the hedging argument, which applies only to attainable claims. This issue has, to a large extent, been neglected in the literature on the pricing of credit-risky securities. The next example illustrates some of the difficulties arising in *incomplete* markets, where most derivatives are not attainable.

Example 10.8 (a model with random recovery). As there is a substantial amount of randomness in real recovery rates, it is interesting to study the impact of random recovery rates on the validity of the above pricing arguments. We consider an extension of the basic one-period default model in which the loss given default may be either 30% or 50%. The price is assumed to be $p_1(0, 1) = 0.941$ and the risk-free simple interest rate is assumed to be $r_{0,1} = 5\%$ as before. The evolution of the price $p_1(\cdot, 1)$ is illustrated in Figure 10.5. We leave the physical measure unspecified—we assume only that all three possible outcomes have strictly positive probability.

We begin our analysis of this model by determining the equivalent martingale measures. Let q_1 be the risk-neutral probability that default occurs and the LGD is 0.5, let q_2 be the risk-neutral probability that default occurs and the LGD is 0.3, and let $q_3 = 1 - q_1 - q_2$. It follows that q_1 and q_2 satisfy the equation

$$p_1(0, 1) = 1.05^{-1}(q_1 \cdot 0.5 + q_2 \cdot 0.7 + (1 - q_1 - q_2) \cdot 1), \tag{10.33}$$

with the restrictions that $q_1 > 0$, $q_2 > 0$, $1 - q_1 - q_2 > 0$. Obviously, Q is no longer unique. It is easily seen from (10.33) that the set $\mathcal{Q}$ of equivalent martingale

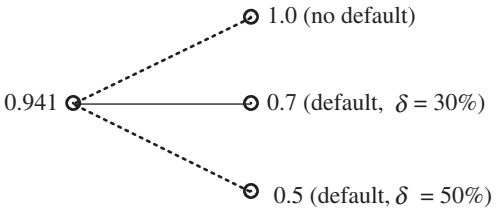

Figure 10.5. Evolution of the price $p_1(\cdot, 1)$ of the defaultable bond in Example 10.8.

measures is given by

$$\mathcal{Q} = \{\boldsymbol{q} \in \mathbb{R}^3 : q_1 \in (0, 0.024),\ q_2 = \tfrac{10}{3}(1 - 1.05 \cdot p_1(0,1) - 0.5 \cdot q_1),\ q_3 = 1 - (q_1 + q_2)\}. \tag{10.34}$$

It is interesting to look at the boundary cases. For $q_1 = 0$ we obtain $q_2 = 4\%$, $q_3 = 96\%$; this is the scenario where the risk-neutral default probability $q = q_1 + q_2$ is maximized. For $q_1 = 2.4\%$ we obtain $q_2 = 0$, $q_3 = 97.6\%$; this is the scenario where q is minimized. Note, however, that the measures $\boldsymbol{q}_0 := (0.024, 0, 0.976)$ and $\boldsymbol{q}_1 := (0, 0.04, 0.96)$ do not belong to $\mathcal{Q}$, as they are not equivalent to the physical measure P.

Consider a derivative security with pay-off H and maturity $T = 1$, such as the default put with pay-off $H = 0$ if $p_1(1, 1) = 1$ (no default) and $H = 1$ otherwise. Every price of the form $H_0 = E^Q(1.05^{-1}H)$ for some $Q \in \mathcal{Q}$ is consistent with no arbitrage and will therefore be called an *admissible value* for the derivative. If $\mathcal{Q}$ contains more than one element, as in our case, there is typically more than one admissible value. For instance, we obtain for the default put option that

$$\inf_{Q\in\mathcal{Q}} E^Q\left(\frac{H}{1.05}\right) \approx 0.023 \quad \text{and} \quad \sup_{Q\in\mathcal{Q}} E^Q\left(\frac{H}{1.05}\right) \approx 0.038; \tag{10.35}$$

obviously, the infimum and supremum in (10.35) correspond to the measures $\boldsymbol{q}_0$ and $\boldsymbol{q}_1$, where q is minimized and maximized, respectively. This non-uniqueness of admissible values reflects the fact that in our three-state model the put is no longer attainable. In fact, the hedging portfolio (ξ_1, ξ_2) now has to solve the following three equations:

$$\left.\begin{aligned} \xi_1 \cdot 0.5 + \xi_2 \cdot 1.05 &= 1 \quad \text{(default, low recovery)},\\ \xi_1 \cdot 0.7 + \xi_2 \cdot 1.05 &= 1 \quad \text{(default, high recovery)},\\ \xi_1 \cdot 1 + \xi_2 \cdot 1.05 &= 0 \quad \text{(no default)}. \end{aligned}\right\} \tag{10.36}$$

It is immediately seen that the system (10.36) of three equations and only two unknowns has no solution, so that the default put is not attainable. This illustrates two fundamental results from modern mathematical finance: a claim with bounded pay-off is attainable if and only if the set of admissible values consists of a single number; an arbitrage-free market is complete if and only if there is exactly one equivalent martingale measure Q. The latter result is known as the *second fundamental theorem of asset pricing*.

Example 10.8 shows that in an incomplete market new issues arise; in particular, it is not obvious how to choose the correct price of a derivative security from the range of admissible values or how to deal with the risk incurred by selling a derivative security. This is unfortunate, as realistic models, which capture the dynamics of financial time series, are typically incomplete. In recent years a number of interesting concepts for the risk management of derivative securities in incomplete markets have been developed. These approaches typically propose mitigating the risk by an appropriate trading strategy and often suggest a pricing formula for the remaining risk. However, a discussion of this work is outside the scope of this book. A brief overview of the existing literature on hedging in (incomplete) credit markets is given in Notes and Comments.

Advantages and limitations of risk-neutral pricing. The risk-neutral pricing approach is a *relative pricing theory*, which explains prices of credit products in terms of observable prices of other securities. If properly applied, it leads to arbitrage-free prices of credit-risky securities, which are consistent with prices quoted in the market. These features make the risk-neutral pricing approach to credit risk the method of choice in an environment where credit risk is actively traded and, in particular, for valuing credit instruments when the market for related products is relatively liquid. On the other hand, since pricing models have to be calibrated to prices of traded credit instruments, they are difficult to apply when we lack sufficient market information. Moreover, in such cases prices quoted using an ad hoc choice of some risk-neutral measure are more or less plucked out of thin air.

This can be contrasted with the more traditional pricing methodology for loans and related credit products, where a loan is taken on the balance sheet if the spread earned on the loan is deemed by the lender to be a sufficient compensation for bearing the default risk of the loan and where the default risk is measured using the real-world measure and historical (default) data. Such an approach is well suited to situations where the market for related credit instruments is relatively illiquid and little or no price information is available; loans to medium or small businesses are a prime example. On the other hand, the traditional pricing methodology does not necessarily lead to prices that are consistent (in the sense of absence of arbitrage) across products or compatible with quoted market prices for credit instruments, so it is less suitable in a trading environment.

Martingale modelling. Recall that, according to the first fundamental theorem of asset pricing, a model for security prices is arbitrage free if and (essentially) only if it admits at least one equivalent martingale measure Q. Moreover, in a complete market, the only thing that matters for the pricing of derivative securities is the Q-dynamics of the traded underlying assets. When building a model for pricing derivatives it is therefore a natural shortcut to model the objects of interest—such as interest rates, default times and the price processes of traded bonds—directly, under some exogenously specified martingale measure Q. In the literature this approach is termed *martingale modelling*.

Martingale modelling is particularly convenient if the value H of the underlying assets at some maturity date T is exogenously given, as in the case of zero-coupon bonds. In that case the price of the underlying asset at time $t < T$ can be computed as the conditional expectation under Q of the discounted value at maturity via the risk-neutral pricing rule (10.30). Model parameters are then determined using the requirement that at time $t = 0$ the price of traded securities, as computed from the model using (10.30), should coincide with the price of those securities as observed in the market; this is known as *calibration* of the model to market data.

Martingale modelling ensures that the resulting model is arbitrage free, which is advantageous if one has to model the prices of many different securities simultaneously. The approach is therefore frequently adopted in default-free term structure models and in reduced-form models for credit-risky securities. Martingale modelling has two drawbacks. First, historical information is, to a large extent, useless in estimating model parameters, as these may change in the transition from the real-world measure to the equivalent martingale measure. Second, as illustrated in Example 10.8, realistic models for pricing credit derivatives are typically incomplete, so that one cannot eliminate all risk by dynamic hedging. In those situations one is interested in the distribution of the remaining risk under the physical measure P, so martingale modelling alone is not sufficient. In summary, the martingale-modelling approach is most suitable in situations where the market for underlying securities is relatively liquid. In that case we have sufficient price information to calibrate our models, and issues of market completeness become less relevant.

10.4.3 Bond Pricing

In this section we discuss the pricing of defaultable zero-coupon bonds in hazard rate models. Note that coupon-paying corporate bonds can be represented as a portfolio of zero-coupon bonds, so our analysis applies to coupon-paying bonds as well.

Recovery models. We begin with a survey of different models for the recovery of defaultable zero-coupon bonds. As in previous sections we denote the price at time t of a defaultable zero-coupon bond with maturity $T \geqslant t$ by $p_1(t, T)$; $p_0(t, T)$ denotes the price of the corresponding default-free zero-coupon bond. The face value of these bonds is always taken to be one. The following recovery models are frequently used in the literature.

(i) *Recovery of Treasury (RT).* The RT model was proposed by Jarrow and Turnbull (1995). Under RT, if default occurs at some point in time $\tau \leqslant T$, the owner of the defaulted bond receives $(1 - \delta_\tau)$ units of the default-free zero-coupon bond $p_0(\cdot, T)$ at time τ, where $\delta_\tau \in [0, 1]$ models the percentage loss given default. At maturity T the holder of the defaultable bond therefore receives the payment $I_{\{\tau > T\}} + (1 - \delta_\tau) I_{\{\tau \leqslant T\}}$.

(ii) *Recovery of Face Value (RF).* Under RF, if default occurs at $\tau \leqslant T$, the holder of the bond receives a (possibly random) recovery payment of size $(1 - \delta_\tau)$

immediately at the default time τ. Note that even with deterministic loss given default $\delta_\tau \equiv \delta$ and deterministic interest rates, the value at maturity of the recovery payment is random as it depends on the exact timing of default.

A further recovery model, the so-called *recovery of market value*, is considered in Section 10.5.3. In real markets, recovery is a complex issue with many legal and institutional features, and all recovery models put forward in the literature are at best a crude approximation of reality. The RF assumption comes closest to legal practice, as debt with the same seniority is assigned the same (fractional) recovery, independent of the maturity. On the other hand, for "extreme" parameter values (long maturities and high risk-free interest rates), RF may lead to negative credit spreads, as we will see in Section 10.6.3. Moreover, the RF model leads to slightly more involved pricing formulas for defaultable bonds than the RT model. Empirical evidence on recovery rates for loans and bonds is discussed in Section 11.2.3.

Bond pricing. Next we turn to pricing formulas for defaultable bonds. We use martingale modelling and work directly under some martingale measure Q. We assume that under Q the default time τ is a random time with deterministic risk-neutral hazard function $\gamma^Q(t)$. The information available to investors at time t is given by $\mathcal{H}_t = \sigma(\{Y_u : u \leqslant t\})$. We take interest rates and recovery rates to be deterministic; the percentage loss given default is denoted by $\delta \in (0, 1)$, and the continuously compounded interest rate is denoted by $r(t) \geqslant 0$. Note that, in this setting, the price of the default-free zero-coupon bond with maturity $T \geqslant t$ equals

$$p_0(t, T) = \exp\left(-\int_t^T r(s)\,\mathrm{d}s\right).$$

This is the simplest type of model that can be calibrated to a given term structure of default-free interest rates and single-name credit spreads; generalizations allowing for stochastic interest rates, recovery rates and hazard rates will be discussed in Section 10.5.

The actual payments of a defaultable zero-coupon bond can be represented as a combination of a *survival claim* that pays one unit at the maturity date T and a recovery payment in case of default. The survival claim has pay-off $I_{\{\tau>T\}}$. Recall from (10.29) that

$$Q(\tau > T \mid \mathcal{H}_t) = I_{\{\tau>t\}} \exp\left(-\int_t^T \gamma^Q(s)\,\mathrm{d}s\right)$$

and define $R(t) = r(t) + \gamma^Q(t)$. The price of a survival claim at time t then equals

$$\begin{aligned} E^Q(p_0(t, T) I_{\{\tau>T\}} \mid \mathcal{H}_t) &= \exp\left(-\int_t^T r(s)\,\mathrm{d}s\right) Q(\tau > T \mid \mathcal{H}_t) \\ &= I_{\{\tau>t\}} \exp\left(-\int_t^T R(s)\,\mathrm{d}s\right). \end{aligned} \tag{10.37}$$

Note that for $\tau > t$, (10.37) can be viewed as the price of a default-free zero-coupon bond with adjusted interest rate $R(t) > r(t)$. A similar relationship between

defaultable and default-free bond prices can be established in many reduced-form credit risk models.

Under the RT model the value of the recovery payment at the maturity date T of the bond is $(1-\delta)I_{\{\tau\leqslant T\}} = (1-\delta) - (1-\delta)I_{\{\tau>T\}}$. Using (10.37), the value of the recovery payment at time $t < T$ is therefore

$$(1-\delta)p_0(t,T) - (1-\delta)I_{\{\tau>t\}}\exp\left(-\int_t^T (r(s)+\gamma^Q(s))\,\mathrm{d}s\right).$$

Under the RF hypothesis the recovery payment takes the form $(1-\delta)I_{\{\tau\leqslant T\}}$, where the payment occurs directly at time τ. Payments of this form will be referred to as *payment-at-default claims*). The value of the recovery payment at time $t \leqslant T$ therefore equals

$$E^Q\left((1-\delta)I_{\{t<\tau\leqslant T\}}\exp\left(-\int_t^\tau r(s)\,\mathrm{d}s\right)\,\middle|\,\mathcal{H}_t\right).$$

The evaluation of this expression is discussed in the following lemma.

Lemma 10.9. *Suppose that τ is a random time with hazard function $\gamma^Q(t)$, and let $R(t) = r(t) + \gamma^Q(t)$ as before. Then we have the identity*

$$E^Q\left(I_{\{t<\tau\leqslant T\}}\exp\left(-\int_t^\tau r(s)\,\mathrm{d}s\right)\,\middle|\,\mathcal{H}_t\right) = I_{\{\tau>t\}}\int_t^T \gamma^Q(s)\exp\left(-\int_t^s R(u)\,\mathrm{d}u\right)\mathrm{d}s.$$

Proof. Using Lemma 10.6 we get that

$$E^Q\left(I_{\{t<\tau\leqslant T\}}\exp\left(-\int_t^\tau r(s)\,\mathrm{d}s\right)\,\middle|\,\mathcal{H}_t\right) = I_{\{\tau>t\}}\frac{E^Q(I_{\{t<\tau\leqslant T\}}\exp(-\int_t^\tau r(s)\,\mathrm{d}s))}{\exp(-\int_0^t \gamma^Q(s)\,\mathrm{d}s)}. \quad (10.38)$$

Since τ has density

$$\gamma^Q(t)\exp\left(-\int_0^t \gamma^Q(s)\,\mathrm{d}s\right),$$

we have

$$E^Q\left(I_{\{t<\tau\leqslant T\}}\exp\left(-\int_t^\tau r(s)\,\mathrm{d}s\right)\right) = \int_t^T \exp\left(-\int_t^s r(u)\,\mathrm{d}u\right)\gamma^Q(s)\exp\left(-\int_0^s \gamma^Q(u)\,\mathrm{d}u\right)\mathrm{d}s.$$

Substitution of the right-hand side into equation (10.38) gives the result. □

10.4.4 CDS Pricing

The CDS market is among the most liquid markets for credit-risky securities, so the task of building a model using CDS spreads as input is frequently encountered in practice. In this section we therefore discuss CDS pricing and the calibration of hazard rate models to observed CDS spreads.

Pricing. We consider the following CDS contract. We take the notional to be one, so that percentage loss given default and absolute loss given default are the same. The premium payments are due at N points in time $0 < t_1 < \cdots < t_N$. If $\tau > t_k$, the protection buyer pays a premium of size $x^*(t_k - t_{k-1})$ at t_k, where x^* denotes the swap spread. After default, no further premium payments are made. If default occurs before the maturity date t_N of the swap, the protection seller makes a default payment of size δ to the buyer at the default time τ. In a standard CDS the protection buyer pays the protection seller at default the part of the premium that has accrued since the last regular premium payment date; here we ignore these accrued premium payments to simplify the exposition.

We use the same set-up as in the analysis of bond pricing in the previous section. As a first step we price the payments made by the protection buyer (the so-called premium payment leg of the swap) and the payments made by the protection seller (the default payment leg) separately, using a generic risk-neutral hazard function γ^Q and a generic spread x. The price of the premium payment leg at $t < t_N$ (the expected discounted value of the payments) is given by

$$\begin{aligned} V_t^{\text{prem}}(x;\gamma^Q) &= E^Q\bigg(\sum_{k:\,t_k>t} \exp\bigg(-\int_t^{t_k} r(u)\,\mathrm{d}u\bigg) x(t_k - t_{k-1}) I_{\{t_k<\tau\}} \;\bigg|\; \mathcal{H}_t\bigg) \\ &= x \sum_{k:\,t_k>t} p_0(t,t_k)(t_k - t_{k-1}) Q(\tau > t_k \mid \mathcal{H}_t), \end{aligned} \tag{10.39}$$

which is easily computed using the formula

$$Q(\tau > t_k \mid \mathcal{H}_t) = I_{\{\tau>t\}} \exp\bigg(-\int_t^{t_k} \gamma^Q(s)\,\mathrm{d}s\bigg).$$

The default payment leg is a typical payment-at-default claim. Using Lemma 10.9 we obtain

$$\begin{aligned} V_t^{\text{def}}(\gamma^Q) &= E^Q\bigg(\exp\bigg(-\int_t^{\tau} r(u)\,\mathrm{d}u\bigg)\delta I_{\{t<\tau\leqslant t_N\}} \;\bigg|\; \mathcal{H}_t\bigg) \\ &= I_{\{t<\tau\}}\delta \int_t^{t_N} \gamma^Q(s) \exp\bigg(-\int_t^s (r(u) + \gamma^Q(u))\,\mathrm{d}u\bigg)\,\mathrm{d}s. \end{aligned} \tag{10.40}$$

According to market convention the CDS spread x_t^* quoted for the contract at time t (the so-called *fair CDS spread* x^*) is chosen such that the value of the contract is equal to zero. Hence x_t^* is defined by the equation $V_t^{\text{prem}}(x_t^*;\gamma^Q) = V_t^{\text{def}}(\gamma^Q)$, which gives

$$x_t^* = I_{\{t<\tau\}} \frac{\delta \int_t^{t_N} \gamma^Q(s) \exp(-\int_t^s (r(u) + \gamma^Q(u))\,\mathrm{d}u)\,\mathrm{d}s}{\sum_{k:\,t_k>t} p_0(t,t_k)(t_k - t_{k-1}) \exp(-\int_t^{t_k} \gamma^Q(s)\,\mathrm{d}s)}. \tag{10.41}$$

Obviously, x_t^* depends on the hazard function γ^Q, as V_t^{prem} and V_t^{def} depend on γ^Q.

Note that in the pricing argument we have neglected the issue of counterparty risk and, in particular, the possibility that the protection seller might default before the maturity of the CDS. A discussion of counterparty risk for CDS contracts is given in Section 17.2.

Calibration. Assume now that we observe spreads quoted in the market for one or more CDSs on the same reference entity. Under the martingale-modelling approach we have to calibrate our model to the available market information: that is, we have to determine the implied risk-neutral hazard function γ^Q, which ensures that the fair CDS spreads implied by the model equal the spreads that are quoted in the market.

Suppose that the market information at time $t = 0$ consists of the fair spread x^* of one CDS with maturity t_N; the risk-neutral hazard function γ^Q is constant, so that, for all $s \geqslant 0$, $\gamma^Q(s) = \bar{\gamma}^Q$ for some $\bar{\gamma}^Q > 0$, which we refer to as the risk-neutral hazard rate. It follows from (10.39) and (10.40) that the implied risk-neutral hazard rate $\bar{\gamma}^Q$ satisfies the equation

$$x^* \sum_{k=1}^{N} p_0(0, t_k)(t_k - t_{k-1})\mathrm{e}^{-\bar{\gamma}^Q t_k} = \delta \bar{\gamma}^Q \int_0^{t_N} p_0(0, t)\mathrm{e}^{-\bar{\gamma}^Q t}\,\mathrm{d}t. \tag{10.42}$$

Here, the left-hand side equals $V_0^{\mathrm{prem}}(x^*, \bar{\gamma}^Q)$ and the right-hand side is obviously equal to $V_0^{\mathrm{def}}(\bar{\gamma}^Q)$. There is a unique implied risk-neutral hazard rate solving equation (10.42). This may be seen by first noting that $V_0^{\mathrm{prem}}(x^*, \bar{\gamma}^Q)$ is a decreasing function of $\bar{\gamma}^Q$ while $V_0^{\mathrm{def}}(\bar{\gamma}^Q)$ is an increasing function of $\bar{\gamma}^Q$. Moreover, $V_0^{\mathrm{def}}(0) = 0$, so the value of the premium payments exceeds the value of the default payment for small values of $\bar{\gamma}^Q$. On the other hand, as $\bar{\gamma}^Q$ tends to infinity, $V_0^{\mathrm{prem}}(x^*, \bar{\gamma}^Q)$ converges to zero, so $V_0^{\mathrm{prem}}(x^*, \bar{\gamma}^Q) < V_0^{\mathrm{def}}(\bar{\gamma}^Q)$ for large values of $\bar{\gamma}^Q$.

If one observes spreads for several CDSs on the same reference entity but with different maturities, a time-independent risk-neutral hazard function is generally not sufficient to calibrate the model to the observed swap spreads. Instead one typically uses hazard functions $\gamma^Q(t)$ that are piecewise constant or piecewise linear. An exception occurs in the special case where (1) the spread curve is *flat* (i.e. all CDSs on the reference entity have the same spread x^*, independent of the maturity), (2) the risk-free interest rate is constant, and (3) the time points t_k are equally spaced ($t_k - t_{k-1} = \Delta t$ for all k). In that case the implied risk-neutral hazard rate $\bar{\gamma}^Q$ is the solution of equation (10.42) in the case where $N = 1$, that is, the solution of

$$x^* \Delta t p_0(0, \Delta t)\mathrm{e}^{-\bar{\gamma}^Q \Delta t} = \delta \bar{\gamma}^Q \int_0^{\Delta t} \mathrm{e}^{-rt}\mathrm{e}^{-\bar{\gamma}^Q t}\,\mathrm{d}t. \tag{10.43}$$

For Δt relatively small (quarterly or semiannual spread payments), a good approximation to the solution of (10.43) is given by $\bar{\gamma}^Q \approx x^*/\delta$, i.e. by the ratio of the fair swap spread and the percentage loss given default. This approximation is frequently used in practice.

Note, finally, that for most issuers the implied hazard rate is relatively small (of the order of a few percentage points). We therefore have the following approximation for the one-year default probability:

$$Q(\tau \leqslant 1) = 1 - \mathrm{e}^{-\bar{\gamma}^Q} \approx \bar{\gamma}^Q \approx x^*/\delta, \tag{10.44}$$

so the quantity x^*/δ can be viewed as a proxy for the risk-neutral one-year default probability.

10.4.5 P versus Q: Empirical Results

We have now assembled the necessary technical tools to discuss some of the empirical work on the relationship between physical and risk-neutral default probabilities. Understanding this relationship is important; it enables market participants to use information about historical default probabilities in pricing credit-risky securities. Conversely, it allows the use of market quotes for CDSs or defaultable bonds as additional inputs in determining historical default probabilities.

In most empirical studies risk-neutral default probabilities are estimated from credit-spread data for CDSs. By comparing these estimates with estimates of the physical default probability—obtained, for instance, from the public-firm EDF methodology introduced in Section 10.3.3—it is possible to gain some empirical evidence on the relationship between physical and risk-neutral default probabilities in real markets. An extensive empirical study along these lines is found in Berndt et al. (2008). The authors carried out a very detailed regression analysis of the observed spreads for five-year CDSs against five-year EDFs for a large pool of firms. The five-year EDF of a firm with publicly traded stock is an annualized estimate of the physical five-year default probability. The computation of EDFs is described in detail in Section 10.3.3, and annualization is a way of expressing EDFs for different time horizons on a common yearly scale.

Berndt et al. (2008) begin by estimating a linear model for the relationship between the observed swap spread $x^*_{t,i}$ of firm i at date t and the five-year EDF of that firm on the same day, labelled $\mathrm{EDF}_{t,i}$. The model takes the form

$$x^*_{t,i} = \alpha + \beta\,\mathrm{EDF}_{t,i} + \varepsilon_{t,i}, \quad (t,i) \in S, \tag{10.45}$$

where S denotes the set of all time points/firms for which there is an observable EDF–CDS pair. The model was fitted to a sample of 33 912 EDF–CDS observations for a large set of publicly traded US firms in the period December 2000 to December 2004. The estimated coefficients were given by $\alpha = 33$ bp and $\beta = 1.6$; the R^2 was 0.73.

Berndt et al. (2008) propose the following interpretation of this regression result. Their model implies that the fair swap spread x^* of a firm increases by approximately 16 basis points for every 10 basis point increase in the five-year EDF of that firm; neglecting the intercept, we thus have that $x^*_{t,i}/\mathrm{EDF}_{t,i} \approx 1.6$. Assuming a fixed loss given default δ, we may use the quantity $q_{t,i} = x^*_{t,i}/\delta$ as a proxy for the risk-neutral default probability of firm i at time t; moreover, $\mathrm{EDF}_{t,i}$ can be viewed as a proxy for the physical default probability of firm i at time t. The ratio of risk-neutral to historical default probabilities is therefore given approximately by

$$\frac{q_{t,i}}{p_{t,i}} \approx \frac{x^*_{t,i}}{\delta\,\mathrm{EDF}_{t,i}} \approx 1.6\delta^{-1}.$$

With $\delta = 0.75$ we obtain $q_{t,i}/p_{t,i} \approx 2.13$; higher recovery rates, i.e. smaller values of δ, would lead to an even higher estimate for $q_{t,i}/p_{t,i}$. The analysis of Berndt et al. (2008) clearly shows that physical and risk-neutral default probabilities can differ substantially, and care must be taken to distinguish between the two concepts.

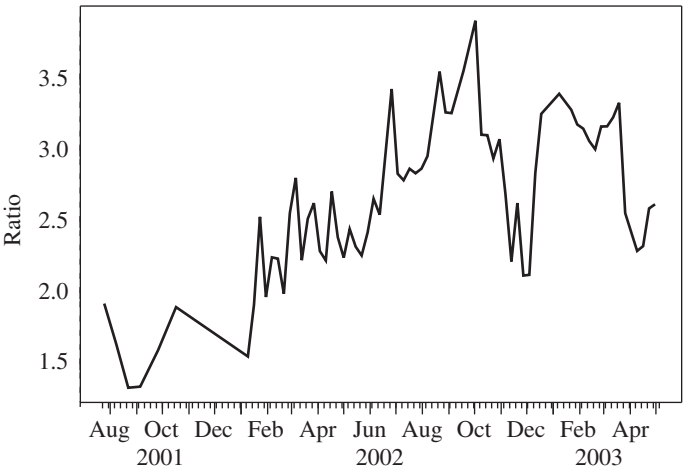

Figure 10.6. Ratio of one-year risk-neutral and historical default probabilities for Vintage Petroleum, as estimated by Berndt et al. (2008).

A careful inspection of the EDF–CDS relationship shows that the simple linear model (10.45) might not be appropriate for a number of reasons. First, the intercept of 33 basis points is implausible, as it would imply that even for a firm with historical default probability p close to zero the swap spread is still of the order of 30 basis points. Second, Berndt et al. (2008) found that the ratio x_i^*/EDF_i varies between industry sectors—reflecting different recovery rates for different industries—and over time, as is illustrated in Figure 10.6. Third, there seems to be some concavity in the relationship between swap spreads and EDFs; in particular, the ratio $x_{t,i}^*/\mathrm{EDF}_{t,i}$ is higher for high-quality firms with low EDF values than for low-quality firms. For these reasons the authors go on to consider more refined logarithmic regression models that fit the data significantly better.

Notes and Comments

Hazard rate models are a common tool in credit risk and survival analysis: see, for example, Bielecki and Rutkowski (2002) or, for a general introduction to survival analysis, the classical textbook by Cox and Oakes (1984). Further useful textbooks are Fleming and Harrington (2005), Marshall and Olkin (2007) and Aalen, Borgan and Gjessing (2010).

The fundamental theorems of asset pricing and the conceptual underpinnings of risk-neutral pricing are discussed in most textbooks on mathematical finance: see, for example, Duffie (2001), Björk (2004), Shreve (2004b) and Delbaen and Schachermayer (2006). The term martingale modelling was coined in Björk (2004) in the context of default-free short-rate models. In recent years a number of interesting approaches to the risk management of derivative securities in incomplete markets have been developed. *Quadratic hedging* approaches were first developed by Föllmer and Sondermann (1986) and Föllmer and Schweizer (1991); Schweizer (2001) is an excellent survey; *utility-based* approaches to pricing and hedging in incomplete markets are discussed in Delbaen et al. (2002) and Becherer (2004), and

the latter paper explicitly considers applications of utility-based hedging strategies to credit risk models. Papers dealing with dynamic hedging and market incompleteness in credit risk models include Bielecki, Jeanblanc and Rutkowski (2004), Bielecki, Jeanblanc and Rutkowski (2007), Frey and Backhaus (2010) and Cont and Kan (2011).

A detailed analysis of CDS pricing can be found in many sources; a good reference is Schönbucher (2003). Theoretical results on the relationship between physical and risk-neutral default probabilities were obtained by Artzner and Delbaen (1995) and Jarrow, Lando and Yu (2005). In their paper, Berndt et al. (2008) go beyond the regression analysis presented in our text and estimate a full time-series model for the joint evolution of risk-neutral and actual default intensities. Further empirical studies of the relationship between actual and risk-neutral default probabilities include Fons (1994), Bohn (2000), Driessen (2005) and Huang and Huang (2012). These results largely corroborate the findings of Berndt et al. (2008).

10.5 Pricing with Stochastic Hazard Rates

In the models with deterministic hazard functions discussed in Section 10.4, the only risk factor affecting a defaultable bond or a CDS is default risk. Hence in these models credit spreads evolve deterministically prior to default, which is clearly unrealistic. Moreover, it is not possible to price options on defaultable bonds or CDSs in such models. In this section we consider models where the hazard function is replaced by a stochastic *hazard process*. In mathematical terms this leads to the notion of doubly stochastic random times, which is discussed in Section 10.5.1. In Section 10.5.2 we derive pricing formulas for certain building blocks that can be used to value many important credit-risky securities. Applications of these formulas are studied in Section 10.5.3.

10.5.1 Doubly Stochastic Random Times

We now consider a situation where additional information affecting the distribution of the random time τ is available. Formally, we represent this additional information by some filtration $(\mathcal{F}_t)$ on the underlying probability space $(\Omega, \mathcal{F}, P)$. In credit risk models this information is typically generated by some background process $(\boldsymbol{\Psi}_t)$ representing, for instance, the risk-free interest rate or various measures of economic activity, so that $\mathcal{F}_t = \sigma(\{\boldsymbol{\Psi}_s : s \leqslant t\})$.

Consider some random time τ on $(\Omega, \mathcal{F}, P)$ with $P(\tau > 0) = 1$ and denote by $Y_t = I_{\{\tau \leqslant t\}}$ the associated jump indicator and by $(\mathcal{H}_t)$ the filtration generated by (Y_t) (see equation (10.26)). We introduce a new filtration $(\mathcal{G}_t)$ by

$$\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{H}_t, \quad t \geqslant 0, \tag{10.46}$$

meaning that $\mathcal{G}_t$ is the smallest σ-algebra that contains $\mathcal{F}_t$ and $\mathcal{H}_t$. We will frequently use the notation $(\mathcal{G}_t) = (\mathcal{F}_t) \vee (\mathcal{H}_t)$ below. The filtration $(\mathcal{G}_t)$ contains information about the background processes and the occurrence or non-occurrence of default up to time t, and thus typically corresponds to the information available

to investors. Obviously, τ is an $(\mathcal{H}_t)$ stopping time and hence also a $(\mathcal{G}_t)$ stopping time. Note, however, that we do not assume that τ is a stopping time with respect to the background filtration $(\mathcal{F}_t)$.

Doubly stochastic random times are a straightforward extension of the models considered in Section 10.4 to the present set-up with additional information.

Definition 10.10 (doubly stochastic random time). A random time τ is said to be doubly stochastic if there exists a positive $(\mathcal{F}_t)$-adapted process (γ_t) such that $\Gamma_t = \int_0^t \gamma_s \, \mathrm{d}s$ is strictly increasing and finite for every $t > 0$ and such that, for all $t \geqslant 0$,

$$P(\tau > t \mid \mathcal{F}_\infty) = \exp\left(-\int_0^t \gamma_s \, \mathrm{d}s\right). \tag{10.47}$$

In that case (γ_t) is referred to as the $(\mathcal{F}_t)$-*conditional hazard process* of τ.

In (10.47) $\mathcal{F}_\infty$ denotes the smallest σ-algebra that contains $\mathcal{F}_t$ for all $t \geqslant 0$: that is, $\mathcal{F}_\infty = \sigma(\bigcup_{t \geqslant 0} \mathcal{F}_t)$. Conditioning on $\mathcal{F}_\infty$ thus means that we know the past and future economic environment and in particular the entire trajectory $(\gamma_s(\omega))_{s \geqslant 0}$ of the hazard rate process. Hence (10.47) implies that, given the economic environment, τ is a random time with deterministic hazard function given by the mapping $s \mapsto \gamma_s(\omega)$. The term *doubly stochastic* obviously refers to the fact that the hazard rate at any time is itself a realization of a stochastic process. In the literature, doubly stochastic random times are also known as *conditional Poisson* or *Cox* random times. Note, finally, that (10.47) implies that $P(\tau \leqslant t \mid \mathcal{F}_\infty)$ is $\mathcal{F}_t$-measurable, so we have the equality

$$P(\tau \leqslant t \mid \mathcal{F}_\infty) = P(\tau \leqslant t \mid \mathcal{F}_t). \tag{10.48}$$

In the next lemma we give an explicit construction of doubly stochastic random times. This construction is very useful for simulation purposes.

Lemma 10.11. *Let X be a standard exponentially distributed rv on $(\Omega, \mathcal{F}, P)$ independent of $\mathcal{F}_\infty$, i.e. $P(X \leqslant t \mid \mathcal{F}_\infty) = 1 - \mathrm{e}^{-t}$ for all $t \geqslant 0$. Let (γ_t) be a positive $(\mathcal{F}_t)$-adapted process such that $\Gamma_t = \int_0^t \gamma_s \, \mathrm{d}s$ is strictly increasing and finite for every $t > 0$. Define the random time τ by*

$$\tau := \Gamma^{\leftarrow}(X) = \inf\{t \geqslant 0 \colon \Gamma_t \geqslant X\}. \tag{10.49}$$

Then τ is doubly stochastic with $(\mathcal{F}_t)$-conditional hazard rate process (γ_t).

Proof. Note that by definition of τ it holds that $\{\tau > t\} = \{\Gamma_t < X\}$. Since Γ_t is $\mathcal{F}_\infty$-measurable and X is independent of $\mathcal{F}_\infty$, we obtain

$$P(\tau > t \mid \mathcal{F}_\infty) = P(\Gamma_t < X \mid \mathcal{F}_\infty) = \mathrm{e}^{-\Gamma_t},$$

which proves the claim. □

Lemma 10.11 has the following converse.

Lemma 10.12. *Let τ be a doubly stochastic random time with $(\mathcal{F}_t)$-conditional hazard process (γ_t). Denote by $\Gamma_t = \int_0^t \gamma_s \, \mathrm{d}s$ the $(\mathcal{F}_t)$-conditional cumulative hazard process of τ and set $X := \Gamma_\tau$. Then the rv X is standard exponentially distributed and independent of $\mathcal{F}_\infty$, and $\tau = \Gamma^{\leftarrow}(X)$ almost surely.*

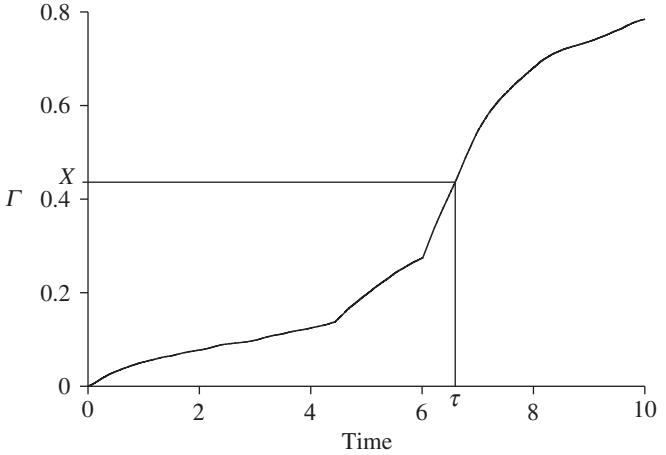

Figure 10.7. A graphical illustration of Algorithm 10.13: $X \approx 0.44$, $\tau \approx 6.59$.

Proof. Since (Γ_t) is strictly increasing by assumption, the relation $\tau = \Gamma^{\leftarrow}(X)$ is clear from the definition of X. To prove that X has the correct distribution we argue as follows:

$$P(X \leqslant t \mid \mathcal{F}_\infty) = P(\Gamma_\tau \leqslant t \mid \mathcal{F}_\infty) = P(\tau \leqslant \Gamma^{\leftarrow}(t) \mid \mathcal{F}_\infty).$$

Since τ is doubly stochastic, the last expression equals $1 - \exp(-\Gamma(\Gamma^{\leftarrow}(t))) = 1 - \mathrm{e}^{-t}$, as Γ is continuous and strictly increasing by assumption. This shows that X is independent of $\mathcal{F}_\infty$ and that it is standard exponentially distributed. □

Lemma 10.11 forms the basis for the following algorithm for the simulation of doubly stochastic random times.

Algorithm 10.13 (univariate threshold simulation).

(1) Generate a trajectory of the hazard process (γ_t). References for suitable simulation approaches are given in Notes and Comments.

(2) Generate a unit exponential rv X independent of (γ_t) (the threshold) and set $\tau = \Gamma^{\leftarrow}(X)$; this step is illustrated in Figure 10.7.

Moreover, Lemmas 10.11 and 10.12 provide an interesting interpretation of doubly stochastic random times in terms of *operational time*. For a given $(\mathcal{F}_t)$-adapted hazard process (γ_t), define a new timescale (operational time) by the associated cumulative hazard process $\Gamma_t = \int_0^t \gamma_s \, \mathrm{d}s$, so that c units of operational time correspond to $\Gamma^{\leftarrow}(c)$ units of real time. Take a standard exponential rv X independent of $\mathcal{F}_\infty$ and measure time in units of operational time. The associated calendar time $\tau := \Gamma^{\leftarrow}(X)$ is then doubly stochastic by Lemma 10.11. Conversely, by Lemma 10.12, if we take a doubly stochastic random time τ, the associated operational time $X := \Gamma_\tau$ is standard exponential, independent of $\mathcal{F}_\infty$. The notion of operational time plays an important role in insurance mathematics (see Section 13.2.7).

Intensity of doubly stochastic random times. We have seen in Proposition 10.7 that the jump indicator process (Y_t) can be turned into an $(\mathcal{H}_t)$-martingale if we subtract the process $\int_0^{t\wedge\tau} \gamma(s)\,\mathrm{d}s$, where $t \wedge \tau$ is a shorthand notation for $\min\{t, \tau\}$. We now generalize this result to doubly stochastic random times.

Proposition 10.14. *Let τ be a doubly stochastic random time with $(\mathcal{F}_t)$-conditional hazard process (γ_t). Then $M_t := Y_t - \int_0^{t\wedge\tau} \gamma_s\,\mathrm{d}s$ is a $(\mathcal{G}_t)$-martingale.*

Proof. Define a new artificial filtration $(\tilde{\mathcal{G}}_t)$ by $\tilde{\mathcal{G}}_t = \mathcal{F}_\infty \vee \mathcal{H}_t$, and note that $\tilde{\mathcal{G}}_0 = \mathcal{F}_\infty$ and $\mathcal{G}_t \subset \tilde{\mathcal{G}}_t$ for all t. As explained above, given $\mathcal{F}_\infty$, τ is a random time with deterministic hazard rate. Proposition 10.7 implies that $M_t := Y_t - \int_0^{t\wedge\tau} \gamma_s\,\mathrm{d}s$ is a martingale with respect to $(\tilde{\mathcal{G}}_t)$. Since (M_t) is $(\mathcal{G}_t)$-adapted and $\mathcal{G}_t \subset \tilde{\mathcal{G}}_t$, (M_t) is also a martingale with respect to $(\mathcal{G}_t)$. □

Finally, we relate Proposition 10.14 to the popular notion of the *intensity* of a random time.

Definition 10.15. Consider a filtration $(\mathcal{G}_t)$ and a random time τ with $(\mathcal{G}_t)$-adapted jump indicator process (Y_t). A non-negative $(\mathcal{G}_t)$-adapted process (λ_t) is called a $(\mathcal{G}_t)$-*intensity* process of the random time τ if $M_t := Y_t - \int_0^{t\wedge\tau} \lambda_s\,\mathrm{d}s$ is a $(\mathcal{G}_t)$-martingale.

In reduced-form credit risk models, (λ_t) is usually called the *default intensity* of the default time τ. It is well known that the intensity (λ_t) is uniquely defined on $\{t < \tau\}$. This is an immediate consequence of general results from stochastic calculus concerning the uniqueness of semimartingale decompositions (see, for example, Chapter 2 of Protter (2005)). Using the terminology of Definition 10.15, we may restate Proposition 10.14 in the following form: "the $(\mathcal{G}_t)$-intensity of a doubly stochastic random time τ is given by its $(\mathcal{F}_t)$-conditional hazard process (γ_t)". At this point a warning is in order: there are random times that admit an intensity in the sense of Definition 10.15 that are not doubly stochastic and for which the pricing formulas derived in Section 10.5.2 below do not hold.

Conditional expectations. Next we discuss the structure of conditional expectations with respect to the full-information σ-algebra $\mathcal{G}_t$; these results are crucial for the derivation of pricing formulas in models with doubly stochastic default times.

Proposition 10.16. *Let τ be an arbitrary random time (not necessarily doubly stochastic) such that $P(\tau > t \mid \mathcal{F}_t) > 0$ for all $t \geqslant 0$. We then have for every integrable rv X that*

$$E(I_{\{\tau>t\}}X \mid \mathcal{G}_t) = I_{\{\tau>t\}} \frac{E(I_{\{\tau>t\}}X \mid \mathcal{F}_t)}{P(\tau > t \mid \mathcal{F}_t)}.$$

Note that Proposition 10.16 allows us to replace certain conditional expectations with respect to $\mathcal{G}_t$ by conditional expectations with respect to the background information $\mathcal{F}_t$. The result is also known as the *Dellacherie formula*. In the special case where the background filtration is trivial, i.e. $\mathcal{F}_t = \{\emptyset, \Omega\}$ for all $t \geqslant 0$, Proposition 10.16 reduces to Lemma 10.6.

Proof. Standard measure-theoretic arguments show that for every $\mathcal{G}_t$-measurable rv X there is some $\mathcal{F}_t$-measurable rv $\tilde{X}$ such that $XI_{\{\tau>t\}} = \tilde{X}I_{\{\tau>t\}}$. This is quite intuitive since prior to default all information is generated by the background filtration $(\mathcal{F}_t)$; a formal proof is given in Section 5.1.1 of Bielecki and Rutkowski (2002). Now $E(I_{\{\tau>t\}}X \mid \mathcal{G}_t)$ is $\mathcal{G}_t$-measurable and zero on $\{\tau \leqslant t\}$. There is therefore an $\mathcal{F}_t$-measurable rv $\tilde{Z}$ such that $E(I_{\{\tau>t\}}X \mid \mathcal{G}_t) = I_{\{\tau>t\}}\tilde{Z}$. Taking conditional expectations with respect to $\mathcal{F}_t$ and noting that $\mathcal{F}_t \subset \mathcal{G}_t$ yields

$$E(I_{\{\tau>t\}}X \mid \mathcal{F}_t) = P(\tau > t \mid \mathcal{F}_t)\tilde{Z}.$$

Hence $\tilde{Z} = E(I_{\{\tau>t\}}X \mid \mathcal{F}_t)/P(\tau > t \mid \mathcal{F}_t)$, which proves the lemma. □

Corollary 10.17. *Let $T > t$ and assume that τ is doubly stochastic with hazard process (γ_t). If the rv $\tilde{X}$ is integrable and $\mathcal{F}_T$-measurable, we have*

$$E(I_{\{\tau>T\}}\tilde{X} \mid \mathcal{G}_t) = I_{\{\tau>t\}}E\left(\exp\left(-\int_t^T \gamma_s \,\mathrm{d}s\right)\tilde{X} \;\middle|\; \mathcal{F}_t\right).$$

Proof. Let $X := I_{\{\tau>T\}}\tilde{X}$. Since $X = I_{\{\tau>t\}}X$ (as $T > t$), Proposition 10.16 yields

$$E(I_{\{\tau>T\}}\tilde{X} \mid \mathcal{G}_t) = E(I_{\{\tau>t\}}X \mid \mathcal{G}_t) = I_{\{\tau>t\}}\mathrm{e}^{\int_0^t \gamma_s \,\mathrm{d}s}E(I_{\{\tau>T\}}\tilde{X} \mid \mathcal{F}_t),$$

where we have used the fact that

$$P(\tau > t \mid \mathcal{F}_t) = \exp\left(-\int_0^t \gamma_s \,\mathrm{d}s\right).$$

Since $\tilde{X}$ is $\mathcal{F}_T$-measurable,

$$E(I_{\{\tau>T\}}\tilde{X} \mid \mathcal{F}_t) = E(\tilde{X}P(\tau > T \mid \mathcal{F}_T) \mid \mathcal{F}_t) = E\left(\tilde{X}\exp\left(-\int_0^T \gamma_s \,\mathrm{d}s\right) \;\middle|\; \mathcal{F}_t\right),$$

and the result follows. □

Corollary 10.17 will be very useful for the pricing of various credit-risky securities in models with doubly stochastic default times. Moreover, the corollary implies that in the above setting γ_t gives a good approximation to the one-year default probability. This follows by setting $T = t + 1$ and $\tilde{X} = 1$ to obtain

$$P(\tau > t + 1 \mid \mathcal{G}_t) = I_{\{\tau>t\}}E\left(\exp\left(-\int_t^{t+1} \gamma_s \,\mathrm{d}s\right) \;\middle|\; \mathcal{F}_t\right). \tag{10.50}$$

Now assume that $\tau > t$ and that the hazard rate remains relatively stable over the time interval $[t, t + 1]$. Under these assumptions the right-hand side of (10.50) is approximated reasonably well by $\mathrm{e}^{-\gamma_t}$ and, if γ_t is small, the one-year default probability satisfies

$$P(\tau \leqslant t + 1 \mid \mathcal{G}_t) \approx 1 - \mathrm{e}^{-\gamma_t} \approx \gamma_t. \tag{10.51}$$

10.5.2 Pricing Formulas

The main result of this section concerns the pricing of three types of contingent claims that can be used as building blocks for constructing the pay-off of many important credit-risky securities. We will show that, for a default time that is doubly stochastic, the computation of prices for these claims can be reduced to a pricing problem for a corresponding default-free claim if we adjust the interest rate and replace the default-free interest rate r_t by the sum $R_t = r_t + \gamma_t$ of the default-free interest rate and the hazard rate of the default time.

The model. We consider a firm whose default time is given by a doubly stochastic random time as in Section 10.5.1. The economic background filtration represents the information generated by an arbitrage-free and complete model for non-defaultable security prices. More precisely, let $(\Omega, \mathcal{F}, (\mathcal{F}_t), Q)$ denote a filtered probability space, where Q is the equivalent martingale measure. Prices of default-free securities such as default-free bonds and the default-free rate of interest (r_t) are $(\mathcal{F}_t)$-adapted processes; $B_t = \exp(\int_0^t r_s \, \mathrm{d}s)$ models the default-free savings account.

Let τ be the default time of some company under consideration and let $Y_t = I_{\{\tau \leqslant t\}}$ be the associated default indicator process. As before we set $\mathcal{H}_t = \sigma(\{Y_s : s \leqslant t\})$ and $\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{H}_t$; we assume that default is observable and that investors have access to the information contained in the background filtration $(\mathcal{F}_t)$, so that the information available to investors at time t is given by $\mathcal{G}_t$. We consider a market for credit products that is liquid enough that we may use the martingale-modelling approach, and we use Q as the pricing measure for defaultable securities. According to (10.30), the price at time t of an arbitrary, non-negative, $\mathcal{G}_T$-measurable contingent claim H is therefore given by

$$H_t = E^Q\left(\exp\left(-\int_t^T r_s \, \mathrm{d}s\right) H \;\middle|\; \mathcal{G}_t\right). \tag{10.52}$$

Finally, we assume that, under Q, the default time τ is a doubly stochastic random time with background filtration $(\mathcal{F}_t)$ and hazard process (γ_t). This latter assumption is crucial for the results that follow.

Definition 10.18. We introduce the following building blocks.

(i) A *survival claim*, i.e. an $\mathcal{F}_T$-measurable promised payment X that is made at time T if there is no default; the actual payment of the survival claim equals $X I_{\{\tau > T\}}$.

(ii) A *risky dividend stream*. Here, we consider a promised dividend stream given by the $(\mathcal{F}_t)$-adapted rate process ν_s, $0 \leqslant s \leqslant T$. The payments of a risky dividend stream stop when default occurs, so that the actual payments of this building block are given by the dividend stream with rate $\nu_t I_{\{\tau > t\}}$, $0 \leqslant t \leqslant T$.

(iii) A *payment-at-default claim* of the form $Z_\tau I_{\{\tau \leqslant T\}}$, where $Z = (Z_t)_{t \geqslant 0}$ is an $(\mathcal{F}_t)$-adapted stochastic process and where Z_τ is short for $Z_{\tau(\omega)}(\omega)$. Note that the payment is made directly at τ, provided that $\tau \leqslant T$, where T is the maturity date of the claim.

Recall from Section 10.4.3 that defaultable bonds can be viewed as portfolios of survival claims and payment-at-default claims. Credit default swaps can also be written as a combination of these claims, as will be shown in Section 10.5.3. A further example is provided by option contracts that are subject to counterparty risk. For concreteness we consider a call option on some default-free security (S_t). Denote the exercise price by K and the maturity date by T and suppose that if the writer defaults at time $\tau \leqslant T$, then the owner of the option receives a fraction $(1 - \delta_\tau)$ of the intrinsic value of the option at the time of default. This can be modelled as a combination of the survival claim $(S_T - K)^+ I_{\{\tau > T\}}$ and the payment-at-default claim $(1 - \delta_\tau)(S_\tau - K)^+ I_{\{\tau \leqslant T\}}$.

Pricing results. In the following theorem we show that the pricing of the building blocks introduced in Definition 10.18 can be reduced to a pricing problem in a default-free security market model with investor information given by the background filtration $(\mathcal{F}_t)$ and with adjusted default-free interest rate.

Theorem 10.19. *Suppose that, under Q, τ is doubly stochastic with background filtration $(\mathcal{F}_t)$ and hazard process (γ_t). Define $R_s := r_s + \gamma_s$. Assume that the rvs $\exp(-\int_t^T r_s \,\mathrm{d}s)|X|$, $\int_t^T |\nu_s| \exp(-\int_t^s r_u \,\mathrm{d}u)\,\mathrm{d}s$ and $\int_t^T |Z_s \gamma_s| \exp(-\int_t^s R_u \,\mathrm{d}u)\,\mathrm{d}s$ are all integrable with respect to Q. Then the following identities hold:*

$$E^Q\left(\exp\left(-\int_t^T r_s \,\mathrm{d}s\right) I_{\{\tau > T\}} X \;\middle|\; \mathcal{G}_t\right) = I_{\{\tau > t\}} E^Q\left(\exp\left(-\int_t^T R_s \,\mathrm{d}s\right) X \;\middle|\; \mathcal{F}_t\right), \tag{10.53}$$

$$E^Q\left(\int_t^T \nu_s I_{\{\tau > s\}} \exp\left(-\int_t^s r_u \,\mathrm{d}u\right) \mathrm{d}s \;\middle|\; \mathcal{G}_t\right) = I_{\{\tau > t\}} E^Q\left(\int_t^T \nu_s \exp\left(-\int_t^s R_u \,\mathrm{d}u\right) \mathrm{d}s \;\middle|\; \mathcal{F}_t\right), \tag{10.54}$$

$$E^Q\left(I_{\{t < \tau \leqslant T\}} \exp\left(-\int_t^\tau r_s \,\mathrm{d}s\right) Z_\tau \;\middle|\; \mathcal{G}_t\right) = I_{\{\tau > t\}} E^Q\left(\int_t^T Z_s \gamma_s \exp\left(-\int_t^s R_u \,\mathrm{d}u\right) \mathrm{d}s \;\middle|\; \mathcal{F}_t\right). \tag{10.55}$$

Proof. The integrability conditions ensure that all conditional expectations are well defined. We start with the pricing formula (10.53) for the vulnerable claim. Define the $\mathcal{F}_T$-measurable rv $\tilde{X} := \exp(-\int_t^T r_s \,\mathrm{d}s) X$. Using Corollary 10.17 with $s = T$ and $\Gamma_t = \int_0^t \gamma_s \,\mathrm{d}s$ we find that

$$E^Q(\tilde{X} I_{\{\tau > T\}} \mid \mathcal{G}_t) = I_{\{\tau > t\}} E^Q(\exp(-(\Gamma_T - \Gamma_t))\tilde{X} \mid \mathcal{F}_t).$$

Noting that $\Gamma_T - \Gamma_t = \int_t^T \gamma_s \,\mathrm{d}s$ and using the definition of $\tilde{X}$, it follows that the right-hand side equals $I_{\{\tau > t\}} E^Q(\exp(-\int_t^T R_s \,\mathrm{d}s) X \mid \mathcal{F}_t)$. The pricing formula (10.54) follows from (10.53) and the Fubini Theorem for conditional expectations. Finally,

we turn to (10.55). Lemma 10.16 implies that

$$E^Q\left(I_{\{\tau>t\}}\exp\left(-\int_t^\tau r_s\,\mathrm{d}s\right)Z_\tau I_{\{\tau\leqslant T\}}\;\middle|\;\mathcal{G}_t\right)$$
$$= I_{\{\tau>t\}}\frac{E^Q(I_{\{\tau>t\}}\exp(-\int_t^\tau r_s\,\mathrm{d}s)Z_\tau I_{\{\tau\leqslant T\}}\mid\mathcal{F}_t)}{P(\tau>t\mid\mathcal{F}_t)}. \quad (10.56)$$

Now note that

$$P(\tau\leqslant t\mid\mathcal{F}_T)=1-\exp\left(-\int_0^t\gamma_s\,\mathrm{d}s\right),$$

so the conditional density of τ given $\mathcal{F}_T$ equals $f_{\tau|\mathcal{F}_T}(t)=\gamma_t\exp(-\int_0^t\gamma_s\,\mathrm{d}s)$. Hence

$$E^Q\left(I_{\{\tau>t\}}\exp\left(-\int_t^\tau r_s\,\mathrm{d}s\right)Z_\tau I_{\{\tau\leqslant T\}}\;\middle|\;\mathcal{F}_T\right)$$
$$=\int_t^T\exp\left(-\int_t^s r_u\,\mathrm{d}u\right)Z_s\gamma_s\exp\left(-\int_0^s\gamma_u\,\mathrm{d}u\right)\mathrm{d}s$$
$$=\exp\left(-\int_0^t\gamma_u\,\mathrm{d}u\right)\int_t^T Z_s\gamma_s\exp\left(-\int_t^s R_u\,\mathrm{d}u\right)\mathrm{d}s.$$

Using iterated conditional expectations we obtain the formula

$$E^Q\left(I_{\{\tau>t\}}\exp\left(-\int_t^\tau r_s\,\mathrm{d}s\right)Z_\tau I_{\{\tau\leqslant T\}}\;\middle|\;\mathcal{F}_t\right)$$
$$=\exp\left(-\int_0^t\gamma_u\,\mathrm{d}u\right)E^Q\left(\int_t^T Z_s\gamma_s\exp\left(-\int_t^s R_u\,\mathrm{d}u\right)\mathrm{d}s\;\middle|\;\mathcal{F}_t\right),$$

and the identity (10.55) follows from (10.56). □

10.5.3 Applications

Credit default swaps. We extend our analysis of Section 10.4.4 and discuss the pricing of CDSs in models where the default time is doubly stochastic. This allows us to incorporate stochastic interest rates, recovery rates and hazard rates into the analysis.

We quickly recall the form of the payments of the CDS contract. As in our previous analysis, the premium payments are due at N points in time $0<t_1<\cdots<t_N$; at a pre-default date t_k, the protection buyer pays a premium of size $x(t_k-t_{k-1})$, where x denotes the swap spread in percentage points (again we take the nominal of the swap to be one). If $\tau\leqslant t_N$, the protection seller makes a default payment of size δ_τ to the buyer at the default time τ, where the percentage loss given default is now a general $(\mathcal{F}_t)$-adapted process. Using Theorem 10.19, both legs of the swap can be priced. The regular premium payments constitute a sequence of survival claims.

Using (10.53) the fair price of the premium leg at $t = 0$ is

$$V^{\text{prem},1} = \sum_{k=1}^{N} E^Q\left(\exp\left(-\int_0^{t_k} r_u \, \mathrm{d}u\right) x(t_k - t_{k-1}) I_{\{t_k < \tau\}}\right)$$
$$= x \sum_{k=1}^{N} (t_k - t_{k-1}) E^Q\left(\exp\left(-\int_0^{t_k} R_u \, \mathrm{d}u\right)\right).$$

The default payment leg is a payment-at-default claim with $Z_s = \delta_s$ and maturity t_N, so its value is given by $V^{\text{def}} = E^Q(\int_0^{t_N} \delta_s \gamma_s \exp(-\int_0^s R_u \, \mathrm{d}u) \, \mathrm{d}s)$. We have therefore reduced the pricing of credit default swaps to a pricing problem in the default-free world. Methods for solving this problem will be discussed in the next section.

Recovery of market value. Recovery of market value, abbreviated RM, is an alternative recovery model for defaultable bonds and other credit-risky securities that has been put forward by Duffie and Singleton (1999); its main virtue is the fact that it leads to particularly simple pricing formulas. Consider a claim whose payoff consists of the survival claim X and a recovery payment at the default time. Under the RM hypothesis it is assumed that this recovery payment is given by $(1 - \delta_\tau) V_\tau I_{\{\tau \leqslant T\}}$, where the $(\mathcal{F}_t)$-adapted process $(\delta_t) \in (0, 1)$ gives the percentage loss given default of the claim and where the $(\mathcal{F}_t)$-adapted process (V_t) gives the pre-default value of the claim. Note that this is a recursive definition, as the pre-default value at time t also depends on the form of the recovery payments in the time period $(t, T]$. Nonetheless, the following result can be established.

Proposition 10.20. *Suppose that, under Q, τ is doubly stochastic with hazard rate process (γ_t). Suppose, moreover, that X is integrable and that the RM assumption holds. Then the pre-default value process (V_t) is uniquely determined and is given by*

$$V_t = E^Q\left(\exp\left(-\int_t^T (r_s + \delta_s \gamma_s) \, \mathrm{d}s\right) X \;\middle|\; \mathcal{F}_t\right), \quad 0 \leqslant t \leqslant T. \tag{10.57}$$

Note that for $\delta_t \equiv 1$ the claim is a standard survival claim; in that case, (10.57) reduces to the formula (10.53). On the other hand, for $\delta_t \equiv 0$ the claim is essentially default free; in that case, (10.57) reduces to the standard pricing formula for the claim X in a default-free security market model. For a proof of Proposition 10.20 we refer to the references given in Notes and Comments.

Credit spreads and hazard rates. With doubly stochastic default times the risk-neutral hazard process (γ_t) and the credit spread

$$c(t, T) = -\frac{1}{T - t}(\ln p_1(t, T) - \ln p_0(t, T))$$

of defaultable bonds are closely related. Analytic results are most easily derived for the instantaneous credit spread given by

$$c(t, t) = \lim_{T \to t} c(t, T) = -\frac{\partial}{\partial T}\bigg|_{T=t} (\ln p_1(t, T) - \ln p_0(t, T)). \tag{10.58}$$

Assume that $\tau > t$, so that $p_1(t,t) = p_0(t,t) = 1$. We therefore obtain

$$\left.\frac{\partial}{\partial T}\right|_{T=t} \ln p_1(t,T) = \left.\frac{\partial}{\partial T}\right|_{T=t} p_1(t,T), \tag{10.59}$$

and similarly for $p_0(t,T)$. To compute the derivative in (10.59) we need to distinguish between the different recovery models. Under the RM hypothesis we can apply Proposition 10.20 with $X = 1$. Exchanging expectation and differentiation we obtain

$$\begin{aligned} -\left.\frac{\partial}{\partial T}\right|_{T=t} p_1(t,T) &= -E^Q\left(\left.\frac{\partial}{\partial T}\right|_{T=t} \exp\left(-\int_t^T (r_s + \delta_s\gamma_s)\,\mathrm{d}s\right) \,\middle|\, \mathcal{F}_t\right) \\ &= r_t + \delta_t\gamma_t. \end{aligned} \tag{10.60}$$

Applying (10.60) with $\delta_t \equiv 0$ yields

$$-\left.\frac{\partial}{\partial T}\right|_{T=t} p_0(t,T) = r_t,$$

so that $c(t,t) = \delta_t\gamma_t$, i.e. the instantaneous credit spread equals the product of the hazard rate and the percentage loss given default, which is quite intuitive from an economic point of view. Under the RF recovery model, $p_1(t,T)$ is given by the sum of the price of a survival claim $I_{\{\tau>T\}}$ and a payment at default of size $(1-\delta_\tau)$. Equation (10.60) with $\delta_t \equiv 1$ shows that the derivative with respect to T of the survival claim at $T = t$ is equal to $-(r_t + \gamma_t)$. For the recovery payment we get

$$\left.\frac{\partial}{\partial T}\right|_{T=t} E\left(\int_t^T \gamma_s(1-\delta_s)\exp\left(-\int_t^s R_u\,\mathrm{d}u\right)\mathrm{d}s \,\middle|\, \mathcal{F}_t\right) = (1-\delta_t)\gamma_t.$$

Hence

$$-\left.\frac{\partial}{\partial T}\right|_{T=t} p_1(t,T) = r_t + \gamma_t - (1-\delta_t)\gamma_t = r_t + \delta_t\gamma_t,$$

so that $c_1(t,t)$ is again equal to $\delta_t\gamma_t$. An analogous computation shows that we also have $c_1(t,t) = \delta_t\gamma_t$ under RT. However, for $T - t > 0$, the credit spread corresponding to the different recovery models differs, as is illustrated in Section 10.6.3.

Notes and Comments

The material discussed in this section is based on many sources. We mention in particular the books by Lando (2004) and Bielecki and Rutkowski (2002). The text by Bielecki and Rutkowski is more technical than our presentation; among other things the authors discuss various probabilistic characterizations of doubly stochastic random times. The threshold-simulation approach for doubly stochastic random times requires the simulation of trajectories of the hazard process. An excellent source for simulation techniques for stochastic processes is Glasserman (2003).

More general reduced-form models where the default time τ is not doubly stochastic are discussed, for example, in Kusuoka (1999), Elliott, Jeanblanc and Yor (2000), Bélanger, Shreve and Wong (2004), Collin-Dufresne, Goldstein and Hugonnier (2004) and Blanchet-Scalliet and Jeanblanc (2004).

Theorem 10.19 is originally due to Lando (1998); related results were obtained by Jarrow and Turnbull (1995) and Jarrow, Lando and Turnbull (1997). Proposition 10.20 is due to Duffie and Singleton (1999); extensions are discussed in Becherer and Schweizer (2005). An excellent text for the overall mathematical background is Jeanblanc, Yor and Chesney (2009).

The analogy with default-free term structure models makes the reduced-form models with doubly stochastic default times relatively easy to apply. However, some care is required in interpreting the results and applying the *linear pricing rules* for corporate debt that the models imply. In particular, one must bear in mind that in these models the default intensity does not explicitly take into account the structure of a firm's outstanding risky debt. A formal analysis of the effect of debt structure on bond values is best carried out in the context of firm-value models, where the default is explicitly modelled in terms of fundamental economic quantities. A good discussion of these issues can be found in Chapter 2 of Lando (2004).

10.6 Affine Models

In order to apply the pricing formulas for doubly stochastic random times obtained in Theorem 10.19 we need effective ways to evaluate the conditional expectations on the right-hand side of equations (10.53), (10.54) and (10.55). In most models, where default is modelled by a doubly stochastic random time, (r_t) and (γ_t) are modelled as functions of some p-dimensional Markovian state variable process $(\boldsymbol{\Psi}_t)$ with state space given by the domain $D \subset \mathbb{R}^p$, so that the natural background filtration is given by $(\mathcal{F}_t) = \sigma(\{\boldsymbol{\Psi}_s : s \leqslant t\})$. Moreover, $R_t := r_t + \gamma_t$ is of the form $R_t = R(\boldsymbol{\Psi}_t)$ for some function $R\colon D \subseteq \mathbb{R}^p \to \mathbb{R}_+$. We therefore have to compute conditional expectations of the form

$$E\left(\exp\left(-\int_t^T R(\boldsymbol{\Psi}_s)\,\mathrm{d}s\right)g(\boldsymbol{\Psi}_T) + \int_t^T h(\boldsymbol{\Psi}_s)\exp\left(-\int_t^s R(\boldsymbol{\Psi}_u)\,\mathrm{d}u\right)\mathrm{d}s \;\middle|\; \mathcal{F}_t\right) \tag{10.61}$$

for generic functions $g, h\colon D \to \mathbb{R}_+$. Since $(\boldsymbol{\Psi}_t)$ is a Markov process, this conditional expectation is given by some function $f(t, \boldsymbol{\Psi}_t)$ of time and the current value $\boldsymbol{\Psi}_t$ of the state variable process. It is well known that under some additional regularity assumptions the function f can be computed as solution of a parabolic partial differential equation (PDE)—this is the celebrated *Feynman–Kac formula*. The Feynman–Kac formula provides a way to determine f using analytical or numerical techniques for PDEs. In particular, it is known that in the case where $(\boldsymbol{\Psi}_t)$ belongs to the class of *affine jump diffusions* (see below), R is an affine function, $g(\boldsymbol{\psi}) = \mathrm{e}^{\boldsymbol{u}'\boldsymbol{\psi}}$ for some $\boldsymbol{u} \in \mathbb{R}^p$ and $h \equiv 0$, the function f takes the form

$$f(t, \boldsymbol{\psi}) = \exp(\alpha(t, T) + \boldsymbol{\beta}(t, T)'\boldsymbol{\psi}) \tag{10.62}$$

for deterministic functions $\alpha\colon [0, T] \to \mathbb{R}$ and $\boldsymbol{\beta}\colon [0, T] \to R^p$; moreover, α and $\boldsymbol{\beta}$ are determined by a $(p+1)$-dimensional ordinary differential equation (ODE) system that is easily solved numerically. Models based on affine jump diffusions and an affine specification of R are therefore relatively easy to implement, which

explains their popularity in practice. A relationship of the form (10.62) is often termed an *affine term structure*, as it implies that continuously compounded yields of bonds at time t are affine functions of $\boldsymbol{\Psi}_t$.

In this section we discuss these results. We concentrate on the case where the state variable process is given by a one-dimensional diffusion; extensions to processes with jumps will be considered briefly at the end.

10.6.1 Basic Results

The PDE characterization of f. We assume that the state variable process (Ψ_t) is the unique solution of the SDE

$$\mathrm{d}\Psi_t = \mu(\Psi_t)\,\mathrm{d}t + \sigma(\Psi_t)\,\mathrm{d}W_t, \quad \Psi_0 = \psi \in D, \tag{10.63}$$

with state space given by the domain $D \subseteq \mathbb{R}$. Here, (W_t) is a standard, one-dimensional Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$, and μ and σ are continuous functions from D to $\mathbb{R}$ and D to $\mathbb{R}_+$, respectively. The next result shows that the conditional expectation (10.61) can be computed as the solution of a parabolic PDE.

Lemma 10.21 (Feynman–Kac). *Consider generic functions $R, g, h\colon D \to \mathbb{R}_+$. Suppose that the function $f\colon [0,T] \times D \to \mathbb{R}$ is continuous, once continuously differentiable in t and twice continuously differentiable in ψ on $[0,T) \times D$, and that f solves the terminal-value problem*

$$\left.\begin{aligned} f_t + \mu(\psi) f_\psi + \tfrac{1}{2}\sigma^2(\psi) f_{\psi\psi} + h(\psi) = R(\psi) f, \quad (t,\psi) \in [0,T) \times D, \\ f(T,\psi) = g(\psi), \quad \psi \in D. \end{aligned}\right\} \tag{10.64}$$

If f is bounded or, more generally, if $\max_{0\leqslant t\leqslant T} f(t,\psi) \leqslant C(1+\psi^2)$ for $\psi \in D$, then

$$E\left(\exp\left(-\int_t^T R(\boldsymbol{\Psi}_s)\,\mathrm{d}s\right) g(\boldsymbol{\Psi}_T) + \int_t^T h(\boldsymbol{\Psi}_s) \exp\left(-\int_t^s R(\boldsymbol{\Psi}_u)\,\mathrm{d}u\right)\mathrm{d}s \;\middle|\; \mathcal{F}_t\right) = f(t,\Psi_t). \tag{10.65}$$

The Feynman–Kac formula is a standard result of stochastic calculus and it is discussed in many textbooks on stochastic processes and financial mathematics, so we omit the proof (references are given in Notes and Comments).

Affine term structure. We begin with the case $h \equiv 0$; in financial terms this means that we concentrate on survival claims. The following assumption ensures that for $h \equiv 0$ the solution of the PDE (10.64), with terminal condition $g(\psi) = \mathrm{e}^{u\psi}$, $u\psi \leqslant 0$, for $\psi \in D$, is of the form (10.62), so that we have an affine term structure. Note that $g \equiv 1$ for $u = 0$; this is the appropriate terminal condition for pricing zero-coupon bonds.

Assumption 10.22. *R, μ and σ^2 are affine functions of ψ, i.e. there are constants ρ^0, ρ^1, k^0, k^1, h^0 and h^1 such that $R(\psi) = \rho^0 + \rho^1\psi$, $\mu(\psi) = k^0 + k^1\psi$ and $\sigma^2(\psi) = h^0 + h^1\psi$. Moreover, for all $\psi \in D$ we have $h^0 + h^1\psi \geqslant 0$ and $\rho_0 + \rho_1\psi \geqslant 0$.*

Fix some $T > 0$. We try to find a solution of (10.64) of the form $\tilde{f}(t,\psi) = \exp(\alpha(t,T) + \beta(t,T)\psi)$ for continuously differentiable functions $\alpha(\cdot,T)$ and $\beta(\cdot,T)$. As $\tilde{f}(T,\psi) = \mathrm{e}^{u\psi}$, we immediately obtain the terminal conditions $\alpha(T,T) = 0$, $\beta(T,T) = u$. Denote by $\alpha'(\cdot,T)$ and $\beta'(\cdot,T)$ the derivatives of α and β with respect to t. Using the special form of $\tilde{f}$ we obtain that

$$\tilde{f}_t = (\alpha' + \beta'\psi)\tilde{f}, \quad \tilde{f}_\psi = \beta\tilde{f} \quad \text{and} \quad \tilde{f}_{\psi\psi} = \beta^2\tilde{f}.$$

Hence, under Assumption 10.22 the PDE (10.64) takes the form

$$(\alpha' + \beta'\psi)\tilde{f} + (k^0 + k^1\psi)\beta\tilde{f} + \tfrac{1}{2}(h^0 + h^1\psi)\beta^2\tilde{f} = (\rho^0 + \rho^1\psi)\tilde{f}.$$

Dividing by $\tilde{f}$ and rearranging we obtain

$$\alpha' + k^0\beta + \tfrac{1}{2}h^0\beta^2 - \rho^0 + (\beta' + k^1\beta + \tfrac{1}{2}h^1\beta^2 - \rho^1)\psi = 0.$$

Since this equation must hold for all $\psi \in D$, we obtain the following ODE system:

$$\beta'(t,T) = \rho^1 - k^1\beta(t,T) - \tfrac{1}{2}h^1\beta^2(t,T), \quad \beta(T,T) = u, \tag{10.66}$$

$$\alpha'(t,T) = \rho^0 - k^0\beta(t,T) - \tfrac{1}{2}h^0\beta^2(t,T), \quad \alpha(T,T) = 0. \tag{10.67}$$

The ODE (10.66) for $\beta(\cdot,T)$ is a so-called *Ricatti equation*. While explicit solutions exist only in certain special cases, the ODE is easily solved numerically. The ODE (10.67) for $\alpha(\cdot,T)$ can be solved by simple (numerical) integration once β has been determined. Summing up, we have the following proposition.

Proposition 10.23. *Suppose that Assumption 10.22 holds, that the ODE system (10.66), (10.67) has a unique solution (α,β) on $[0,T]$, and that there is some C such that $\beta(t,T)\psi \leqslant C$ for all $t \in [0,T]$, $\psi \in D$. Then*

$$E\left(\exp\left(-\int_t^T R(\Psi_s)\,\mathrm{d}s\right)\mathrm{e}^{u\Psi_T} \;\middle|\; \mathcal{F}_t\right) = \exp(\alpha(t,T) + \beta(t,T)\Psi_t).$$

Proof. The result follows immediately from Lemma 10.21, as our assumption on β implies that $\tilde{f}(t,\psi) = \exp(\alpha(t,T) + \beta(t,T)\psi)$ is bounded. □

10.6.2 The CIR Square-Root Diffusion

A very popular affine model is the square-root diffusion model proposed by Cox, Ingersoll and Ross (1985) as a model for the short rate of interest. In this model (Ψ_t) is given by the solution of the SDE

$$\mathrm{d}\Psi_t = \kappa(\bar{\theta} - \Psi_t)\,\mathrm{d}t + \sigma\sqrt{\Psi_t}\,\mathrm{d}W_t, \quad \Psi_0 = \psi > 0, \tag{10.68}$$

for parameters $\kappa, \bar{\theta}, \sigma > 0$ and state space $D = [0,\infty)$. Clearly, (10.68) is an affine model in the sense of Assumption 10.22; the parameters are given by $k^0 = \kappa\bar{\theta}$, $k^1 = -\kappa$, $h^0 = 0$ and $h^1 = \sigma^2$.

It is well known that the SDE (10.68) admits a global solution (see Notes and Comments for a reference). This issue is non-trivial since the square-root function is not Lipschitz and since one has to ensure that the solution remains in D for all $t > 0$. Note that (10.68) implies that (Ψ_t) is a *mean-reverting process*: if Ψ_t deviates from the mean-reversion level $\bar{\theta}$, the process is pulled back towards $\bar{\theta}$. Moreover, if the mean reversion is sufficiently strong relative to the volatility, trajectories never reach zero. More precisely, let $\tau_0(\Psi) := \inf\{t \geqslant 0 \colon \Psi_t = 0\}$. It is well known that for $\kappa\bar{\theta} \geqslant \frac{1}{2}\sigma^2$ one has $P(\tau_0(\Psi) < \infty) = 0$, whereas for $\kappa\bar{\theta} < \frac{1}{2}\sigma^2$ one has $P(\tau_0(\Psi) < \infty) = 1$.

In the CIR square-root model the Ricatti equations (10.66) and (10.67) can be solved explicitly. Using Proposition 10.23 it can be computed that

$$E\left(\exp\left(-\int_t^T (\rho^0 + \rho^1 \Psi_s)\,\mathrm{d}s\right) \,\middle|\, \mathcal{F}_t\right) = \exp(\alpha(T-t) + \beta(T-t)\Psi_t),$$

with

$$\beta(\tau) = \frac{-2\rho^1(\mathrm{e}^{\gamma\tau} - 1)}{\gamma - \kappa + \mathrm{e}^{\gamma\tau}(\gamma + \kappa)}, \tag{10.69}$$

$$\alpha(\tau) = -\rho^0\tau + 2\frac{\kappa\bar{\theta}}{\sigma^2}\ln\left(\frac{2\gamma\mathrm{e}^{\tau(\gamma+\kappa)/2}}{\gamma - \kappa + \mathrm{e}^{\gamma\tau}(\gamma + \kappa)}\right), \tag{10.70}$$

and $\tau := T - t$, $\gamma := \sqrt{\kappa^2 + 2\sigma^2\rho^1}$. These formulas are the key to pricing bonds in models where the risk-free short rate and the default intensities are affine functions of independent square-root processes, as is shown in the next example.

Example 10.24 (a three-factor model). We now consider the pricing of zero-coupon bonds in a three-factor model similar to models that are frequently used in the literature. We assume that $\boldsymbol{\Psi}_t = (\Psi_{t,1}, \Psi_{t,2}, \Psi_{t,3})'$ is a vector of three independent square-root diffusions with dynamics $\mathrm{d}\Psi_{t,i} = \kappa_i(\bar{\theta}_i - \Psi_{t,i})\,\mathrm{d}t + \sigma_i\sqrt{\Psi_{t,i}}\,\mathrm{d}W_{t,i}$ for independent Brownian motions $(W_{t,i})$, $i = 1, 2, 3$. The risk-free short rate of interest is given by $r_t = r_0 + \Psi_{t,2} - \Psi_{t,1}$ for a constant $r_0 \geqslant 0$; the hazard rate of the counterparty under consideration is given by $\gamma_t = \gamma_1\Psi_{t,1} + \Psi_{t,3}$ for some constant $\gamma_1 > 0$. This parametrization allows for negative instantaneous correlation between (r_t) and (γ_t), which is in line with empirical evidence. Note, however, that this negative correlation comes at the expense of possibly negative risk-free interest rates. In this context the price of a default-free zero-coupon bond is given by

$$\begin{aligned} &p_0(t,T) \\ &\quad = E\left(\exp\left(-\int_t^T r_s\,\mathrm{d}s\right) \,\middle|\, \mathcal{F}_t\right) \\ &\quad = \mathrm{e}^{-r_0(T-t)} E\left(\exp\left(-\int_t^T \Psi_{s,2}\,\mathrm{d}s\right) \,\middle|\, \mathcal{F}_t\right) E\left(\exp\left(\int_t^T \Psi_{s,1}\,\mathrm{d}s\right) \,\middle|\, \mathcal{F}_t\right), \end{aligned} \tag{10.71}$$

where we have used the independence of $(\Psi_{t,1})$, $(\Psi_{t,2})$, $(\Psi_{t,3})$. Each of the terms in (10.71) can be evaluated using the above formulas for α and β (equations (10.69)

and (10.70)). Assuming that we have recovery of treasury in default (see Section 10.4.3) and a deterministic percentage loss given default δ, we find that the price of a defaultable zero-coupon bond is given by

$$p_1(t,T) = (1-\delta)\,p_0(t,T) + \delta E\left(\exp\left(-\int_t^T (r_s+\gamma_s)\,\mathrm{d}s\right)\,\middle|\,\mathcal{F}_t\right).$$

By definition of r_t and γ_t the last term on the right-hand side equals

$$\delta E\left(\exp\left(-\int_t^T (r_0 + (\gamma_1 - 1)\Psi_{s,1} + \Psi_{s,2} + \Psi_{s,3})\,\mathrm{d}s\right)\,\middle|\,\mathcal{F}_t\right),$$

which can be evaluated in a similar way to the evaluation of expression (10.71). In the next section we will show how to deal with more complicated recovery models, such as recovery of face value.

10.6.3 Extensions

A jump-diffusion model for (Ψ_t). We briefly discuss an extension of the basic model (10.63), where the economic factor process (Ψ_t) follows a diffusion with jumps. Adding jumps to the dynamics of (Ψ_t) provides more flexibility for modelling default correlations in models with conditionally independent defaults (see Section 17.3.2), and it also leads to more realistic credit spread dynamics.

In this section we assume that (Ψ_t) is the unique solution of the SDE

$$\mathrm{d}\Psi_t = \mu(\Psi_t)\,\mathrm{d}t + \sigma(\Psi_t)\,\mathrm{d}W_t + \mathrm{d}Z_t, \quad \Psi_0 = \psi \in D. \tag{10.72}$$

Here, (Z_t) is a pure jump process whose jump intensity at time t is equal to $\lambda^Z(\Psi_t)$ for some function $\lambda^Z\colon D \to \mathbb{R}_+$ and whose jump-size distribution has df ν on $\mathbb{R}$. Intuitively this means that, given the trajectory $(\Psi_t(\omega))_{t\geqslant 0}$ of the factor process, (Z_t) jumps at the jump times of an inhomogeneous Poisson process (see Section 13.2.7) with time-varying intensity $\lambda^Z(t,\Psi_t)$; the size of the jumps has df ν.

Suppose now that Assumption 10.22 holds, and that $\lambda^Z(\psi) = l^0 + l^1\psi$ for constants l^0, l^1 such that $\lambda^Z(\psi) > 0$ for all $\psi \in D$. In that case we say that (Ψ_t) follows an *affine jump diffusion*. For $x \in \mathbb{R}$ denote by $\hat{\nu}(x) = \int_{\mathbb{R}} \mathrm{e}^{-xy}\,\mathrm{d}\nu(y) \in (0,\infty]$ the extended Laplace–Stieltjes transform of ν (with domain $\mathbb{R}$ instead of the usual domain $[0,\infty)$). Consider the following extension of the ODE system (10.66), (10.67):

$$\beta'(t,T) = \rho^1 - k^1\beta(t,T) - \tfrac{1}{2}h^1\beta^2(t,T) - l^1(\hat{\nu}(-\beta(t,T)) - 1), \tag{10.73}$$

$$\alpha'(t,T) = \rho^0 - k^0\beta(t,T) - \tfrac{1}{2}h^0\beta^2(t,T) - l^0(\hat{\nu}(-\beta(t,T)) - 1), \tag{10.74}$$

with terminal condition $\beta(T,T) = u$ for some $u \leqslant 0$ and $\alpha(T,T) = 0$. Suppose that the system described by (10.74) and (10.73) has a unique solution α, β and that $\beta(t,T)\psi \leqslant C$ for all $t \in [0,T]$, $\psi \in D$ (for $l^0 \neq 0$ or $l^1 \neq 0$ this implicitly implies that $\hat{\nu}(-\beta(t,T)) < \infty$ for all t). Define $\tilde{f}(t,\psi) = \exp(\alpha(t,T) + \beta(t,T)\psi)$. Using similar arguments to those above it can then be shown that the conditional expectation $E(\exp(-\int_t^T R(\Psi_s)\,\mathrm{d}s)\mathrm{e}^{u\Psi_T} \mid \mathcal{F}_t)$ equals $\tilde{f}(t,\Psi_t)$.

Table 10.4. Parameters used in the model of Duffie and Gârleanu (2001). Recall that l^0 gives the intensity of jump in the factor process, μ gives the average jump size. With these parameters the average waiting time for a jump in the systematic factor process is $1/l^0 = 5$ years.

κ	$\bar{\theta}$	σ	l^0	μ
0.6	0.02	0.14	0.2	0.1

Example 10.25 (the model of Duffie and Gârleanu (2001)). The following jump-diffusion model has been used in the literature on CDO pricing. The dynamics of (Ψ_t) are given by

$$\mathrm{d}\Psi_t = \kappa(\bar{\theta} - \Psi_t)\,\mathrm{d}t + \sigma\sqrt{\Psi_t}\,\mathrm{d}W_t + \mathrm{d}Z_t \tag{10.75}$$

for parameters $\kappa, \bar{\theta}, \sigma > 0$ and a jump process (Z_t) with constant jump intensity $l^0 > 0$ and exponentially distributed jump sizes with parameter $1/\mu$. Following Duffie and Gârleanu, we will sometimes call the model (10.75) a *basic affine jump diffusion*. Note that these assumptions imply that the mean of ν is equal to μ and that ν has support $[0, \infty)$, so that (Ψ_t) has only upwards jumps. The existence of a solution to (10.75) therefore follows from the existence of solutions in the pure diffusion case. It is relatively easy to show that for $t \to \infty$ we obtain $E(\Psi_t) \to \bar{\theta} + l^0\mu/\kappa$. For illustrative purposes we present the parameter values used in Duffie and Gârleanu (2001) in Table 10.4; a typical trajectory of (Ψ_t) is simulated in Figure 10.8. Next we compute the Laplace–Stieltjes transform $\hat{\nu}$. We obtain for $u > -1/\mu$ that

$$\hat{\nu}(u) = \int_0^\infty \mathrm{e}^{-ux}(1/\mu)\mathrm{e}^{-x/\mu}\,\mathrm{d}x = \frac{1}{1+\mu u};$$

for $u \leqslant -1/\mu$ we get $\hat{\nu}(u) = \infty$. We therefore have all the necessary ingredients to set up the Ricatti equations (10.74) and (10.73). In the case of the model (10.75) it is in fact possible to solve these equations explicitly (see, for example, Chapter 11 of Duffie and Singleton (2003)). However, the explicit solution is given by a very lengthy expression so we omit the details.

Application to payment-at-default claims. According to Theorem 10.19, in a model with a doubly stochastic default time τ with risk-neutral hazard rate $\gamma(\Psi_t)$, the price at t of a payment-at-default claim of constant size $\delta > 0$ equals

$$\delta E\left(\int_t^T \gamma(\Psi_s)\exp\left(-\int_t^s R(\Psi_u)\,\mathrm{d}u\right)\mathrm{d}s \;\middle|\; \mathcal{F}_t\right), \tag{10.76}$$

where again $R(\psi) = r(\psi) + \gamma(\psi)$. Using the Fubini Theorem this equals

$$\delta\int_t^T E\left(\gamma(\Psi_s)\exp\left(-\int_t^s R(\Psi_u)\,\mathrm{d}u\right) \;\middle|\; \mathcal{F}_t\right)\mathrm{d}s. \tag{10.77}$$

Suppose now that $\gamma(\psi) = \gamma^0 + \gamma^1\psi$, that $R(\psi) = \rho^0 + \rho^1\psi$ and that (Ψ_t) is given by an affine jump diffusion as introduced above. In that case the inner expectation in (10.77) is given by a function $F(t, s, \Psi_t)$. This function can be computed using an extension of the basic affine methodology, so that (10.77) can be

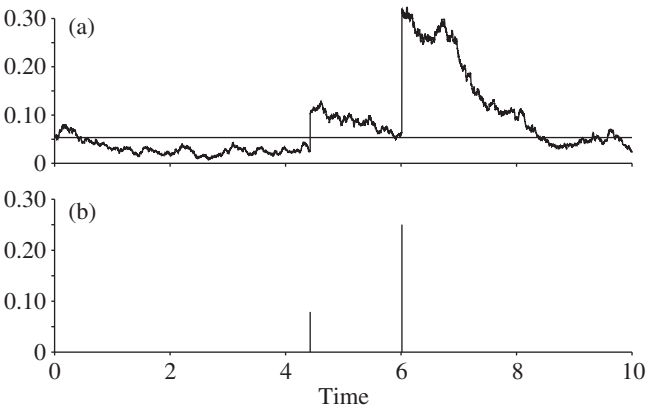

Figure 10.8. (a) A typical trajectory of the basic affine jump diffusion model (10.75) and (b) the corresponding jumps of (Z_t). The parameter values used are given in Table 10.4; the initial value Ψ_0 is equal to the long-run mean $\bar{\theta} + (l^0\mu)/\kappa$, which is marked by the horizontal line.

computed by one-dimensional numerical integration. Define for $0 \leqslant t \leqslant s$ the function $\tilde{f}(t, s, \psi) = \exp(\alpha(t, s) + \beta(t, s)\psi)$, where $\alpha(\cdot, s)$ and $\beta(\cdot, s)$ solve the ODEs (10.74), (10.73) with terminal condition $\alpha(s, s) = \beta(s, s) = 0$. Denote by $\hat{\nu}'(x)$ the derivative of the Laplace–Stieltjes transform of ν. It is then a straightforward application of standard calculus to show that, modulo some integrability conditions, $F(t, s, \psi) = \tilde{f}(t, s, \psi)(A(t, s) + B(t, s)\psi)$, where $A(\cdot, s)$ and $B(\cdot, s)$ solve the following ODE system:

$$B'(t, s) + k^1 B(t, s) + h^1 \beta B(t, s) - l^1 \hat{\nu}'(-\beta) B(t, s) = 0, \tag{10.78}$$

$$A'(t, s) + k^0 B(t, s) + h^0 \beta B(t, s) - l^0 \hat{\nu}'(-\beta) B(t, s) = 0, \tag{10.79}$$

with terminal conditions $A(s, s) = \gamma_0$, $B(s, s) = \gamma_1$. Again, (10.78) and (10.79) are straightforward to evaluate numerically.

It is of course possible to compute the conditional expectation (10.76) by using the Feynman–Kac formula (10.65) with $g \equiv 0$ and $h = \gamma$. However, in most cases the ensuing PDE needs to be solved numerically.

Example 10.26 (defaultable zero-coupon bonds and CDSs). We now have all the necessary ingredients to compute prices and credit spreads of defaultable zero-coupon bonds and CDS spreads in a model with a doubly stochastic default time with hazard rate $\gamma_t = \Psi_t$ for a one-dimensional affine jump diffusion (Ψ_t). In Figure 10.9 we plot the credit spread for defaultable bonds for the recovery assumptions discussed in Section 10.4.3. Note that, for $T \to t$, i.e. for time to maturity close to zero, the spread tends to $c(t, t) = \delta\Psi_t > 0$, as claimed in Section 10.5.3; in particular, the credit spread does not vanish as $T \to t$. This is in stark contrast to firm-value models, where typically $c(t, t) = 0$, as was shown in Section 10.3.1. Note further that, for $T - t$ large, under the RF assumption we obtain *negative* credit spreads, which is clearly unrealistic. These negative credit spreads are caused by the fact that under RF we obtain a payment of fixed size $1 - \delta$ immediately at default. If the

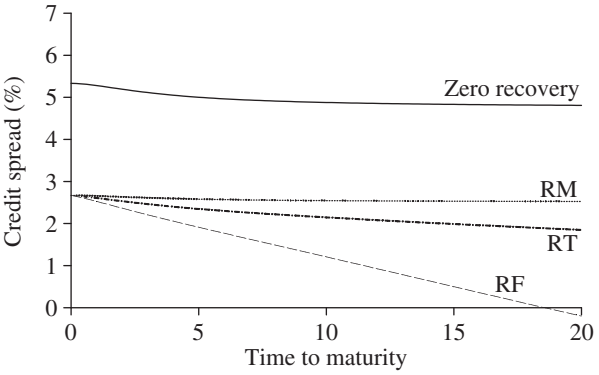

Figure 10.9. Spreads of defaultable zero-coupon bonds in the Duffie–Gârleanu model (10.75) for various recovery assumptions. The parameters of (Ψ_t) are given in Table 10.4; the initial value is $\Psi_0 \approx 0.0533$. The risk-free interest rate and the loss given default are deterministic and are given by $r = 6\%$ and $\delta = 0.5$. Note that under the RF recovery model, the spread becomes negative for large times to maturity.

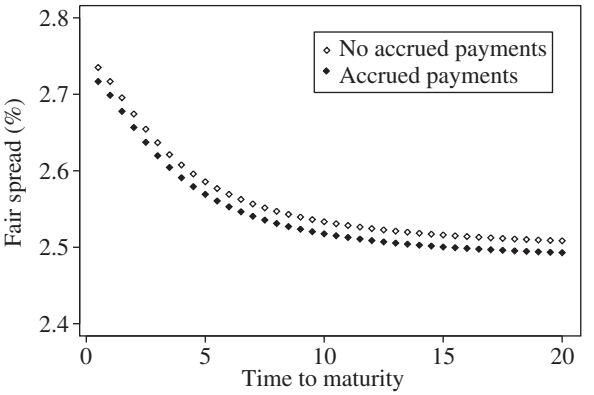

Figure 10.10. Fair CDS spreads in the Duffie–Gârleanu model (10.75) for a CDS contract with semiannual premium payments and varying time to maturity. The parameters of (Ψ_t) are given in Table 10.4; the initial value is $\Psi_0 \approx 0.0533$. The risk-free interest rate and the loss given default are deterministic and are given by $r = 6\%$ and $\delta = 0.5$. Note that, for small time to maturity, the fair swap spread is approximately equal to $\delta\Psi_0 \approx 2.7\%$.

default-free interest rate r is relatively large, it may happen that

$$E^Q\left(\exp\left(-\int_0^\tau r_s\,\mathrm{d}s\right)(1-\delta)I_{\{\tau\leqslant T\}}\right) > E^Q\left(\exp\left(-\int_0^T r_s\,\mathrm{d}s\right)I_{\{\tau\leqslant T\}}\right),$$

even if $\delta > 0$. This stems from the fact that on the right-hand side discounting is done over the whole period $[0, T]$ (as opposed to $[0, \tau]$), so that discounting has a large impact on the value of the right-hand side, compensating the higher terminal pay-off. In Figure 10.10 we have plotted the fair spreads for CDSs with and without accrued payments for varying maturities, assuming that the risk-neutral hazard process is a basic affine jump diffusion.

Notes and Comments

The Feynman–Kac formula is discussed, for example, in Section 5.5 of Björk (2004) and, at a slightly more technical level, in Section 5.7 of Karatzas and Shreve (1988).

Important original papers on affine models in term structure modelling are Duffie and Kan (1996) for diffusion models and Duffie, Pan and Singleton (2000) for jump diffusions. The latter paper also contains other applications of affine models, such as the pricing of equity options under stochastic volatility and econometric issues related to affine models. It should be mentioned that there is also a converse to Proposition 10.23; if the conditional expectations $E(\exp(-\int_t^T R(\Psi_s)\,\mathrm{d}s)\mathrm{e}^{u\Psi_T} \mid \mathcal{F}_t)$ are all exponentially affine functions of Ψ_t, the process (Ψ_t) is necessarily affine (see Duffie and Kan (1996), Duffie, Filipović and Schachermayer (2003) or Chapter 10 of Filipović (2009) for details).

The mathematical properties of the CIR model are discussed in, for example, Chapter 6.2 of Lamberton and Lapeyre (1996), where the explicit solution of the Ricatti equations in the CIR model (summarized by (10.69) and (10.70)) is also derived. The model studied in Example 10.24 is akin to models proposed by Duffie and Singleton (1999). Problems related to the modelling of negative correlation between state variable processes in an affine setting are discussed in Section 5.8 of Lando (2004). Empirical work on affine models for defaultable bonds includes the publications of Duffee (1999) and Driessen (2005).

11

Portfolio Credit Risk Management

This chapter is concerned with one-period models for credit portfolios with a view towards credit risk management issues for portfolios of largely non-traded credit products, such as the retail and commercial loans in the banking book of a typical bank.

The main theme in our analysis is the modelling of the dependence structure of the default events. In fact, default dependence has a profound impact on the upper tail of the credit loss distribution for a large portfolio. This is illustrated in Figure 11.1, where we compare the loss distribution for a portfolio of 1000 firms that default independently (portfolio 1) with a more realistic portfolio of the same size where defaults are dependent (portfolio 2). In portfolio 2 defaults are weakly dependent, in the sense that the correlation between default events is approximately 0.5%. In both cases the default probability is approximately 1%, so, on average, we expect ten defaults. As will be seen in Section 11.5, the model applied to portfolio 2 can be viewed as a realistic model for the loss distribution of a homogeneous portfolio of 1000 loans with a Standard & Poor's rating of BB. We see clearly from Figure 11.1 that the loss distribution of portfolio 2 is skewed and its right tail is substantially heavier than the right tail of the loss distribution of portfolio 1, illustrating the dramatic impact of default dependence.

Note that there are good economic reasons for expecting dependence between defaults of different obligors. Most importantly, the financial health of a firm varies with randomly fluctuating macroeconomic factors such as changes in economic growth. Since different firms are affected by common macroeconomic factors, this creates dependence between their defaults.

We begin our analysis with a discussion of threshold models in Section 11.1. These can be viewed as multivariate extensions of the Merton model considered in Chapter 10. In Section 11.2 we consider so-called mixture models in which defaults are assumed to be conditionally independent events given a set of common factors. The factors are usually interpreted as macroeconomic variables and are also modelled stochastically. Mixture models are commonly used in practice, essentially for tractability reasons; many threshold models also have convenient representations as mixture models.

Sections 11.3 and 11.4 are concerned with the calculation or approximation of the portfolio loss distribution and related measures of tail risk. We give asymptotic approximations for tail probabilities and quantiles that hold in large, relatively

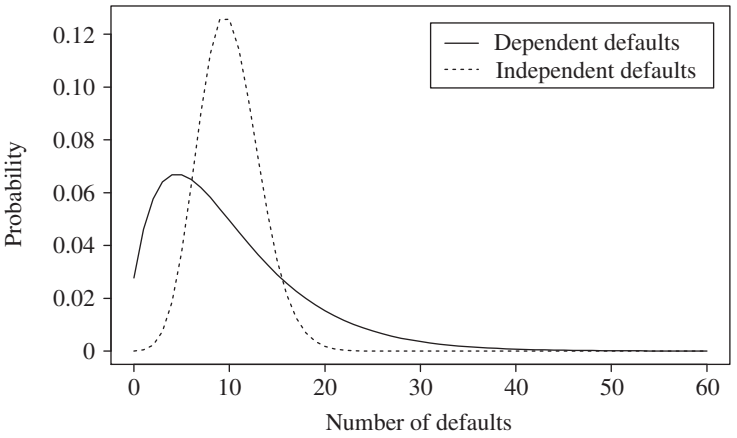

Figure 11.1. Comparison of loss distributions for two homogeneous portfolios of 1000 loans with a default probability of 1% and different dependence structures. In portfolio 1 defaults are assumed to be independent; in portfolio 2 we assume a default correlation of 0.5%. Portfolio 2 can be considered as representative for BB-rated loans. We clearly see that the default dependence generates a loss distribution with a heavier right tail and a shift of the mode to the left.

homogeneous portfolios (Section 11.3) and we discuss Monte Carlo methods for estimating tail probabilities, risk measures and capital allocations in mixture models (Section 11.4). Finally, in Section 11.5 we consider the important issue of statistical inference for credit models.

11.1 Threshold Models

The models of this section are one-period models for portfolio credit risk inspired by the firm-value models of Section 10.3. Their defining attribute is the idea that default occurs for a company i over the period $[0, T]$ if some critical rv X_i lies below some deterministic threshold d_i. The variable X_i can take on different interpretations. In a multivariate version of Merton's model, $X_i = X_{T,i}$ is a lognormally distributed asset value at the time horizon T and d_i represents liabilities to be repaid at T. More abstractly, X_i is frequently viewed as a latent variable representing the credit quality or "ability-to-pay" of obligor i.

The dependence among defaults stems from the dependence among the components of the vector $\boldsymbol{X} = (X_1, \ldots, X_m)'$. The distributions assumed for $\boldsymbol{X}$ can, in principle, be completely general, and indeed a major issue of this section will be the influence of the copula of $\boldsymbol{X}$ on the risk of the portfolio.

11.1.1 Notation for One-Period Portfolio Models

It is convenient to introduce some notation for one-period portfolio models that will be in force throughout the remainder of the chapter. We consider a portfolio of m obligors and fix a time horizon T. For $1 \leqslant i \leqslant m$, we let the rv R_i be a state indicator for obligor i at time T and assume that R_i takes integer values in the set $\{0, 1, \ldots, n\}$ representing, for example, rating classes; as in Section 10.2.1

we interpret the value 0 as default and non-zero values as states of increasing credit quality. At time $t = 0$ obligors are assumed to be in some non-default state.

Mostly we will concentrate on the binary outcomes of default and non-default and ignore the finer categorization of non-defaulted companies. In this case we write Y_i for the default indicator variables so that $Y_i = 1 \iff R_i = 0$ and $Y_i = 0 \iff R_i > 0$. The random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_m)'$ is a vector of default indicators for the portfolio, and $p(\boldsymbol{y}) = P(Y_1 = y_1, \ldots, Y_m = y_m)$, $\boldsymbol{y} \in \{0, 1\}^m$, is its joint probability function; the marginal default probabilities are denoted by $p_i = P(Y_i = 1)$, $i = 1, \ldots, m$.

The *default* or *event correlations* will be of particular interest to us; they are defined to be the correlations of the default indicators. Because

$$\operatorname{var}(Y_i) = E(Y_i^2) - p_i^2 = E(Y_i) - p_i^2 = p_i - p_i^2,$$

we obtain, for firms i and j with $i \neq j$, the formula

$$\rho(Y_i, Y_j) = \frac{E(Y_i Y_j) - p_i p_j}{\sqrt{(p_i - p_i^2)(p_j - p_j^2)}}. \tag{11.1}$$

We count the number of defaulted obligors at time T with the rv $M := \sum_{i=1}^m Y_i$. The actual loss if company i defaults—termed the *loss given default* (LGD) in practice—is modelled by the random quantity $\delta_i e_i$, where e_i represents the overall exposure to company i and $0 \leqslant \delta_i \leqslant 1$ represents a random proportion of the exposure that is lost in the event of default. We will denote the overall loss by $L := \sum_{i=1}^m \delta_i e_i Y_i$ and make further assumptions about the e_i and δ_i variables as and when we need them.

It is possible to set up different credit risk models leading to the same multivariate distribution for $\boldsymbol{R}$ or $\boldsymbol{Y}$. Since this distribution is the main object of interest in the analysis of portfolio credit risk, we call two models with state vectors $\boldsymbol{R}$ and $\tilde{\boldsymbol{R}}$ (or $\boldsymbol{Y}$ and $\tilde{\boldsymbol{Y}}$) *equivalent* if $\boldsymbol{R} \overset{\mathrm{d}}{=} \tilde{\boldsymbol{R}}$ (or $\boldsymbol{Y} \overset{\mathrm{d}}{=} \tilde{\boldsymbol{Y}}$).

The exchangeable special case. To simplify the analysis we will often assume that the state indicator vector $\boldsymbol{R}$, and thus the default indicator vector $\boldsymbol{Y}$, are *exchangeable* random vectors. This is one way to mathematically formalize the notion of *homogeneous* groups that is used in practice. Recall that a random vector $\boldsymbol{R}$ is said to be exchangeable if $(R_1, \ldots, R_m) \overset{\mathrm{d}}{=} (R_{\Pi(1)}, \ldots, R_{\Pi(m)})$ for any permutation $(\Pi(1), \ldots, \Pi(m))$ of $(1, \ldots, m)$. Exchangeability implies in particular that, for any $k \in \{1, \ldots, m-1\}$, all of the $\binom{m}{k}$ possible k-dimensional marginal distributions of $\boldsymbol{R}$ are identical. In this situation we introduce a simple notation for default probabilities where $\pi := P(Y_i = 1)$, $i \in \{1, \ldots, m\}$, is the default probability of any firm and

$$\pi_k := P(Y_{i_1} = 1, \ldots, Y_{i_k} = 1), \quad \{i_1, \ldots, i_k\} \subset \{1, \ldots, m\},\ 2 \leqslant k \leqslant m, \tag{11.2}$$

is the joint default probability for any k firms. In other words, π_k is the probability that an arbitrarily selected subgroup of k companies defaults in $[0, T]$. When default

indicators are exchangeable, we get

$$E(Y_i) = E(Y_i^2) = P(Y_i = 1) = \pi, \quad \forall i,$$
$$E(Y_i Y_j) = P(Y_i = 1, Y_j = 1) = \pi_2, \quad \forall i \neq j,$$

so that $\operatorname{cov}(Y_i, Y_j) = \pi_2 - \pi^2$; this implies that the default correlation in (11.1) is given by

$$\rho_Y := \rho(Y_i, Y_j) = \frac{\pi_2 - \pi^2}{\pi - \pi^2}, \quad i \neq j, \tag{11.3}$$

which is a simple function of the first- and second-order default probabilities.

11.1.2 Threshold Models and Copulas

We start with a general definition of a threshold model before discussing the link to copulas.

Definition 11.1. Let $\boldsymbol{X} = (X_1, \ldots, X_m)'$ be an m-dimensional random vector and let $D \in \mathbb{R}^{m \times n}$ be a deterministic matrix with elements d_{ij} such that, for every i, the elements of the ith row form a set of increasing thresholds satisfying $d_{i1} < \cdots < d_{in}$. Augment these thresholds by setting $d_{i0} = -\infty$ and $d_{i(n+1)} = \infty$ for all obligors and then set

$$R_i = j \iff d_{ij} < X_i \leqslant d_{i(j+1)}, \quad j \in \{0, \ldots, n\},\ i \in \{1, \ldots, m\}.$$

Then $(\boldsymbol{X}, D)$ is said to define a threshold model for the state vector $\boldsymbol{R} = (R_1, \ldots, R_m)'$.

We refer to $\boldsymbol{X}$ as the vector of *critical variables* and denote its marginal dfs by $F_{X_i}(x) = P(X_i \leqslant x)$. The ith row of D contains the *critical thresholds* for firm i. By definition, default (corresponding to the event $R_i = 0$) occurs if $X_i \leqslant d_{i1}$, so the default probability of company i is given by $p_i = F_{X_i}(d_{i1})$. When working with a default-only model we simply write $d_i = d_{i1}$ and denote the threshold model by $(\boldsymbol{X}, \boldsymbol{d})$.

In the context of such models it is important to distinguish the default correlation $\rho(Y_i, Y_j)$ of two firms $i \neq j$ from the correlation of the critical variables X_i and X_j. Since the critical variables are often interpreted in terms of asset values, the latter correlation is often referred to as *asset correlation*. For given default probabilities, $\rho(Y_i, Y_j)$ is determined by $E(Y_i Y_j)$ according to (11.1). Moreover, in a threshold model, $E(Y_i Y_j) = P(X_i \leqslant d_{i1},\ X_j \leqslant d_{j1})$, which implies that default correlation depends on the joint distribution of X_i and X_j. If $\boldsymbol{X}$ is multivariate normal, as in many models used in practice, the correlation of X_i and X_j determines the copula of their joint distribution and hence the default correlation (see Lemma 11.2 below). For general critical variables outside the multivariate normal class, the correlation of the critical variables does not fully determine the default correlation; this can have serious implications for the tail of the distribution of $M = \sum_{i=1}^m Y_i$, as will be shown in Section 11.1.5.

We now give a simple criterion for the equivalence of two threshold models in terms of the marginal distributions of the state vector $\boldsymbol{R}$ and the copula of $\boldsymbol{X}$. This

result clarifies the central role of copulas in threshold models. For the necessary background information on copulas we refer to Chapter 7.

Lemma 11.2. *Let $(\boldsymbol{X}, D)$ and $(\tilde{\boldsymbol{X}}, \tilde{D})$ be a pair of threshold models with state vectors $\boldsymbol{R} = (R_1, \ldots, R_m)'$ and $\tilde{\boldsymbol{R}} = (\tilde{R}_1, \ldots, \tilde{R}_m)'$, respectively. The models are equivalent if the following conditions hold.*

(i) *The marginal distributions of the random vectors $\boldsymbol{R}$ and $\tilde{\boldsymbol{R}}$ coincide, i.e.*

$$P(R_i = j) = P(\tilde{R}_i = j), \quad j \in \{1, \ldots, n\},\ i \in \{1, \ldots, m\}.$$

(ii) *$\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ admit the same copula C.*

Proof. According to Definition 11.1, $\boldsymbol{R} \overset{\mathrm{d}}{=} \tilde{\boldsymbol{R}}$ if and only if, for all $j_1, \ldots, j_m \in \{1, \ldots, n\}$,

$$\begin{aligned} &P(d_{1j_1} < X_1 \leqslant d_{1(j_1+1)}, \ldots, d_{mj_m} < X_m \leqslant d_{m(j_m+1)}) \\ &\qquad = P(\tilde{d}_{1j_1} < \tilde{X}_1 \leqslant \tilde{d}_{1(j_1+1)}, \ldots, \tilde{d}_{mj_m} < \tilde{X}_m \leqslant \tilde{d}_{m(j_m+1)}). \end{aligned}$$

By standard measure-theoretic arguments this holds if, for all $j_1, \ldots, j_m \in \{1, \ldots, n\}$,

$$P(X_1 \leqslant d_{1j_1}, \ldots, X_m \leqslant d_{mj_m}) = P(\tilde{X}_1 \leqslant \tilde{d}_{1j_1}, \ldots, \tilde{X}_m \leqslant \tilde{d}_{mj_m}).$$

By Sklar's Theorem (Theorem 7.3) this is equivalent to

$$C(F_{X_1}(d_{1j_1}), \ldots, F_{X_m}(d_{mj_m})) = C(F_{\tilde{X}_1}(\tilde{d}_{1j_1}), \ldots, F_{\tilde{X}_m}(\tilde{d}_{mj_m})),$$

where C is the copula of $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ (using condition (ii)). Condition (i) implies that $F_{X_i}(d_{ij}) = F_{\tilde{X}_i}(\tilde{d}_{ij})$ for all $j \in \{1, \ldots, n\}$, $i \in \{1, \ldots, m\}$, and the claim follows. □

The result shows that in a threshold model the copula of the critical variables determines the link between marginal probabilities of migration for individual firms and joint probabilities of migration for groups of firms. To illustrate this further, consider for simplicity a two-state model for default and non-default and a subgroup of k companies $\{i_1, \ldots, i_k\} \subset \{1, \ldots, m\}$ with individual default probabilities $p_{i_1}, \ldots, p_{i_k}$. Then

$$\begin{aligned} P(Y_{i_1} = 1, \ldots, Y_{i_k} = 1) &= P(X_{i_1} \leqslant d_{i_1 1}, \ldots, X_{i_k} \leqslant d_{i_k 1}) \\ &= C_{i_1 \cdots i_k}(p_{i_1}, \ldots, p_{i_k}), \end{aligned} \tag{11.4}$$

where $C_{i_1 \cdots i_k}$ denotes the corresponding k-dimensional margin of C. As a special case consider now a model for a single homogeneous group. We assume that $\boldsymbol{X}$ has an exchangeable copula (i.e. a copula of the form (7.20)) and that all individual default probabilities are equal to some constant π so that the default indicator vector $\boldsymbol{Y}$ is exchangeable. The formula (11.4) reduces to the useful formula

$$\pi_k = C_{1\cdots k}(\pi, \ldots, \pi), \quad 2 \leqslant k \leqslant m, \tag{11.5}$$

which will be used for the calibration of some copula models later on.

11.1.3 Gaussian Threshold Models

In this section we discuss the case where the critical variables have a Gauss copula.

Multivariate Merton model. It is straightforward to generalize the Merton model of Section 10.3.1 to a portfolio of m firms. We assume that the multivariate asset-value process $(\boldsymbol{V}_t)$ with $\boldsymbol{V}_t = (V_{t,1}, \dots, V_{t,m})'$ follows an m-dimensional geometric Brownian motion with drift vector $\boldsymbol{\mu}_V = (\mu_1, \dots, \mu_m)'$, vector of volatilities $\boldsymbol{\sigma}_V = (\sigma_1, \dots, \sigma_m)'$, and instantaneous correlation matrix P. This means that $(\boldsymbol{V}_t)$ solves the stochastic differential equations

$$\mathrm{d}V_{t,i} = \mu_i V_{t,i}\,\mathrm{d}t + \sigma_i V_{t,i}\,\mathrm{d}W_{t,i}, \quad i = 1, \dots, m,$$

for correlated Brownian motions with correlation $\rho(W_{t,i}, W_{t,j}) = \rho_{ij}$, $t \geqslant 0$. For all i the asset value $V_{T,i}$ is thus of the form

$$V_{T,i} = V_{0,i} \exp((\mu_i - \tfrac{1}{2}\sigma_i^2)T + \sigma_i W_{T,i}),$$

and $\boldsymbol{W}_T := (W_{T,1}, \dots, W_{T,m})'$ is a multivariate normal random vector satisfying $\boldsymbol{W}_T \sim N_m(\mathbf{0}, TP)$. In its basic form the Merton model is a default-only model in which the firm defaults if $V_{T,i} \leqslant B_i$ and B_i is the liability of firm i. Writing $\boldsymbol{B} = (B_1, \dots, B_m)'$, the threshold model representation is thus given by $(\boldsymbol{V}_T, \boldsymbol{B})$. Since in a threshold model the default event is invariant under strictly increasing transformations of critical variables and thresholds, this model is equivalent to a threshold model $(\boldsymbol{X}, \boldsymbol{d})$ with

$$X_i := \frac{\ln V_{T,i} - \ln V_{0,i} - (\mu_i - \frac{1}{2}\sigma_i^2)T}{\sigma_i\sqrt{T}},$$
$$d_i := \frac{\ln B_i - \ln V_{0,i} - (\mu_i - \frac{1}{2}\sigma_i^2)T}{\sigma_i\sqrt{T}}.$$

The transformed variables satisfy $\boldsymbol{X} \sim N_m(\mathbf{0}, P)$ and their copula is the Gauss copula C_P^{Ga}.

Gaussian threshold models in practice. In practice it is usual to start directly with threshold models of the form $(\boldsymbol{X}, \boldsymbol{d})$ with $\boldsymbol{X} \sim N_m(\mathbf{0}, P)$. There are two practical challenges: first, one has to calibrate the threshold vector $\boldsymbol{d}$ (or, in the case of a multi-state model, the threshold matrix D) in line with exogenously given default and transition probabilities; second, one needs to calibrate the correlation matrix P in a parsimonious way. The problem of calibrating the obligor-specific rows of the threshold matrix D to (rating) state transition probabilities was discussed in Section 10.3.4. In particular, in a default-only model we set $d_i = \Phi^{-1}(p_i)$ for given default probabilities p_i, for $i = 1, \dots, m$. Since $\boldsymbol{X}$ has standard normal margins, P is also the covariance matrix of $\boldsymbol{X}$.

In its most general form P has $m(m-1)/2$ distinct parameters. In portfolio credit risk applications m is typically large and it is important to use a more parsimonious parametrization of this matrix based on a factor model of the kind described in Section 6.4.1. Factor models also lend themselves to economic interpretation, and the

factors are commonly interpreted as country and industry effects. We now describe the mathematical form of typical factor models used in credit risk. The calibration of these models in practice is discussed in Section 11.5.1.

Factor models. We assume that

$$X_i = \sqrt{\beta_i}\tilde{F}_i + \sqrt{1-\beta_i}\varepsilon_i, \tag{11.6}$$

where $\tilde{F}_i$ and $\varepsilon_1, \ldots, \varepsilon_m$ are independent standard normal variables, and where $0 \leqslant \beta_i \leqslant 1$ for all i. In this formulation the $\tilde{F}_i$ are the *systematic* variables, which are correlated, and the ε_i are *idiosyncratic* variables. It follows that β_i can be viewed as a measure of the *systematic risk* of X_i: that is, the part of the variance of X_i that is explained by the systematic variable.

The systematic variables are assumed to be of the form $\tilde{F}_i = \boldsymbol{a}_i'\boldsymbol{F}$, where $\boldsymbol{F}$ is a vector of common factors satisfying $\boldsymbol{F} \sim N_p(\boldsymbol{0}, \Omega)$ with $p < m$, and where Ω is a correlation matrix. These factors typically represent country and industry effects. The assumption that $\mathrm{var}(\tilde{F}_i) = 1$ imposes the constraint that $\boldsymbol{a}_i'\Omega\boldsymbol{a}_i = 1$ for all i. Since $\mathrm{var}(X_i) = 1$ and since $\tilde{F}_i$ and $\varepsilon_1, \ldots, \varepsilon_m$ are independent and standard normal, the asset correlations in this model are given by

$$\rho(X_i, X_j) = \mathrm{cov}(X_i, X_j) = \sqrt{\beta_i\beta_j}\,\mathrm{cov}(\tilde{F}_i, \tilde{F}_j) = \sqrt{\beta_i\beta_j}\boldsymbol{a}_i'\Omega\boldsymbol{a}_j.$$

In order to set up the model we have to determine $\boldsymbol{a}_i$ and β_i for each obligor as well as Ω, with the additional constraint that $\boldsymbol{a}_i'\Omega\boldsymbol{a}_i = 1$ for all i. Since Ω has $p(p-1)/2$ parameters, the loading vectors $\boldsymbol{a}_i$ and coefficients β_i have a combined total of $mp + m$ parameters, and we are applying m constraints, the dimension of the calibration problem is $mp + p(p-1)/2$. In particular, the number of parameters grows linearly rather than quadratically in m.

Note that this factor model does fit into the general framework developed in Section 6.4.1. $\boldsymbol{X}$ can be written as

$$\boldsymbol{X} = B\boldsymbol{F} + \tilde{\boldsymbol{\varepsilon}}, \tag{11.7}$$

where $B = DA$, $D = \mathrm{diag}(\sqrt{\beta}_1, \ldots, \sqrt{\beta}_m)$, $A \in \mathbb{R}^{m\times p}$ is the matrix with ith row given by $\boldsymbol{a}_i'$, and $\tilde{\varepsilon}_i = \sqrt{1-\beta_i}\varepsilon_i$.

We often consider the special case of a one-factor model. This corresponds to a model where $\tilde{F}_i = F$ for a single common standard normal factor so that the equation in (11.6) takes the form

$$X_i = \sqrt{\beta_i}F + \sqrt{1-\beta_i}\varepsilon_i. \tag{11.8}$$

If, moreover, every obligor has the same systematic variance $\beta_i = \rho$, we get that $\rho(X_i, X_j) = \rho$ for all $i \neq j$. This model is often referred to as an equicorrelation model and was introduced previously in Example 6.32 and equation (6.53).

11.1.4 Models Based on Alternative Copulas

While most threshold models used in industry are based explicitly or implicitly on the Gauss copula, there is no reason why we have to assume a Gauss copula. In fact,

simulations presented in Section 11.1.5 show that the choice of copula may be very critical to the tail of the distribution of the number of defaults M. We now look at threshold models based on alternative copulas.

The first class of models attempts to preserve some of the flexibility of the Gaussian threshold models, which do have the appealing feature that they can accommodate a wide range of different correlation structures for the critical variables. This is clearly an advantage in modelling a portfolio where obligors are exposed to several risk factors and where the exposure to different risk factors differs markedly across obligors, such as a portfolio of loans to companies from different industry sectors or countries.

Example 11.3 (normal mean–variance mixtures). For the distribution of the critical variables we consider the kind of model described in Section 6.2.2. We start with an m-dimensional multivariate normal vector $\boldsymbol{Z} \sim N_m(\boldsymbol{0}, \Sigma)$ and a positive, scalar rv W, which is independent of $\boldsymbol{Z}$. The vector of critical variables $\boldsymbol{X}$ is assumed to have the structure

$$\boldsymbol{X} = \boldsymbol{m}(W) + \sqrt{W}\boldsymbol{Z}, \tag{11.9}$$

where $\boldsymbol{m}\colon [0, \infty) \to \mathbb{R}^m$ is a measurable function. In the special case where $\boldsymbol{m}(W)$ takes a constant value $\boldsymbol{\mu}$ not depending on W, the distribution is called a normal variance mixture. An important example of a normal variance mixture is the multivariate t distribution, as discussed in Example 6.7, which is obtained when W has an inverse gamma distribution, $W \sim \mathrm{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$, or equivalently when $\nu/W \sim \chi^2_\nu$. An example of a general mean–variance mixture is the GH distribution discussed in Section 6.2.3.

In a normal mean–variance mixture model the default condition may be written in the form

$$X_i \leqslant d_{i1} \iff Z_i \leqslant \frac{d_{i1}}{\sqrt{W}} - \frac{m_i(W)}{\sqrt{W}} =: \tilde{D}_i, \tag{11.10}$$

where $m_i(W)$ is the ith component of $\boldsymbol{m}(W)$. A possible economic interpretation of the model (11.9) is to consider Z_i as the asset value of company i and d_{i1} as an a priori estimate of the corresponding default threshold. The actual default threshold is *stochastic* and is represented by $\tilde{D}_i$, which is obtained by applying a multiplicative shock and an additive shock to the estimate d_{i1}. If we interpret this shock as a stylized representation of global factors such as the overall liquidity and risk appetite in the banking system, it makes sense to assume that the shocks to the default thresholds of different obligors are driven by the same rv W.

Normal variance mixtures, such as the multivariate t, provide the most tractable examples of normal mean–variance mixtures; they admit a calibration approach using linear factor models that is similar to the approach used for models based on the Gauss copula. In normal variance mixture models the correlation matrices of $\boldsymbol{X}$ (when defined) and $\boldsymbol{Z}$ coincide. Moreover, if $\boldsymbol{Z}$ follows a linear factor model (11.7), then $\boldsymbol{X}$ inherits the linear factor structure from $\boldsymbol{Z}$. Note, however, that the systematic factors $\sqrt{W}\boldsymbol{F}$ and the idiosyncratic factors $\sqrt{W}\boldsymbol{\varepsilon}$ are no longer independent but merely uncorrelated.

The class of threshold models based on the t copula can be thought of as containing the Gaussian threshold models as limiting cases when $\nu \to \infty$. However, the additional parameter ν adds a great deal of flexibility. We will come back to this point in Section 11.1.5.

Another class of parametric copulas that could be used in threshold models is the Archimedean family of Section 7.4.

Example 11.4 (Archimedean copulas). Recall that an Archimedean copula is the distribution function of a uniform random vector of the form

$$C(u_1, \ldots, u_m) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_m)), \tag{11.11}$$

where $\psi : [0, \infty) \to [0, 1]$ is a continuous, decreasing function, known as the copula generator, satisfying $\psi(0) = 1$ and $\lim_{t\to\infty} \psi(t) = 0$, and ψ^{-1} is its inverse. We assume that ψ is completely monotonic (see equation (7.47) and surrounding discussion). As explained in Section 7.4, these conditions ensure that (11.11) defines a copula for any portfolio size m. Our main example in this chapter is the Clayton copula. Recall from Section 7.4 that this copula has generator $\psi_\theta(t) = (1 + \theta t)^{-1/\theta}$, where $\theta > 0$, leading to the expression

$$C_\theta^{\mathrm{Cl}}(u_1, \ldots, u_m) = (u_1^{-\theta} + \cdots + u_m^{-\theta} + 1 - m)^{-1/\theta}. \tag{11.12}$$

As discussed in Section 7.4, exchangeable Archimedean copulas suffer from the deficiency that they are not rich in parameters and can model only exchangeable dependence and not a fully flexible dependence structure for the critical variables. Nonetheless, they yield useful parsimonious models for relatively small homogeneous portfolios, which are easy to calibrate and simulate, as we discuss in more detail in Section 11.2.4.

Suppose that $\boldsymbol{X}$ is a random vector with an Archimedean copula and marginal distributions F_{X_i}, $1 \leqslant i \leqslant m$, so that $(\boldsymbol{X}, \boldsymbol{d})$ specifies a threshold model with individual default probabilities $F_{X_i}(d_i)$. As a particular example consider the Clayton copula and assume a homogeneous situation where all individual default probabilities are identical to π. Using equations (11.5) and (11.12) we can calculate that the probability that an arbitrarily selected group of k obligors from a portfolio of m such obligors defaults over the time horizon is given by $\pi_k = (k\pi^{-\theta} - k + 1)^{-1/\theta}$. Essentially, the dependent default mechanism of the homogeneous group is now determined by this equation and the parameters π and θ. We study this Clayton copula model further in Example 11.13.

11.1.5 Model Risk Issues

Model risk is the risk associated with working with misspecified models—in our case, models that are a poor representation of the true mechanism governing defaults and migrations in a credit portfolio. For example, if we intend to use our models to estimate measures of tail risk, like VaR and expected shortfall, then we should be particularly concerned with the possibility that they might underestimate the tail of the portfolio loss distribution.

Table 11.1. Results of simulation study. We tabulate the estimated 95th and 99th percentiles of the distribution of M in an exchangeable model with 10 000 firms. The values for the default probability π and the asset correlation ρ corresponding to the three groups A, B and C are given in the text.

	$q_{0.95}(M)$			$q_{0.99}(M)$		
Group	$\nu = \infty$	$\nu = 50$	$\nu = 10$	$\nu = \infty$	$\nu = 50$	$\nu = 10$
A	14	23	24	21	49	118
B	109	153	239	157	261	589
C	1618	1723	2085	2206	2400	3067

As we have seen, a threshold model essentially consists of a collection of default (and migration) probabilities for individual firms and a copula that describes the dependence of certain critical variables. In discussing model risk in this context we will concentrate on models for default only and assume that individual default probabilities have been satisfactorily determined. It is much more difficult to determine the copula describing default dependence and we will look at model risk associated with the misspecification of this component of the threshold model. See also Section 8.4.4 for a discussion of the issue of dependence uncertainty.

The impact of the choice of copula. Since most threshold models used in practice use the Gauss copula, we are particularly interested in the sensitivity of the distribution of the number of defaults M with respect to the assumption of Gaussian dependence. Our interest is motivated by the observation made in Section 7.3.1 that, by assuming a Gaussian dependence structure, we may underestimate the probability of joint large movements of risk factors, with potentially drastic implications for the performance of risk-management models.

We compare a simple exchangeable model with multivariate normal critical variables and a model where the critical variables are multivariate t. Given a standard normal rv F, an iid sequence $\varepsilon_1, \ldots, \varepsilon_m$ of standard normal variates independent of F, and an asset correlation parameter $\rho \in [0, 1]$, we define a random vector $\boldsymbol{Z}$ by $Z_i = \sqrt{\rho}F + \sqrt{1-\rho}\varepsilon_i$. Observe that this is the equicorrelation special case of the factor model (11.8).

In the t copula case we define the critical variables $X_i := \sqrt{W}Z_i$, where $W \sim \mathrm{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$ is independent of $\boldsymbol{Z}$, so that $\boldsymbol{X}$ has a multivariate t distribution. In the Gauss copula case we simply set $\boldsymbol{X} := \boldsymbol{Z}$. In both cases we choose thresholds so that $P(Y_i = 1) = \pi$ for all i and for some $\pi \in (0, 1)$. Note that the correlation matrix P of $\boldsymbol{X}$ (the asset correlation matrix) is identical in both models and is given by an equicorrelation matrix with off-diagonal element ρ. However, the copula of $\boldsymbol{X}$ differs, and we expect more joint defaults in the t model due to the higher level of dependence in the joint tail of the t copula.

We consider three portfolios of decreasing credit quality, labelled A, B and C. In group A we set $\pi = 0.06\%$ and $\rho = 2.58\%$; in group B we set $\pi = 0.50\%$ and $\rho = 3.80\%$; in group C we set $\pi = 7.50\%$ and $\rho = 9.21\%$. We consider a portfolio of size $m = 10\,000$. For each group we vary the degrees-of-freedom

Table 11.2. Results of simulation study. Estimated 95th and 99th percentiles of the distribution of M in an exchangeable model for varying values of asset correlation ρ.

Quantile	$\rho = 2.58\%$	$\rho = 3.80\%$	$\rho = 9.21\%$
$q_{0.95}(M)$	98	109	148
$q_{0.99}(M)$	133	157	250

parameter ν. In order to represent the tail of the number of defaults M, we use simulations to determine (approximately) the 95% and 99% quantiles, $q_{0.95}(M)$ and $q_{0.99}(M)$, and tabulate them in Table 11.1. The actual simulation was performed using a representation of threshold models as Bernoulli mixture models that is discussed later in Section 11.2.4.

Table 11.1 shows that ν clearly has a massive influence on the high quantiles. For the important 99% quantile the impact is most pronounced for group A, where $q_{0.99}(M)$ is increased by a factor of almost six when we go from a Gaussian model to a model with $\nu = 10$.

The impact of changing asset correlation. Here we retain the assumption that $\boldsymbol{X}$ has a Gauss copula and study the impact of the factor structure of the asset returns on joint default events and hence on the tail of M. More specifically, we increase the systematic risk component of the critical variables for the obligors in our portfolio and analyse how this affects the tail of M. We use the exchangeable model introduced above. We fix the default probability at $\pi = 0.50\%$ (the value for group B above) and vary the asset correlation ρ using the values $\rho = 2.58\%$, $\rho = 3.80\%$ and $\rho = 9.21\%$. In Table 11.2 we tabulate $q_{0.95}(M)$ and $q_{0.99}(M)$ for a portfolio with 10 000 counterparties. Clearly, varying ρ also has a sizeable effect on the quantiles of M. However, this effect is less dramatic and, in particular, less surprising than the impact of varying the copula in our previous experiment.

Both simulation experiments suggest that the loss distributions implied by threshold models are very sensitive to the copula of the critical variables. For this reason a substantial effort should be devoted to the calibration of the dependence model for the critical variables. Moreover, it is important to conduct sensitivity analyses to understand the implications of model risk for risk capital calculations.

Notes and Comments

Our presentation of threshold models is based, to a large extent, on Frey and McNeil (2001, 2003). In those papers we referred to the models as "latent variable" models, because of structural similarities with statistical models of that name (see Joe 1997). However, whereas in statistical latent variable models the critical variables are treated as unobserved, in credit models they are often formally identified, e.g. as asset values or asset-value returns.

To see how the factor modelling approaches used in industry correspond to our presentation in Section 11.1.3 readers should consult the CreditMetrics technical document (RiskMetrics Group 1997) and the description of Moody's GCorr model in Huang et al. (2012). The latter model is used to model correlations between

changes in credit quality for many different kinds of obligor including publicly traded firms, private firms, small and medium-sized enterprises (SMEs) and retail borrowers. For public firms, weekly asset returns, calculated as part of the public-firm EDF methodology described in Section 10.3.3, are used as the measure of changing credit quality or "ability-to-pay".

The first systematic study of model risk for credit portfolio models is Gordy (2000). Our analysis of the impact of the copula of $\boldsymbol{X}$ on the tail of M follows Frey, McNeil and Nyfeler (2001). For an excellent discussion of various aspects of model risk in risk management in general, we refer to Gibson (2000).

11.2 Mixture Models

In a mixture model the default risk of an obligor is assumed to depend on a set of common factors, usually interpreted as macroeconomic variables, which are also modelled stochastically. Given a realization of the factors, defaults of individual firms are assumed to be independent. Dependence between defaults stems from the dependence of individual default probabilities on the set of common factors. We start with a general definition of a Bernoulli mixture model in Section 11.2.1 before looking in detail at the important special case of one-factor Bernoulli mixture models in Section 11.2.2. In Section 11.2.4 we show that many threshold models can be represented as Bernoulli mixtures, and in Section 11.2.5 we discuss the approximation of Bernoulli mixture models through Poisson mixture models and the important example of CreditRisk$^+$.

11.2.1 Bernoulli Mixture Models

Definition 11.5 (Bernoulli mixture model). Given some $p < m$ and a p-dimensional random vector $\boldsymbol{\Psi} = (\Psi_1, \ldots, \Psi_p)'$, the random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_m)'$ follows a Bernoulli mixture model with factor vector $\boldsymbol{\Psi}$ if there are functions $p_i : \mathbb{R}^p \to [0, 1]$, $1 \leqslant i \leqslant m$, such that, conditional on $\boldsymbol{\Psi}$, the components of $\boldsymbol{Y}$ are independent Bernoulli rvs satisfying $P(Y_i = 1 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = p_i(\boldsymbol{\psi})$.

For $\boldsymbol{y} = (y_1, \ldots, y_m)'$ in $\{0, 1\}^m$ we have that

$$P(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = \prod_{i=1}^{m} p_i(\boldsymbol{\psi})^{y_i}(1 - p_i(\boldsymbol{\psi}))^{1-y_i}, \tag{11.13}$$

and the unconditional distribution of the default indicator vector $\boldsymbol{Y}$ is obtained by integrating over the distribution of the factor vector $\boldsymbol{\Psi}$. In particular, the default probability of company i is given by $p_i = P(Y_i = 1) = E(p_i(\boldsymbol{\Psi}))$.

Note that the two-stage hierarchical structure of a Bernoulli mixture model facilitates sampling from the model: first we generate the economic factor realizations, then we generate the pattern of defaults conditional on those realizations. The second step is easy because of the conditional independence assumption.

In general, Bernoulli mixture models have a number of computational advantages. Consider the portfolio loss $L = \sum_{i=1}^{m} e_i \delta_i Y_i$ in the case where the exposures e_i and LGDs δ_i are deterministic. While it is difficult to compute the df F_L of L, it is easy to

use the conditional independence of the defaults to show that the Laplace–Stieltjes transform of F_L for $t \in \mathbb{R}$ is given by

$$\begin{aligned}\hat{F}_L(t) = E(\mathrm{e}^{-tL}) = E(E(\mathrm{e}^{-tL} \mid \boldsymbol{\Psi})) &= E\left(E\left(\exp\left(-t\sum_{i=1}^{m} e_i\delta_i Y_i\right) \,\middle|\, \boldsymbol{\Psi}\right)\right) \\ &= E\left(\prod_{i=1}^{m} E(\mathrm{e}^{-te_i\delta_i Y_i} \mid \boldsymbol{\Psi})\right) \\ &= E\left(\prod_{i=1}^{m} (p_i(\boldsymbol{\Psi})\mathrm{e}^{-te_i\delta_i} + 1 - p_i(\boldsymbol{\Psi}))\right),\end{aligned}$$

which can also be obtained by integrating over the distribution of the factors $\boldsymbol{\Psi}$. The Laplace–Stieltjes transform is useful in a number of practical tasks relating to Bernoulli mixture models, as follows.

- To implement an efficient Monte Carlo scheme for sampling losses from a Bernoulli mixture model we often use importance sampling. For this we need the moment-generating function of L, which can be calculated from the Laplace–Stieltjes transform according to $M_L(t) = E(\mathrm{e}^{Lt}) = \hat{F}_L(-t)$ (see Section 11.4 for more details).
- The probability mass function of L may be calculated by using the inverse Fourier transform to invert the characteristic function of L given by $\phi_L(t) = E(\mathrm{e}^{\mathrm{i}Lt})$. The characteristic function has the same functional form as the Laplace–Stieltjes transform $\hat{F}_L(t)$ but with the imaginary argument $-\mathrm{i}t$ (see also the discussion after Theorem 11.16).

11.2.2 One-Factor Bernoulli Mixture Models

One-factor models, i.e. models where Ψ is one dimensional, are particularly important special cases because of their tractability. Their behaviour for large portfolios is particularly easy to understand, as will be shown in Section 11.3, and this has had an influence on the Basel capital framework. Moreover, they have relatively few parameters and are thus easier to estimate from data. Throughout this section we consider an rv Ψ with values in $\mathbb{R}$ and functions $p_i(\Psi)\colon \mathbb{R} \to [0, 1]$ such that, conditional on Ψ, the default indicator $\boldsymbol{Y}$ is a vector of independent Bernoulli rvs with $P(Y_i = 1 \mid \Psi = \psi) = p_i(\psi)$. We now consider a variety of special cases.

Exchangeable Bernoulli mixture models. A further simplification occurs if the functions p_i are all identical. In this case the Bernoulli mixture model is termed *exchangeable*, since the random vector $\boldsymbol{Y}$ is exchangeable. It is convenient to introduce the rv $Q := p_1(\Psi)$ and to denote the distribution function of this mixing variable by $G(q)$. Conditional on $Q = q$, the number of defaults M is the sum of m independent Bernoulli variables with parameter q and it therefore has a binomial distribution with parameters q and m, i.e. $P(M = k \mid Q = q) = \binom{m}{k} q^k (1-q)^{m-k}$. The unconditional distribution of M is obtained by integrating over q. We have

$$P(M = k) = \binom{m}{k} \int_0^1 q^k (1-q)^{m-k} \, \mathrm{d}G(q). \tag{11.14}$$

Using the notation of Section 11.1.1 we can calculate default probabilities and joint default probabilities for the exchangeable group. Simple calculations give $\pi = E(Y_1) = E(E(Y_1 \mid Q)) = E(Q)$ and, more generally,

$$\pi_k = P(Y_1 = 1, \ldots, Y_k = 1) = E(E(Y_1 \cdots Y_k \mid Q)) = E(Q^k), \qquad (11.15)$$

so that unconditional default probabilities of first and higher order are seen to be moments of the mixing distribution. Moreover, for $i \neq j$, $\operatorname{cov}(Y_i, Y_j) = \pi_2 - \pi^2 = \operatorname{var}(Q) \geqslant 0$, which means that in an exchangeable Bernoulli mixture model the default correlation ρ_Y defined in (11.3) is always non-negative. Any value of ρ_Y in $[0, 1]$ can be obtained by an appropriate choice of the mixing distribution G. In particular, if $\rho_Y = \operatorname{var}(Q) = 0$, the rv Q has a degenerate distribution with all mass concentrated on the point π and the default indicators are independent. The case $\rho_Y = 1$ corresponds to a model where $\pi = \pi_2$ and the distribution of Q is concentrated on the points 0 and 1.

Example 11.6 (beta, probit-normal and logit-normal mixtures). The following mixing distributions are frequently encountered in Bernoulli mixture models.

Beta mixing distribution. In this model $Q \sim \text{Beta}(a, b)$ for some parameters $a > 0$ and $b > 0$. See Section A.2.1 for more details concerning the beta distribution.

Probit-normal mixing distribution. Here, $Q = \Phi(\mu + \sigma\Psi)$ for $\Psi \sim N(0, 1)$, $\mu \in \mathbb{R}$ and $\sigma > 0$, where Φ is the standard normal distribution function. We show later, in Section 11.2.4, that this model is equivalent to an exchangeable version of the one-factor Gaussian threshold model in (11.8).

Logit-normal mixing distribution. Here, $Q = F(\mu + \sigma\Psi)$ for $\Psi \sim N(0, 1)$, $\mu \in \mathbb{R}$ and $\sigma > 0$, where $F(x) = (1 + \mathrm{e}^{-x})^{-1}$ is the df of a so-called logistic distribution.

In the model with beta mixing distribution, the higher-order default probabilities π_k and the distribution of M can be computed explicitly (see Example 11.7 below). Calculations for the logit-normal, probit-normal and other models generally require numerical evaluation of the integrals in (11.14) and (11.15). If we fix any two of π, π_2 and ρ_Y in a beta, logit-normal or probit-normal model, then this fixes the parameters a and b or μ and σ of the mixing distribution, and higher-order joint default probabilities are automatically determined.

Example 11.7 (beta mixing distribution). By definition, the density of a beta distribution is given by

$$g(q) = \frac{1}{\beta(a, b)} q^{a-1}(1 - q)^{b-1}, \quad a, b > 0,\ 0 < q < 1,$$

where $\beta(a, b)$ denotes the beta function. Below we use the fact that the beta function satisfies the recursion formula $\beta(a + 1, b) = (a/(a + b))\beta(a, b)$; this is easily established from the representation of the beta function in terms of the gamma

function in Section A.2.1. Using (11.15), for the higher-order default probabilities we obtain

$$\pi_k = \frac{1}{\beta(a,b)} \int_0^1 q^k q^{a-1}(1-q)^{b-1}\,\mathrm{d}q = \frac{\beta(a+k,b)}{\beta(a,b)}, \quad k = 1, 2, \ldots.$$

The recursion formula for the beta function yields $\pi_k = \prod_{j=0}^{k-1}(a+j)/(a+b+j)$; in particular, $\pi = a/(a+b)$, $\pi_2 = \pi(a+1)/(a+b+1)$ and $\rho_Y = (a+b+1)^{-1}$. The rv M has a so-called beta-binomial distribution. From (11.14) we obtain

$$\begin{aligned} P(M=k) &= \binom{m}{k} \frac{1}{\beta(a,b)} \int_0^1 q^{k+a-1}(1-q)^{m-k+b-1}\,\mathrm{d}q \\ &= \binom{m}{k} \frac{\beta(a+k, b+m-k)}{\beta(a,b)}. \end{aligned} \tag{11.16}$$

One-factor models with covariates. It is straightforward to extend the one-factor probit-normal and logit-normal mixture models to include covariates that influence default probability and default correlation; these covariates might be indicators for group membership, such as a rating class or industry sector, or key ratios taken from a company's balance sheet.

Writing $\boldsymbol{x}_i \in \mathbb{R}^k$ for a vector of deterministic covariates, a general model for the conditional default probabilities $p_i(\Psi)$ in (11.13) would be to assume that

$$\left.\begin{aligned} p_i(\Psi) &= h(\mu_i + \sigma_i \Psi), \\ \mu_i &= \mu + \boldsymbol{\beta}' \boldsymbol{x}_i, \\ \sigma_i &= \exp(\delta + \boldsymbol{\gamma}' \boldsymbol{x}_i), \end{aligned}\right\} \tag{11.17}$$

where $\Psi \sim N(0,1)$, $h(x) = \Phi(x)$ or $h(x) = (1+\mathrm{e}^{-x})^{-1}$, the vectors $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)'$ contain regression parameters, and $\mu \in \mathbb{R}$ and $\delta \in \mathbb{R}$ are intercept parameters. Similar specifications are commonly used in the class of generalized linear models in statistics (see Section 11.5.3).

Clearly, if $\boldsymbol{x}_i = \boldsymbol{x}$ for all i, so that all risks have the same covariates, then we are back in the situation of full exchangeability. Note also that, since the function $p_i(\Psi)$ is increasing in Ψ, the conditional default probabilities $(p_1(\Psi), \ldots, p_m(\Psi))$ form a comonotonic random vector; hence, in a state of the world where the default probability is comparatively high for one counterparty, it is high for all counterparties. For a discussion of comonotonicity we refer to Section 7.2.1.

Example 11.8 (model for several exchangeable groups). The regression structure in (11.17) includes partially exchangeable models, where we define a number of groups within which risks are exchangeable. These groups might represent rating classes according to some internal or rating-agency classification.

If the covariates $\boldsymbol{x}_i$ are simply k-dimensional unit vectors of the form $\boldsymbol{x}_i = \boldsymbol{e}_{r(i)}$, where $r(i) \in \{1, \ldots, k\}$ indicates, say, the rating class of firm i, then the model (11.17) can be written in the form

$$p_i(\Psi) = h(\mu_{r(i)} + \sigma_{r(i)} \Psi) \tag{11.18}$$

for parameters $\mu_r := \mu + \beta_r$ and $\sigma_r := \mathrm{e}^{\delta+\gamma_r}$ for $r = 1, \ldots, k$.

Inserting this specification into (11.13) allows us to find the conditional distribution of the default indicator vector. Suppose there are m_r obligors in rating category r for $r = 1, \ldots, k$, and write M_r for the number of defaults. The conditional distribution of the vector $\boldsymbol{M} = (M_1, \ldots, M_k)'$ is given by

$$P(\boldsymbol{M} = \boldsymbol{l} \mid \Psi = \psi) = \prod_{r=1}^{k} \binom{m_r}{l_r} (h(\mu_r + \sigma_r \psi))^{l_r} (1 - h(\mu_r + \sigma_r \psi))^{m_r - l_r}, \tag{11.19}$$

where $\boldsymbol{l} = (l_1, \ldots, l_k)'$. A model of the form (11.19) with $\sigma_1 = \cdots = \sigma_k$ will be fitted to Standard & Poor's default data in Section 11.5.4. The asymptotic behaviour of such a model (when m is large) is investigated in Example 11.20.

11.2.3 Recovery Risk in Mixture Models

In standard portfolio risk models it is assumed that the loss given default is independent of the default event. This is likely to be an oversimplification, as economic intuition suggests that recovery rates depend on risk factors similar to those for default probabilities; in that case one speaks of *systematic recovery risk*. Consider, for instance, the market for mortgages. During a property crisis many mortgages default. At the same time property prices are low, so that real estate can be sold only for very low prices in a foreclosure (a forced sale in which a bank liquidates a property it holds as collateral), so that recovery rates are low.

The presence of systematic recovery risk is confirmed in a number of empirical studies. Among others, Frye (2000) has carried out a formal empirical analysis using recovery data collected by Moody's on rated corporate bonds. He found that recovery rates are substantially lower than average in times of economic recession. To quote from his paper:

> Using that data [the Moody's data] to estimate an appropriate credit model, we can extrapolate that in a severe economic downturn recoveries might decline 20–25 percentage points from the normal-year average. This could cause loss given default to increase by nearly 100% and to have a similar effect on economic capital. Such systematic recovery risk is absent from first-generation credit risk models. Therefore these models may significantly understate the capital required at banking institutions.

In a similar vein, Hamilton et al. (2005) estimated formal models for the relationship between one-year default rate q and recovery rate R for corporate bonds; according to their analysis the best-fitting relationship is $R(q) \approx (0.52 - 6.9q)^+$.

Clearly, these findings call for the inclusion of systematic recovery risk in standard credit risk models. This is easily accomplished in the mixture-model framework—we replace the constant δ_i with some function $\delta_i(\boldsymbol{\psi})$—but the challenge lies in the estimation of the function $\delta_i(\cdot)$ describing the relationship between loss given default and the systematic factors.

11.2.4 Threshold Models as Mixture Models

Although the mixture models of this section seem, at first glance, to be different in structure from the threshold models of Section 11.1, it is important to realize that the majority of useful threshold models, including all the examples we have given, can be represented as Bernoulli mixture models. This is a very useful insight, because the Bernoulli mixture format has a number of advantages over the threshold format.

- Bernoulli mixture models lend themselves to Monte Carlo risk studies. From the analyses of this section we obtain methods for sampling from many of the models we have discussed, such as the t copula threshold model used in Section 11.1.5.
- Mixture models are arguably more convenient for statistical fitting purposes. We show in Section 11.5.3 that statistical techniques for generalized linear mixed models can be used to fit mixture models to empirical default data gathered over several time periods.
- The large-portfolio behaviour of Bernoulli mixtures can be analysed and understood in terms of the behaviour of the distribution of the common economic factors, as will be shown in Section 11.3.

To motivate the subsequent analysis we begin by computing the mixture model representation of the simple one-factor Gaussian threshold model in (11.8). It is convenient to identify the variable Ψ in the mixture representation with minus the factor F in the threshold representation; this yields conditional default probabilities that are increasing in Ψ and leads to formulas that are in line with the Basel IRB formula. With $F = -\Psi$ the one-factor model takes the form

$$X_i = -\sqrt{\beta_i}\Psi + \sqrt{1-\beta_i}\varepsilon_i.$$

By definition, company i defaults if and only if $X_i \leqslant d_i$ and hence if and only if $\sqrt{1-\beta_i}\varepsilon_i \leqslant d_i + \sqrt{\beta_i}\Psi$. Since the variables $\varepsilon_1, \ldots, \varepsilon_m$ and Ψ are independent, default events are independent conditional on Ψ and we can compute

$$\begin{aligned} p_i(\psi) = P(Y_i = 1 \mid \Psi = \psi) &= P(\sqrt{1-\beta_i}\varepsilon_i \leqslant d_i + \sqrt{\beta_i}\Psi \mid \Psi = \psi) \\ &= \Phi\left(\frac{d_i + \sqrt{\beta_i}\psi}{\sqrt{1-\beta_i}}\right), \end{aligned} \tag{11.20}$$

where we have used the fact that ε_i is standard normally distributed. The threshold is typically set so that the default probability matches an exogenously chosen value p_i, so that $d_i = \Phi^{-1}(p_i)$. In that case we obtain

$$p_i(\psi) = \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\beta_i}\psi}{\sqrt{1-\beta_i}}\right). \tag{11.21}$$

In the following we want to extend this idea to more general threshold models with a factor structure for the critical variables. We give a condition that ensures that a threshold model can be written as a Bernoulli mixture model.

Definition 11.9. A random vector $\boldsymbol{X}$ has a p-dimensional *conditional independence structure* with conditioning variable $\boldsymbol{\Psi}$ if there is some $p < m$ and a p-dimensional random vector $\boldsymbol{\Psi} = (\Psi_1, \ldots, \Psi_p)'$ such that, conditional on $\boldsymbol{\Psi}$, the rvs $X_1, \ldots, X_m$ are independent.

In the motivating example the conditioning variable was taken to be $\Psi = -F$. The next lemma generalizes the computations in (11.20) to any threshold model with a conditional independence structure.

Lemma 11.10. *Let $(\boldsymbol{X}, \boldsymbol{d})$ be a threshold model for an m-dimensional random vector $\boldsymbol{X}$. If $\boldsymbol{X}$ has a p-dimensional conditional independence structure with conditioning variable $\boldsymbol{\Psi}$, then the default indicators $Y_i = I_{\{X_i \leqslant d_i\}}$ follow a Bernoulli mixture model with factor $\boldsymbol{\Psi}$, where the conditional default probabilities are given by $p_i(\boldsymbol{\psi}) = P(X_i \leqslant d_i \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$.*

Proof. For $\boldsymbol{y} \in \{0,1\}^m$ define the set $B := \{1 \leqslant i \leqslant m \colon y_i = 1\}$ and let $B^{\mathrm{c}} = \{1, \ldots, m\} \backslash B$. We have

$$\begin{aligned} P(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) &= P\Bigg(\bigcap_{i \in B} \{X_i \leqslant d_i\} \bigcap_{i \in B^{\mathrm{c}}} \{X_i > d_i\} \;\Bigg|\; \boldsymbol{\Psi} = \boldsymbol{\psi} \Bigg) \\ &= \prod_{i \in B} P(X_i \leqslant d_i \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) \prod_{i \in B^{\mathrm{c}}} (1 - P(X_i \leqslant d_i \mid \boldsymbol{\Psi} = \boldsymbol{\psi})). \end{aligned}$$

Hence, conditional on $\boldsymbol{\Psi} = \boldsymbol{\psi}$, the Y_i are independent Bernoulli variables with success probability $p_i(\boldsymbol{\psi}) := P(X_i \leqslant d_i \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$. □

We now consider a number of examples.

Example 11.11 (Gaussian threshold model). Consider the general Gaussian threshold model with the factor structure in (11.6), which takes the form

$$X_i = \sqrt{\beta_i}\boldsymbol{a}_i'\boldsymbol{F} + \sqrt{1 - \beta_i}\varepsilon_i, \tag{11.22}$$

where $\varepsilon_1, \ldots, \varepsilon_m$ are iid standard normal and where $\mathrm{var}(\boldsymbol{a}_i'\boldsymbol{F}) = 1$ for all i. Conditional on $\boldsymbol{\Psi} = -\boldsymbol{F}$, the vector $\boldsymbol{X}$ is normally distributed with diagonal covariance matrix and thus has conditional independence structure. With $d_i = \Phi^{-1}(p_i)$ the conditional default probabilities are given by

$$\begin{aligned} p_i(\boldsymbol{\psi}) &= P(Y_i = 1 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = P(\sqrt{1 - \beta_i}\varepsilon_i \leqslant d_i + \sqrt{\beta_i}\boldsymbol{a}_i'\boldsymbol{\psi}) \\ &= \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\beta_i}\boldsymbol{a}_i'\boldsymbol{\psi}}{\sqrt{1 - \beta_i}} \right). \end{aligned} \tag{11.23}$$

By comparison with Example 11.6 we see that the individual stochastic default probabilities $p_i(\boldsymbol{\Psi})$ have a probit-normal distribution with parameters μ_i and σ_i given by

$$\mu_i = \Phi^{-1}(p_i)/\sqrt{1 - \beta_i} \quad \text{and} \quad \sigma_i^2 = \beta_i/(1 - \beta_i).$$

Example 11.12 (Student t threshold model). Now consider the case where the critical variables are of the form $\boldsymbol{X} = \sqrt{W}\boldsymbol{Z}$, where $\boldsymbol{Z}$ follows the Gaussian factor model in (11.22) and $W \sim \mathrm{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$. The vector $\boldsymbol{X}$ has a multivariate t distribution with ν degrees of freedom and standard univariate t margins with ν degrees of freedom.

This time we condition on $\boldsymbol{\Psi} = (-\boldsymbol{F}', W)'$. Given $\boldsymbol{\Psi} = (\tilde{\boldsymbol{\psi}}, w)$, the vector $\boldsymbol{X}$ has a multivariate normal distribution with independent components, and a computation similar to that in the previous example gives

$$p_i(\boldsymbol{\psi}) = p_i(\tilde{\boldsymbol{\psi}}, w) = \Phi\left(\frac{t_\nu^{-1}(p_i)w^{-1/2} + \sqrt{\beta_i}\boldsymbol{a}_i'\tilde{\boldsymbol{\psi}}}{\sqrt{1-\beta_i}}\right). \tag{11.24}$$

The formulas (11.23) and (11.24) are useful for Monte Carlo simulation of the corresponding threshold models. For example, rather than simulating an m-dimensional t distribution to implement the t model, one only needs to simulate a p-dimensional normal vector $\tilde{\boldsymbol{\Psi}}$ with $p \ll m$ and an independent gamma-distributed variate $V = W^{-1}$. In the second step of the simulation one simply conducts a series of independent Bernoulli experiments with default probabilities $p_i(\boldsymbol{\Psi})$ to decide whether individual companies default.

Application to Archimedean copula models. Another class of threshold models with an equivalent mixture representation is provided by models where the critical variables have an exchangeable LT-Archimedean copula in the sense of Definition 7.52. Consider a threshold model $(\boldsymbol{X}, \boldsymbol{d})$, where $\boldsymbol{X}$ has an exchangeable LT-Archimedean copula C with generator given by the Laplace transform $\hat{G}$ of some df G on $[0, \infty)$ with $G(0) = 0$. Let $\boldsymbol{p} = (p_1, \ldots, p_m)'$ denote the vector of default probabilities.

Consider now a non-negative rv $\Psi \sim G$ and rvs $U_1, \ldots, U_m$ that are conditionally independent given Ψ with conditional distribution function $P(U_i \leqslant u \mid \Psi = \psi) = \exp(-\psi \hat{G}^{-1}(u))$ for $u \in [0, 1]$. Proposition 7.51 then shows that $\boldsymbol{U}$ has df C. Moreover, by Lemma 11.2, $(\boldsymbol{X}, \boldsymbol{d})$ and $(\boldsymbol{U}, \boldsymbol{p})$ are two equivalent threshold models for default. By construction, $\boldsymbol{U}$ has a one-dimensional conditional independence structure with conditioning variable Ψ, and the conditional default probabilities are given by

$$p_i(\psi) = P(U_i \leqslant p_i \mid \Psi = \psi) = \exp(-\psi \hat{G}^{-1}(p_i)). \tag{11.25}$$

In order to simulate from a threshold model based on an LT-Archimedean copula we may therefore use the following efficient and simple approach. In a first step we simulate a realization ψ of Ψ and then we conduct m independent Bernoulli experiments with default probabilities $p_i(\psi)$ as in (11.25) to simulate a realization of the defaulting counterparties.

Example 11.13 (Clayton copula). As an example consider the Clayton copula with parameter $\theta > 0$. Suppose we wish to construct an exchangeable Bernoulli mixture model with default probability π and joint default probability π_2 that is equivalent to a threshold model with the Clayton copula for the critical variables. As mentioned

in Algorithm 7.53, a gamma-distributed rv $\Psi \sim \mathrm{Ga}(1/\theta, 1)$ (see Section A.2.4 for a definition) has Laplace transform $\hat{G}(t) = (1+t)^{-1/\theta}$. Using (11.25), the mixing variable of the equivalent Bernoulli mixture model can be defined by setting $Q = p_1(\Psi) = \exp(-\Psi(\pi^{-\theta} - 1))$.

Using (11.4), the required value of θ to give the desired joint default probabilities is the solution to the equation $\pi_2 = C_\theta(\pi, \pi) = (2\pi^{-\theta} - 1)^{-1/\theta}$, $\theta > 0$. It is easily seen that π_2 and, hence, the default correlation in our exchangeable Bernoulli mixture model are increasing in θ; for $\theta \to 0$ we obtain independent defaults and for $\theta \to \infty$ defaults become comonotonic and default correlation tends to one.

11.2.5 Poisson Mixture Models and CreditRisk$^+$

Since default is typically a rare event, it is possible to approximate Bernoulli indicator rvs for default with Poisson rvs and to approximate Bernoulli mixture models with Poisson mixture models. By choosing independent gamma distributions for the economic factors $\boldsymbol{\Psi}$ and using the Poisson approximation, we obtain a particularly tractable model for portfolio losses, known as CreditRisk$^+$.

Poisson approximation and Poisson mixture models. To be more precise, assume that, given the factors $\boldsymbol{\Psi}$, the default indicator variables $Y_1, \ldots, Y_m$ for a particular time horizon are conditionally independent Bernoulli variables satisfying $P(Y_i = 1 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = p_i(\boldsymbol{\psi})$. Moreover, assume that the distribution of $\boldsymbol{\Psi}$ is such that the conditional default probabilities $p_i(\boldsymbol{\psi})$ tend to be very small. In this case the Y_i variables can be approximated by conditionally independent Poisson variables $\tilde{Y}_i$ satisfying $\tilde{Y}_i \mid \boldsymbol{\Psi} = \boldsymbol{\psi} \sim \mathrm{Poi}(p_i(\boldsymbol{\psi}))$, since

$$P(\tilde{Y}_i = 0 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = \mathrm{e}^{-p_i(\boldsymbol{\psi})} \approx 1 - p_i(\boldsymbol{\psi}),$$
$$P(\tilde{Y}_i = 1 \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = p_i(\boldsymbol{\psi})\mathrm{e}^{-p_i(\boldsymbol{\psi})} \approx p_i(\boldsymbol{\psi}).$$

Moreover, the portfolio loss $L = \sum_{i=1}^m e_i \delta_i Y_i$ can be approximated by $\tilde{L} = \sum_{i=1}^m e_i \delta_i \tilde{Y}_i$. Of course, it is possible for a company to "default more than once" in the approximating Poisson model, albeit with a very low probability.

We now give a formal definition of a Poisson mixture model for counting variables that parallels the definition of a Bernoulli mixture model in Section 11.2.1.

Definition 11.14 (Poisson mixture model). Given some $p < m$ and a p-dimensional random vector $\boldsymbol{\Psi} = (\Psi_1, \ldots, \Psi_p)'$, the random vector $\tilde{\boldsymbol{Y}} = (\tilde{Y}_1, \ldots, \tilde{Y}_m)'$ follows a Poisson mixture model with factors $\boldsymbol{\Psi}$ if there are functions $\lambda_i : \mathbb{R}^p \to (0, \infty)$, $1 \leqslant i \leqslant m$, such that, conditional on $\boldsymbol{\Psi} = \boldsymbol{\psi}$, the random vector $\tilde{\boldsymbol{Y}}$ is a vector of independent Poisson distributed rvs with rate parameter $\lambda_i(\boldsymbol{\psi})$.

If $\tilde{\boldsymbol{Y}}$ follows a Poisson mixture model and if we define the indicators $Y_i = I_{\{\tilde{Y}_i \geqslant 1\}}$, then $\boldsymbol{Y}$ follows a Bernoulli mixture model and the mixing variables are related by $p_i(\cdot) = 1 - \mathrm{e}^{-\lambda_i(\cdot)}$.

The CreditRisk$^+$ model. The CreditRisk$^+$ model for credit risk was proposed by Credit Suisse Financial Products in 1997 (see Credit Suisse Financial Products 1997). It has the structure of the Poisson mixture model in Definition 11.14, where

the factor vector $\boldsymbol{\Psi}$ consists of p independent gamma-distributed rvs. The distributional assumptions and functional forms imposed in CreditRisk$^+$ make it possible to compute the distribution of the number of defaults and the aggregate portfolio loss fairly explicitly using techniques for compound distributions and mixture distributions that are well known in actuarial mathematics and which are also discussed in Chapter 13 (see Sections 13.2.2 and 13.2.4 in particular).

The (stochastic) parameter $\lambda_i(\boldsymbol{\Psi})$ of the conditional Poisson distribution for firm i is assumed to take the form

$$\lambda_i(\boldsymbol{\Psi}) = k_i \boldsymbol{w}_i' \boldsymbol{\Psi} \tag{11.26}$$

for a constant $k_i > 0$, for non-negative factor weights $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{ip})'$ satisfying $\sum_j w_{ij} = 1$, and for p independent Ga(α_j, β_j)-distributed factors $\Psi_1, \ldots, \Psi_p$ with parameters set to be $\alpha_j = \beta_j = \sigma_j^{-2}$ for $\sigma_j > 0$ and $j = 1, \ldots, p$. This parametrization of the gamma variables ensures that we have $E(\Psi_j) = 1$ and $\operatorname{var}(\Psi_j) = \sigma_j^2$.

It is easy to verify that

$$E(\tilde{Y}_i) = E(E(\tilde{Y}_i \mid \boldsymbol{\Psi})) = E(\lambda_i(\boldsymbol{\Psi})) = k_i E(\boldsymbol{w}_i' \boldsymbol{\Psi}) = k_i,$$

so that k_i is the expected number of defaults for obligor i over the time period. Setting $Y_i = I_{\{\tilde{Y}_i \geqslant 1\}}$ we also observe that

$$P(Y_i = 1) = E(P(\tilde{Y}_i > 0 \mid \boldsymbol{\Psi})) = E(1 - \exp(-k_i \boldsymbol{w}_i' \boldsymbol{\Psi})) \approx k_i E(\boldsymbol{w}_i' \boldsymbol{\Psi}) = k_i,$$

for k_i small, so that k_i is approximately equal to the default probability.

Remark 11.15. The exchangeable version of CreditRisk$^+$ is extremely close to an exchangeable Bernoulli mixture model with beta mixing distribution. To see this, observe that in the exchangeable case the implied Bernoulli mixture model for $\boldsymbol{Y}$ has mixing variable Q given by $Q = 1 - \mathrm{e}^{-k\Psi}$ for some $\Psi \sim \text{Ga}(\alpha, \beta)$, $k > 0$ and $\alpha = \beta$. For $q \in (0, 1)$ we therefore obtain

$$P(Q \leqslant q) = P(1 - \mathrm{e}^{-k\Psi} \leqslant q) = P\left(\Psi \leqslant -\frac{\ln(1-q)}{k}\right),$$

so that the densities g_Q and g_Ψ are related by $g_Q(q) = g_\Psi(-\ln(1-q)/k)/(k(1-q))$. Using the form of the density of the Ga(α, β) distribution we obtain

$$\begin{aligned} g_Q(q) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{k(1-q)} \left(\frac{-\ln(1-q)}{k}\right)^{\alpha-1} \exp\left(\frac{\beta \ln(1-q)}{k}\right) \\ &= \left(\frac{\beta}{k}\right)^\alpha \frac{1}{\Gamma(\alpha)} (-\ln(1-q))^{\alpha-1} (1-q)^{(\beta/k)-1}. \end{aligned}$$

In a realistic credit risk model the parameters are chosen in such a way that the mass of the distribution of Q is concentrated on values of q close to zero, since default is typically a rare event. For small q we may use the approximation $-\ln(1-q) \approx q$ to observe that the functional form of g_Q is extremely close to that of a beta distribution with parameters α and β/k, where we again recall that the model is parametrized to have $\alpha = \beta$.

Distribution of the number of defaults. In CreditRisk$^+$ we have that, given $\boldsymbol{\Psi} = \boldsymbol{\psi}$, $\tilde{Y}_i \sim \text{Poi}(k_i \boldsymbol{w}_i' \boldsymbol{\psi})$, which implies that the distribution of the number of defaults $\tilde{M} := \sum_{i=1}^m \tilde{Y}_i$ satisfies

$$\tilde{M} \mid \boldsymbol{\Psi} = \boldsymbol{\psi} \sim \text{Poi}\left(\sum_{i=1}^m k_i \boldsymbol{w}_i' \boldsymbol{\psi} \right), \tag{11.27}$$

since the sum of independent Poisson variables is also a Poisson variable with a rate parameter given by the sum of the rate parameters of the independent variables.

To compute the unconditional distribution of $\tilde{M}$ we require a well-known result on mixed Poisson distributions, which appears as Proposition 13.21 in a discussion of relevant actuarial methodology for quantitative risk management in Chapter 13. This result says that if the rv N is conditionally Poisson with a gamma-distributed rate parameter $\Lambda \sim \text{Ga}(\alpha, \beta)$, then N has a negative binomial distribution, $N \sim \text{NB}(\alpha, \beta/(\beta + 1))$.

In the case when $p = 1$ we may apply this result directly to (11.27) to deduce that $\tilde{M}$ has a negative binomial distribution (since a constant times a gamma variable remains gamma distributed). For arbitrary p we now show that $\tilde{M}$ is equal in distribution to a sum of p independent negative binomial rvs. This follows by observing that

$$\sum_{i=1}^m k_i \boldsymbol{w}_i' \boldsymbol{\Psi} = \sum_{i=1}^m k_i \sum_{j=1}^p w_{ij} \Psi_j = \sum_{j=1}^p \Psi_j \left(\sum_{i=1}^m k_i w_{ij} \right).$$

Now consider rvs $\tilde{M}_1, \ldots, \tilde{M}_p$ such that $\tilde{M}_j$ is conditionally Poisson with mean $(\sum_{i=1}^m k_i w_{ij}) \psi_j$ conditional on $\Psi_j = \psi_j$. The independence of the components $\Psi_1, \ldots, \Psi_p$ implies that the $\tilde{M}_j$ are independent, and by construction we have $\tilde{M} \stackrel{\text{d}}{=} \sum_{j=1}^p \tilde{M}_j$. Moreover, the rvs $(\sum_{i=1}^m k_i w_{ij}) \Psi_j$ are gamma distributed, so that each of the $\tilde{M}_j$ has a negative binomial distribution by Proposition 13.21.

Distribution of the aggregate loss. To obtain a tractable model, exposures are discretized in CreditRisk$^+$ using the concept of exposure bands. The CreditRisk$^+$ documentation (see Credit Suisse Financial Products 1997) suggests that the LGD can be subsumed in the exposure by multiplying the actual exposure by a value for the LGD that is typical for an obligor with the same credit rating. We will adopt this approach and assume that the losses arising from the individual obligors are of the form $\tilde{L}_i = e_i \tilde{Y}_i$, where the e_i are known (LGD-adjusted) exposures.

For all i, we discretize e_i in units of an amount ϵ; we replace e_i by a value $\ell_i \epsilon \geqslant e_i$, where ℓ_i is a positive integer multiplier. We now define exposure bands $b = 1, \ldots, n$ corresponding to the distinct values $\ell^{(1)}, \ldots, \ell^{(n)}$ for the multipliers. In other words, we group obligors in exposure bands according to the values of their discretized exposures.

Let s_b denote the set of indices for the obligors in exposure band b; this means that $i \in s_b \iff \ell_i = \ell^{(b)}$. Let $\tilde{L}^{(b)} = \sum_{i \in s_b} \epsilon \ell_i \tilde{Y}_i$ denote the aggregate loss in exposure band b. We have $\tilde{L}^{(b)} = \epsilon \ell^{(b)} \tilde{M}^{(b)}$, where $\tilde{M}^{(b)} = \sum_{i \in s_b} \tilde{Y}_i$ denotes the number of defaults in exposure band b. Let $\tilde{L} = \sum_{b=1}^n \tilde{L}^{(b)}$ denote the aggregate portfolio loss.

We now want to determine the distribution of $\tilde{L}$. The following theorem gives the necessary information for achieving this with Fourier inversion.

Theorem 11.16. *Let $\tilde{L}$ represent the aggregate loss in the general p-factor CreditRisk$^+$ model with exposures discretized into exposure bands as described above. The following then hold.*

(i) *The Laplace–Stieltjes transform of the df of $\tilde{L}$ is given by*

$$\hat{F}_{\tilde{L}}(s) = \prod_{j=1}^{p} \left(1 + \sigma_j^2 \sum_{i=1}^{m} k_i w_{ij} \left(1 - \sum_{b=1}^{n} \mathrm{e}^{-s\epsilon\ell^{(b)}} q_{jb}\right)\right)^{-\sigma_j^{-2}}, \qquad (11.28)$$

where $q_{jb} = \sum_{i \in s_b} k_i w_{ij} / \sum_{i=1}^{m} k_i w_{ij}$ for $b = 1, \ldots, n$.

(ii) *The distribution of $\tilde{L}$ has the structure $\tilde{L} \stackrel{d}{=} \sum_{j=1}^{p} Z_j$, where the Z_j are independent variables that follow a compound negative binomial distribution. More precisely, it holds that $Z_j \sim \mathrm{CNB}(\sigma_j^{-2}, \theta_j, G_{X_j})$ with $\theta_j = (1 + \sigma_j^2 \sum_{i=1}^{m} k_i w_{ij})^{-1}$ and G_{X_j} the df of a multinomial random variable X_j taking the value $\epsilon\ell^{(b)}$ with probability q_{jb}.*

Proof. The proof requires the mathematics of compound distributions as described in Section 13.2.2.

(i) Conditional on $\boldsymbol{\Psi} = \boldsymbol{\psi}$, $M^{(b)} = \sum_{i \in s_b} \tilde{Y}_i$ has a Poisson distribution with parameter

$$\lambda^{(b)}(\boldsymbol{\psi}) := \sum_{i \in s_b} \lambda_i(\boldsymbol{\psi}) = \sum_{i \in s_b} k_i \boldsymbol{w}_i' \boldsymbol{\psi},$$

and the loss $\tilde{L}^{(b)} = \epsilon\ell^{(b)} \tilde{M}^{(b)}$ in exposure band b has a compound Poisson distribution given by

$$\tilde{L}^{(b)} \mid \boldsymbol{\Psi} = \boldsymbol{\psi} \sim \mathrm{CPoi}(\lambda^{(b)}(\boldsymbol{\psi}), G^{(b)}),$$

where $G^{(b)}$ is the df of point mass at $\epsilon\ell^{(b)}$. It follows from Proposition 13.10 on sums of compound Poisson variables that

$$\tilde{L} \mid \boldsymbol{\Psi} = \boldsymbol{\psi} \sim \mathrm{CPoi}\left(\lambda(\boldsymbol{\psi}), \sum_{b=1}^{n} \frac{\lambda^{(b)}(\boldsymbol{\psi})}{\lambda(\boldsymbol{\psi})} G^{(b)}\right), \qquad (11.29)$$

where $\lambda(\boldsymbol{\psi}) := \sum_{b=1}^{n} \lambda^{(b)}(\boldsymbol{\psi})$. The (conditional) severity distribution in (11.29) is the distribution function of a multinomial random variable that takes the values $\epsilon\ell^{(b)}$ with probabilities $\lambda^{(b)}(\boldsymbol{\psi})/\lambda(\boldsymbol{\psi})$ for $b = 1, \ldots, n$.

Writing $\hat{F}_{L|\boldsymbol{\Psi}}(s \mid \boldsymbol{\psi})$ for the Laplace–Stieltjes transform of the conditional distribution function of $\tilde{L}$ given $\boldsymbol{\Psi}$, we can use equation (13.11) and Example 13.5 to

infer that

$$\begin{aligned}\hat{F}_{\tilde{L}|\boldsymbol{\Psi}}(s \mid \boldsymbol{\psi}) &= \exp\left(-\lambda(\boldsymbol{\psi})\left(1 - \sum_{b=1}^{n} \frac{\lambda^{(b)}(\boldsymbol{\psi})}{\lambda(\boldsymbol{\psi})} \mathrm{e}^{-s\epsilon\ell^{(b)}}\right)\right) \\ &= \exp\left(-\left(\sum_{i=1}^{m} k_i \sum_{j=1}^{p} w_{ij}\psi_j - \sum_{b=1}^{n} \mathrm{e}^{-s\epsilon\ell^{(b)}} \sum_{i\in s_b} k_i \sum_{j=1}^{p} w_{ij}\psi_j\right)\right) \\ &= \exp\left(-\sum_{j=1}^{p} \psi_j \left(\sum_{i=1}^{m} k_i w_{ij} - \sum_{b=1}^{n} \mathrm{e}^{-s\epsilon\ell^{(b)}} \sum_{i\in s_b} k_i w_{ij}\right)\right) \\ &= \prod_{j=1}^{p} \exp\left(-\psi_j \sum_{i=1}^{m} k_i w_{ij}\left(1 - \sum_{b=1}^{n} \mathrm{e}^{-s\epsilon\ell^{(b)}} q_{jb}\right)\right).\end{aligned}$$

Writing g_{Ψ_j} for the density of the factor Ψ_j and using the independence of the factors, it follows that

$$\hat{F}_{\tilde{L}}(s) = \prod_{j=1}^{p} \int_0^\infty \exp\left(-\psi_j \sum_{i=1}^{m} k_i w_{ij}\left(1 - \sum_{b=1}^{n} \mathrm{e}^{-s\epsilon\ell^{(b)}} q_{jb}\right)\right) g_{\Psi_j}(\psi_j)\,\mathrm{d}\psi_j,$$

and equation (11.28) is derived by evaluating the integrals and recalling that the parameters of the gamma distribution are chosen to be $\alpha_j = \beta_j = \sigma_j^{-2}$.

(ii) To see that this is the Laplace–Stieltjes transform of the df of a sum of independent compound negative binomial distributions, observe that (11.28) may be written as

$$\hat{F}_{\tilde{L}}(s) = \prod_{j=1}^{p} \left(1 + \sigma_j^2 \sum_{i=1}^{m} k_i w_{ij}(1 - \hat{G}_{X_j}(s))\right)^{-\sigma_j^{-2}}, \tag{11.30}$$

where $\hat{G}_{X_j}$ is the Laplace–Stieltjes transform of G_{X_j}. Substituting $\theta_j = (1 + \sigma_j^2 \sum_{i=1}^{m} k_i w_{ij})^{-1}$ we obtain

$$\hat{F}_{\tilde{L}}(s) = \prod_{j=1}^{p} \left(\frac{\theta_j}{1 - \hat{G}_{X_j}(s)(1-\theta_j)}\right)^{\sigma_j^{-2}},$$

and, by comparing with Example 13.6, this can be seen to be the product of Laplace–Stieltjes transforms of the dfs of variables $Z_j \sim \mathrm{CNB}(\sigma_j^{-2}, \theta_j, G_{X_j})$. □

This theorem gives the key information required to evaluate the distribution of $\tilde{L}$ by applying Fourier inversion to the characteristic function of $\tilde{L}$. Indeed, we have

$$\phi_{\tilde{L}}(s) = \prod_{j=1}^{p} \left(\frac{\theta_j}{1 - \phi_{X_j}(s)(1-\theta_j)}\right)^{\sigma_j^{-2}},$$

where ϕ_{X_j} is the characteristic function of the multinomial severity distribution. It is now straightforward to compute ϕ_{X_j} using the fast Fourier transform and hence to compute the cf of $\tilde{L}$. The probability mass function of $\tilde{L}$ can then be computed by inverting $\phi_{\tilde{L}}$ using the inverse fast Fourier transform.

Notes and Comments

The logit-normal mixture model can be thought of as a one-factor version of the CreditPortfolioView model of Wilson (1997a,b). Details of this model can be found in Section 5 of Crouhy, Galai and Mark (2000). Further details of the beta-binomial distribution can be found in Joe (1997).

The rating agency Moody's uses a so-called binomial expansion technique to model default dependence in a simplistic way. The method, which is very popular with practitioners, is not based on a formal default risk model but is related to binomial distributions. The basic idea is to approximate a portfolio of m dependent counterparties by a homogeneous portfolio of $d < m$ independent counterparties with adjusted exposures and identical default probabilities; the index d is called the *diversity score* and is chosen according to rules defined by Moody's. For further information we refer to Davis and Lo (2001) and to Section 9.2.7 of Lando (2004).

The equivalence between threshold models and mixture models has been observed by Koyluoglu and Hickman (1998) and Gordy (2000) for the special case of CreditMetrics and CreditRisk$^+$. Applications of Proposition 7.51 to credit risk modelling are also discussed in Schönbucher (2005). The study of mixture representations for sequences of exchangeable Bernoulli rvs is related to a well-known result of de Finetti, which states that any *infinite* sequence $Y_1, Y_2, \ldots$ of exchangeable Bernoulli rvs has a representation as an exchangeable Bernoulli mixture; see, for instance, Theorem 35.10 in Billingsley (1995) for a precise statement. Any exchangeable model for $\boldsymbol{Y}$ that can be extended to arbitrary portfolio size m therefore has a representation as an exchangeable Bernoulli mixture model

A comprehensive description of CreditRisk$^+$ is given in its original documentation (Credit Suisse Financial Products 1997). An excellent discussion of the model structure from a more academic viewpoint is provided in Gordy (2000). Both sources also provide further information concerning the calibration of the factor variances σ_i and factor weights w_{ij}. The derivation of recursion formulas for the probabilities $P(\tilde{M} = k)$, $k = 0, 1, \ldots$, via Panjer recursion is given in Appendix A10 of the CreditRisk$^+$ documentation. In Gordy (2002) an alternative approach to the computation of the loss distribution in CreditRisk$^+$ is proposed using the saddlepoint approximation (see, for example, Jensen 1995). Further numerical work for CreditRisk$^+$ can be found in papers by Kurth and Tasche (2003), Glasserman (2004) and Haaf, Reiss and Schoenmakers (2004). Importance-sampling techniques for CreditRisk$^+$ are discussed in Glasserman and Li (2005).

11.3 Asymptotics for Large Portfolios

We now provide some asymptotic results for large portfolios in Bernoulli mixture models. These results can be used to approximate the credit loss distribution and associated risk measures in a large portfolio. Moreover, they are useful for identifying the key sources of model risk in a Bernoulli mixture model. In particular, we will see that in one-factor models the tail of the loss distribution is essentially determined by the tail of the mixing distribution, which has direct consequences for

the analysis of model risk in mixture models and for the setting of capital adequacy rules for loan books.

11.3.1 Exchangeable Models

We begin our discussion of the asymptotic properties of Bernoulli mixture models with the special case of an exchangeable model. We consider an infinite sequence of obligors indexed by $i \in \mathbb{N}$ with identical exposures $e_i = e$ and LGD equal to 100%. We assume that, given a mixing variable $Q \in [0, 1]$, the default indicators Y_i are independent Bernoulli random variables with conditional default probability $P(Y_i = 1 \mid Q = q) = q$. This simple model can be viewed an as idealization of a large pool of homogeneous obligors.

We are interested in the asymptotic behaviour of the relative loss (the loss expressed as a proportion of total exposure). Writing $L^{(m)} = \sum_{i=1}^{m} eY_i$ for the total loss of the first m companies, the corresponding relative loss is given by

$$\frac{L^{(m)}}{me} = \frac{1}{m} \sum_{i=1}^{m} Y_i.$$

Conditioning on $Q = q$, the Y_i are independent with mean q and the strong law of large numbers implies that, given $Q = q$, $\lim_{m\to\infty} L^{(m)}/(me) = q$ almost surely. This shows that, for large m, the behaviour of the relative loss is essentially governed by the mixing distribution $G(q)$ of Q. In particular, it can be shown that, for G strictly increasing,

$$\lim_{m\to\infty} \mathrm{VaR}_\alpha \left(\frac{L^{(m)}}{me} \right) = q_\alpha(Q) \tag{11.31}$$

(see Proposition 11.18 below).

These results can be used to analyse model risk in exchangeable Bernoulli mixture models. We consider the risk related to the choice of mixing distribution under the constraint that the default probability π and the default correlation ρ_Y (or, equivalently, π and π_2) are known and fixed.

According to (11.31), for large m the tail of $L^{(m)}$ is essentially determined by the tail of the mixing variable Q. In Figure 11.2 we plot the tail function of the probit-normal distribution (corresponding to the Gaussian threshold model), the logit-normal distribution, the beta distribution (close to CreditRisk$^+$; see Remark 11.15) and the mixture distribution corresponding to the Clayton copula (see Example 11.13). The plots are shown on a logarithmic scale and in all cases the first two moments have the values $\pi = 0.049$ and $\pi_2 = 0.003\,13$, which correspond roughly to Standard & Poor's rating category B; the parameter values for each of the models can be found in Table 11.3.

Inspection of Figure 11.2 shows that the tail functions differ significantly only after the 99% quantile, the logit-normal distribution being the one with the heaviest tail. From a practical point of view this means that the particular parametric form of the mixing distribution in a Bernoulli mixture model is of lesser importance once π and ρ_Y have been fixed. Of course this does not mean that Bernoulli mixtures are

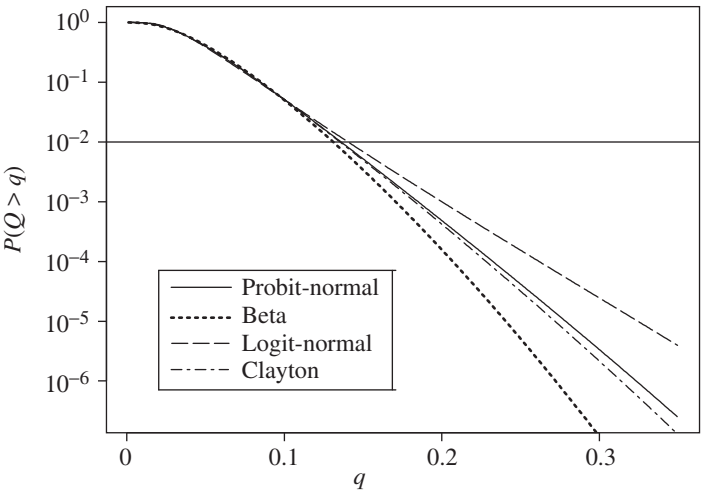

Figure 11.2. The tail of the mixing distribution of Q in four different exchangeable Bernoulli-mixture models: beta, probit-normal, logit-normal and Clayton. In all cases the first two moments have the values $\pi = 0.049$ and $\pi_2 = 0.003\,13$, which correspond roughly to Standard & Poor's rating category B; the actual parameter values can be found in Table 11.3. The horizontal line at 10^{-2} shows that the models only really start to differ around the 99th percentile of the mixing distribution.

Table 11.3. Parameter values for various exchangeable Bernoulli mixture models with identical values of π and π_2 (and ρ_Y). The values of π and π_2 correspond roughly to Standard & Poor's ratings CCC, B and BB (in fact, they have been estimated from 20 years of Standard & Poor's default data using the simple moment estimator in (11.50)). This table is used in the model-risk study of Section 11.1.5. and the simulation study of Section 11.5.2.

Model	Parameter	CCC	B	BB
All models	π	0.188	0.049	0.011 2
	π_2	0.042	0.003 13	0.000 197
	ρ_Y	0.044 6	0.015 7	0.006 43
Beta	a	4.02	3.08	1.73
	b	17.4	59.8	153
Probit-normal	μ	−0.93	−1.71	−2.37
	σ	0.316	0.264	0.272
Logit-normal	μ	−1.56	−3.1	−4.71
	σ	0.553	0.556	0.691
Clayton	π	0.188	0.049	0.011 2
	θ	0.070 4	0.032	0.024 7

immune to model risk; the tail of $L^{(m)}$ is quite sensitive to π and, in particular, to ρ_Y, and these parameters are not easily estimated (see Section 11.5.4 for a discussion of statistical inference for mixture models).

11.3.2 General Results

Since we are interested in asymptotic properties of the overall loss distribution, we also consider exposures and losses given default. Let $(e_i)_{i\in\mathbb{N}}$ be an infinite sequence of positive deterministic exposures, let $(Y_i)_{i\in\mathbb{N}}$ be the corresponding sequence of default indicators, and let $(\delta_i)_{i\in\mathbb{N}}$ be a sequence of rvs with values in $(0, 1]$ representing percentage losses given that default occurs. In this setting the loss for a portfolio of size m is given by $L^{(m)} = \sum_{i=1}^m L_i$, where $L_i = e_i\delta_i Y_i$ are the individual losses. We are interested in results for the relative loss $L^{(m)}/\sum_{i=1}^m e_i$, which expresses the loss as a proportion of total exposure. We introduce the notation $a_m = \sum_{i=1}^m e_i$ for the aggregate exposure to the first m obligors.

We now make some technical assumptions for our model.

(A1) There is a p-dimensional random vector $\boldsymbol{\Psi}$ such that, conditional on $\boldsymbol{\Psi}$, the $(L_i)_{i\in\mathbb{N}}$ form a sequence of independent rvs.

In this assumption the conditional independence structure is extended from the default indicators to the losses. Note that (A1) allows for the situation where δ_i depends on the systematic factors $\boldsymbol{\Psi}$. This extension is relevant from an empirical viewpoint since evidence suggests that losses given default tend to depend on the state of the underlying economy (see Section 11.2.3).

(A2) There is a function $\bar{\ell}\colon \mathbb{R}^p \to [0, 1]$ such that

$$\lim_{m\to\infty} \frac{1}{a_m} E(L^{(m)} \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) = \bar{\ell}(\boldsymbol{\psi}) \tag{11.32}$$

for all $\boldsymbol{\psi} \in \mathbb{R}^p$. We call $\bar{\ell}(\boldsymbol{\psi})$ the asymptotic relative loss function.

Assumption (A2) implies that we preserve the essential composition of the portfolio as we allow it to grow (see, for instance, Example 11.20).

(A3) The sequence of exposures satisfies $\lim_{m\to\infty} a_m = \infty$ and $\sum_{i=1}^\infty (e_i/a_i)^2 < \infty$.

This is a very weak technical assumption that would be satisfied by any realistic sequence of exposures. For example, if $e_i = e$ for all i, then $\sum_{i=1}^\infty (e_i/a_i)^2 = \sum_{i=1}^\infty i^{-2} < \infty$. Even in the case where $e_i = i$, so that exposures grow linearly with portfolio size, we have $\sum_{i=1}^\infty (e_i/a_i)^2 = \sum_{i=1}^\infty (2/(i+1))^2 < \infty$, since $a_i = \sum_{k=1}^i k = i(i+1)/2$. To find a counterexample we would need to take a sequence of exposures where the cumulative sum grows at the same rate as the maximum of the first m exposures. In intuitive terms this means that the portfolio is dominated by a few large exposures (name concentration).

The following result shows that under these assumptions the average portfolio loss is essentially determined by the asymptotic relative loss function $\bar{\ell}$ and by the realization of the factor random vector $\boldsymbol{\Psi}$.

Proposition 11.17. *Consider a sequence $L^{(m)} = \sum_{i=1}^m L_i$ satisfying Assumptions (A1)–(A3) above. Denote by $P(\cdot \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$ the conditional distribution of*

the sequence $(L_i)_{i\in\mathbb{N}}$ *given* $\boldsymbol{\Psi} = \boldsymbol{\psi}$. *Then*

$$\lim_{m\to\infty} \frac{1}{a_m} L^{(m)} = \bar{\ell}(\boldsymbol{\psi}), \quad P(\cdot \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) \text{ a.s.}$$

Proof. The proof is based on the following version of the law of large numbers for non-identically distributed random variables given by Petrov (1995, Theorem 6.7). If $(Z_i)_{i\in\mathbb{N}}$ is a sequence of independent random variables and $(a_i)_{i\in\mathbb{N}}$ is a sequence of positive constants satisfying $\lim_{m\to\infty} a_m = \infty$ and $\sum_{i=1}^{\infty}(\operatorname{var}(Z_i)/a_i^2) < \infty$, then, as $m \to \infty$,

$$\frac{1}{a_m}\left(\sum_{i=1}^{m} Z_i - E\left(\sum_{i=1}^{m} Z_i\right)\right) \to 0 \text{ a.s.}$$

We set $Z_i = L_i = e_i\delta_i Y_i$ and $a_i = \sum_{k=1}^{i} e_k$ as before. We apply this result conditional on $\boldsymbol{\Psi} = \boldsymbol{\psi}$; that is, we work under the measure $P(\cdot \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$. Under this measure the L_i are independent by Assumption (A1). Note that $\operatorname{var}(L_i) = e_i^2 \operatorname{var}(\delta_i Y_i) \leqslant e_i^2$, since $\delta_i Y_i$ is an rv on [0, 1]. Using Assumption (A3) we verify that

$$\sum_{i=1}^{\infty} \frac{\operatorname{var}(Z_i)}{a_i^2} \leqslant \sum_{i=1}^{\infty} \left(\frac{e_i}{a_i}\right)^2 < \infty.$$

Applying Petrov's result and Assumption (A2) we get

$$\lim_{m\to\infty} \frac{1}{a_m} L^{(m)} - \bar{\ell}(\boldsymbol{\psi}) = \lim_{m\to\infty} \left(\frac{1}{a_m}\sum_{i=1}^{m} L_i - \frac{1}{a_m} E\left(\sum_{i=1}^{m} L_i \mid \boldsymbol{\Psi} = \boldsymbol{\psi}\right)\right) = 0.$$

□

For one-factor Bernoulli mixture models a stronger result can be obtained that links the quantiles of the relative portfolio loss $L^{(m)}/a_m$ to quantiles of the mixing distribution.

Proposition 11.18. *Consider a sequence* $L^{(m)} = \sum_{i=1}^{m} L_i$ *satisfying Assumptions (A1)–(A3) with a one-dimensional mixing variable* Ψ *with df* G. *Assume that the conditional asymptotic loss function* $\bar{\ell}(\psi)$ *is strictly increasing and continuous and that* G *is strictly increasing at* $q_\alpha(\Psi)$, *i.e. that* $G(q_\alpha(\Psi)+\delta) > \alpha$ *for every* $\delta > 0$. *Then*

$$\lim_{m\to\infty} \operatorname{VaR}_\alpha\left(\frac{1}{a_m} L^{(m)}\right) = \bar{\ell}(q_\alpha(\Psi)). \tag{11.33}$$

The assumption that $\bar{\ell}$ is strictly increasing makes sense if it is assumed that low values of Ψ correspond to good states of the world with lower conditional default probabilities and lower losses given default than average, while high values of Ψ correspond to bad states with correspondingly higher losses given default.

Proof. The proof is based on the following simple intuition. Since $L^{(m)}/a_m$ converges to $\bar{\ell}(\Psi)$ and since $\bar{\ell}$ is strictly increasing by assumption, we have for large m

$$q_\alpha\left(\frac{L^{(m)}}{a_m}\right) \approx q_\alpha(\bar{\ell}(\Psi)) = \bar{\ell}(q_\alpha(\Psi)).$$

To turn this into a formal argument we use the following continuity result for quantiles, a proof of which may be found in Fristedt and Gray (1997, Proposition 5, p. 250) or Resnick (2008, Proposition 0.1).

Lemma 11.19. *Consider a sequence of random variables $(Z_m)_{m\in\mathbb{N}}$ that converges in distribution to a random variable Z. Then $\lim_{m\to\infty} q_u(Z_m) = q_u(Z)$ at all points of continuity u of the quantile function $u \mapsto q_u(Z)$.*

In our case, $L^{(m)}/a_m$ converges to $\bar{\ell}(\Psi)$ almost surely, and hence in distribution. The assumption that the function $\bar{\ell}$ is strictly increasing and that G is strictly increasing at $q_\alpha(\Psi)$ ensures that the distribution function of $\bar{\ell}(\Psi)$ is strictly increasing at that point so that the quantile function of the rv $\bar{\ell}(\Psi)$ is continuous at α. This shows that

$$\lim_{m\to\infty} q_\alpha\left(\frac{1}{a_m}L^{(m)}\right) = q_\alpha(\bar{\ell}(\Psi)).$$

Finally, the equality $q_\alpha(\bar{\ell}(\Psi)) = \bar{\ell}(q_\alpha(\Psi))$ follows from Proposition A.3. □

Example 11.20. Consider the one-factor Bernoulli mixture model for k exchangeable groups defined by (11.18). Denote by $r(i)$ the group of obligor i and assume that, within each group r, the exposures, LGDs and conditional default probabilities are identical and are given by e_r, δ_r and $p_r(\psi)$, respectively.

Suppose that we allow the portfolio to grow and that we write $m_r^{(m)}$ for the number of obligors in group r when the portfolio size is m. The relative exposure to group r is given by $\lambda_r^{(m)} = e_r m_r^{(m)} / \sum_{r=1}^k e_r m_r^{(m)}$, and we assume that $\lambda_r^{(m)} \to \lambda_r$ as $m \to \infty$. In this case the asymptotic relative loss function in equation (11.32) is

$$\begin{aligned}
\bar{\ell}(\psi) &= \lim_{m\to\infty} \sum_{i=1}^m \frac{e_{r(i)}}{\sum_{i=1}^m e_{r(i)}} \delta_{r(i)} p_{r(i)}(\psi) \\
&= \lim_{m\to\infty} \sum_{r=1}^k \frac{e_r m_r^{(m)}}{\sum_{r=1}^k e_r m_r^{(m)}} \delta_r h(\mu_r + \sigma\psi) \\
&= \sum_{r=1}^k \lambda_r \delta_r h(\mu_r + \sigma\psi).
\end{aligned}$$

Since Ψ is assumed to have a standard normal distribution, (11.33) implies that

$$\lim_{m\to\infty} q_\alpha\left(\frac{L^{(m)}}{\sum_{i=1}^m e_i}\right) = \sum_{r=1}^k \lambda_r \delta_r h(\mu_r + \sigma\Phi^{-1}(\alpha)). \tag{11.34}$$

For large m, since $\lambda_r \sum_{1=1}^m e_i \approx m_r^{(m)} e_r$, we get that

$$\mathrm{VaR}_\alpha(L^{(m)}) \approx \sum_{r=1}^k m_r^{(m)} e_r \delta_r h(\mu_r + \sigma_r\Phi^{-1}(\alpha)). \tag{11.35}$$

11.3.3 The Basel IRB Formula

In this section we examine how the considerations of Sections 11.3.1 and 11.3.2 have influenced the Basel capital adequacy framework, which was discussed in more general terms in Section 1.3. Under this framework a bank is required to hold 8% of the so-called *risk-weighted assets* (RWA) of its credit portfolio as risk capital. The RWA of a portfolio is given by the sum of the RWA of the individual risks in the portfolio, i.e. $\text{RWA}^{\text{portfolio}} = \sum_{i=1}^{m} \text{RWA}_i$. The quantity RWA_i reflects the exposure size and the riskiness of obligor i and takes the form $\text{RWA}_i = w_i e_i$, where w_i is a risk weight and e_i denotes exposure size.

Banks may choose between two options for determining the risk weight w_i, which must then be implemented for the entire portfolio. Under the simpler *standardized approach*, the risk weight w_i is determined by the type (sovereign, bank or corporate) and the credit rating of counterparty i. For instance, $w_i = 50\%$ for a corporation with a Moody's rating in the range A+ to A−. Under the more advanced *internal-ratings-based* (IRB) approach, the risk weight takes the form

$$w_i = (0.08)^{-1} c \delta_i \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\beta_i}\Phi^{-1}(0.999)}{\sqrt{1-\beta_i}} \right). \tag{11.36}$$

Here, c is a technical adjustment factor that is of minor interest to us, p_i represents the default probability, and δ_i is the percentage loss given default of obligor i. The parameter $\beta_i \in (0.12, 0.24)$ measures the systematic risk of obligor i. Estimates for p_i and (under the so-called advanced IRB approach) for δ_i and e_i are provided by the individual bank; the adjustment factor c and, most importantly, the value of β_i are determined by fixed rules within the Basel II Accord independently of the structure of the specific portfolio under consideration. The risk capital to be held for counterparty i is thus given by

$$\text{RC}_i = 0.08\text{RWA}_i = c \delta_i e_i \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\beta_i}\Phi^{-1}(0.999)}{\sqrt{1-\beta_i}} \right). \tag{11.37}$$

The interesting part of equation (11.37) is, of course, the expression involving the standard normal df, and we now give a derivation.

Consider a one-factor Gaussian threshold with default probabilities $p_1, \ldots, p_m$ and critical variables given by

$$X_i = \sqrt{\beta_i} F + \sqrt{1-\beta_i}\varepsilon_i \tag{11.38}$$

for iid standard normal rvs $F, \varepsilon_1, \ldots, \varepsilon_m$. By taking $\Psi = -F$, the equivalent Bernoulli mixture model was shown in Section 11.2.4 to have conditional default probabilities $p_i(\psi) = \Phi((\Phi^{-1}(p_i) + \sqrt{\beta_i}\psi)/\sqrt{1-\beta_i})$. Note that this is of the form $p_i(\psi) = h(\mu_i + \sigma_i \psi)$ for $h = \Phi$, $\mu_i = \Phi^{-1}(p_i)/\sqrt{1-\beta_i}$ and $\sigma_i = \sqrt{\beta_i/(1-\beta_i)}$. Assume, moreover, that the portfolio has a homogeneous group structure consisting of a few large groups with (approximately) identical exposures, PDs, LGDs and factor weights within the groups, as in Example 11.20.

Applying the analysis of that example and, in particular, equation (11.35), it follows that

$$\begin{aligned}\mathrm{VaR}_\alpha(L) &\approx \sum_{i=1}^{m} \delta_i e_i \, p_i(q_\alpha(\Psi)) \\ &= \sum_{i=1}^{m} \delta_i e_i \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\beta_i}\Phi^{-1}(\alpha)}{\sqrt{1-\beta_i}}\right).\end{aligned}$$

For $c = 1$ the risk capital RC_i in (11.37) can thus be considered as the asymptotic contribution of risk i to the 99.9% VaR of the overall portfolio in a one-factor Gaussian threshold model with homogeneous group structure. Note, further, that β_i can be viewed as the asset correlation for firms i within the same group.

While formula (11.36) is influenced by portfolio-theoretic considerations, the Basel framework falls short of reflecting the true dependence structure of a bank's credit portfolio for a number of reasons. First, in the Basel framework the parameters β_i are specified ad hoc by regulatory rules irrespective of the composition of the portfolio at hand. Second, the homogeneous group structure and the simple one-factor model (11.38) are typically oversimplified representations of the factor structure underlying default dependence, particularly for internationally active banks. Third, the rule is based on an asymptotic result. Moreover, historical default experience for the portfolio under consideration has no formal role to play in setting capital adequacy standards. These deficiencies should be weighed against the relative simplicity of the IRB approach, which makes it suitable for use in a supervisory setting. For economic capital purposes, on the other hand, most banks develop fully internal models with more sophisticated factor models to describe dependencies.

Notes and Comments

The results in Section 11.3 are an amalgamation of results from Frey and McNeil (2003) and Gordy (2003). The first limit result for large portfolios was obtained in Vasicek (1997) for a probit-normal mixture model equivalent to the KMV model. Asymptotic results for credit portfolios related to the theory of large deviations are discussed in Dembo, Deuschel and Duffie (2004). For details of the IRB approach, and the Basel II Capital Accord in general, we refer to the website of the Basel Committee: www.bis.org/bcbs. Our discussion in Section 11.3.3 is related to the analysis by Gordy (2003).

There have been a number of papers on second-order corrections or so-called *granularity adjustments* to the large-portfolio results in Propositions 11.17 and 11.18. While the results assume that idiosyncratic risk diversifies away in sufficiently large portfolios, these corrections take into account the fact that a certain amount of idiosyncratic risk and name concentration will remain in real portfolios. References include Martin and Wilde (2002), Gordy (2004), Gordy and Marrone (2012), Gordy and Lütkebohmert (2013) and Gagliardini and Gouriéroux (2013). See also the book by Lütkebohmert (2009) on concentration risk in credit portfolios.

11.4 Monte Carlo Methods

In this section we consider a Bernoulli mixture model for a loan portfolio and assume that the overall loss is of the form $L = \sum_{i=1}^{m} L_i$, where the L_i are conditionally independent given some economic factor vector $\boldsymbol{\Psi}$. A possible method for calculating risk measures and related quantities such as capital allocations is to use Monte Carlo (MC) simulation, although the problem of *rare-event simulation* arises. Suppose, for example, that we wish to compute expected shortfall and expected shortfall contributions at the confidence level α for our portfolio. We need to evaluate the conditional expectations

$$E(L \mid L \geqslant q_\alpha(L)) \quad \text{and} \quad E(L_i \mid L \geqslant q_\alpha(L)). \tag{11.39}$$

If $\alpha = 0.99$, say, then only 1% of our standard Monte Carlo draws will lead to a portfolio loss higher than $q_{0.99}(L)$. The standard MC estimator of (11.39), which consists of averaging the simulated values of L or L_i over all draws, leading to a simulated portfolio loss $L \geqslant q_\alpha(L)$, will be unstable and subject to high variability, unless the number of simulations is very large. The problem is of course that most simulations are "wasted", in that they lead to a value of L that is smaller than $q_\alpha(L)$. Fortunately, there exists a variance-reduction technique known as *importance sampling* (IS), which is well suited to such problems.

11.4.1 Basics of Importance Sampling

Consider an rv X on some probability space $(\Omega, \mathcal{F}, P)$ and assume that it has an absolutely continuous df with density f. A generalization to general probability spaces is discussed below. The problem we consider is the computation of the expected value

$$\theta = E(h(X)) = \int_{-\infty}^{\infty} h(x) f(x)\, \mathrm{d}x \tag{11.40}$$

for some known function h. To calculate the probability of an event we consider a function of the form $h(x) = I_{\{x \in A\}}$ for some set $A \subset \mathbb{R}$; for expected shortfall computation we consider functions of the form $h(x) = x I_{\{x \geqslant c\}}$ for some $c \in \mathbb{R}$. Where the analytical evaluation of (11.40) is difficult, due to the complexity of the distribution of X, we can resort to an MC approach, for which we only have to be able to simulate variates from the distribution with density f.

Algorithm 11.21 (Monte Carlo integration).

(1) Generate $X_1, \ldots, X_n$ independently from density f.

(2) Compute the standard MC estimate $\hat{\theta}_n^{\mathrm{MC}} = (1/n) \sum_{i=1}^{n} h(X_i)$.

The MC estimator converges to θ by the strong law of large numbers, but the speed of convergence may not be particularly fast, particularly when we are dealing with rare-event simulation.

Importance sampling is based on an alternative representation of the integral in (11.40). Consider a second probability density g (whose support should contain that of f) and define the *likelihood ratio* $r(x)$ by $r(x) := f(x)/g(x)$ whenever

$g(x) > 0$, and $r(x) = 0$ otherwise. The integral (11.40) may be written in terms of the likelihood ratio as

$$\theta = \int_{-\infty}^{\infty} h(x)r(x)g(x)\,\mathrm{d}x = E_g(h(X)r(X)), \tag{11.41}$$

where E_g denotes expectation with respect to the density g. We can therefore approximate the integral with the following algorithm.

Algorithm 11.22 (importance sampling).

(1) Generate $X_1, \ldots, X_n$ independently from density g.

(2) Compute the IS estimate $\hat{\theta}_n^{\mathrm{IS}} = (1/n)\sum_{i=1}^n h(X_i)\, r(X_i)$.

The density g is often termed the *importance-sampling density*. The art (or science) of importance sampling is in choosing an importance-sampling density such that, for fixed n, the variance of the IS estimator is considerably smaller than that of the standard Monte Carlo estimator. In this way we can hope to obtain a prescribed accuracy in evaluating the integral of interest using far fewer random draws than are required in standard Monte Carlo simulation. The variances of the estimators are given by

$$\begin{aligned}\mathrm{var}_g(\hat{\theta}_n^{\mathrm{IS}}) &= (1/n)(E_g(h(X)^2 r(X)^2) - \theta^2),\\ \mathrm{var}(\hat{\theta}_n^{\mathrm{MC}}) &= (1/n)(E(h(X)^2) - \theta^2),\end{aligned}$$

so the aim is to make $E_g(h(X)^2 r(X)^2)$ small compared with $E(h(X)^2)$. In theory, the variance of $\hat{\theta}^{\mathrm{IS}}$ can be reduced to zero by choosing an optimal g. To see this, suppose for the moment that h is non-negative and set

$$g^*(x) = f(x)h(x)/E(h(X)). \tag{11.42}$$

With this choice, the likelihood ratio becomes $r(x) = E(h(X))/h(x)$. Hence $\hat{\theta}_1^{\mathrm{IS}} = h(X_1)r(X_1) = E(h(X))$, and the IS estimator gives the correct answer in a single draw. In practice, it is of course impossible to choose an IS density of the form (11.42), as this requires knowledge of the quantity $E(h(X))$ that one wants to compute; nonetheless, (11.42) can provide useful guidance in choosing an IS density, as we will see in the next section.

Consider the case of estimating a rare-event probability corresponding to $h(x) = I_{\{x \geqslant c\}}$ for c significantly larger than the mean of X. Then we have that $E(h(X)^2) = P(X \geqslant c)$ and, using (11.41), that

$$E_g(h(X)^2 r(X)^2) = E_g(r(X)^2; X \geqslant c) = E(r(X); X \geqslant c). \tag{11.43}$$

Clearly, we should try to choose g such that the likelihood ratio $r(x) = f(x)/g(x)$ is small for $x \geqslant c$; in other words, we should make the event $\{X \geqslant c\}$ more likely under the IS density g than it is under the original density f.

Exponential tilting. We now describe a useful way of finding IS densities when X is light tailed. For $t \in \mathbb{R}$ we write $M_X(t) = E(\mathrm{e}^{tX}) = \int_{-\infty}^{\infty} \mathrm{e}^{tx} f(x)\,\mathrm{d}x$ for the moment-generating function of X, which we assume is finite for $t \in \mathbb{R}$. If $M_X(t)$ is finite, we can define an IS density by $g_t(x) := \mathrm{e}^{tx} f(x)/M_X(t)$. The likelihood ratio is $r_t(x) = f(x)/g_t(x) = M_X(t)\mathrm{e}^{-tx}$. Define μ_t to be the mean of X with respect to the density g_t, i.e.

$$\mu_t := E_{g_t}(X) = E(X\mathrm{e}^{tX})/M_X(t). \tag{11.44}$$

How can we choose t optimally for a particular IS problem? We consider the case of tail probability estimation and recall from (11.43) that the objective is to make

$$E(r(X); X \geqslant c) = E(I_{\{X \geqslant c\}} M_X(t)\mathrm{e}^{-tX}) \tag{11.45}$$

small. Now observe that $\mathrm{e}^{-tx} \leqslant \mathrm{e}^{-tc}$ for $x \geqslant c$ and $t \geqslant 0$, so

$$E(I_{\{X \geqslant c\}} M_X(t)\mathrm{e}^{-tX}) \leqslant M_X(t)\mathrm{e}^{-tc}.$$

Instead of solving the (difficult) problem of minimizing (11.45) over t, we choose t such that this bound becomes minimal. Equivalently, we try to find t minimizing $\ln M_X(t) - tc$. Using (11.44) we obtain that

$$\frac{\mathrm{d}}{\mathrm{d}t} \ln M_X(t) - tc = \frac{E(X\mathrm{e}^{tX})}{M_X(t)} - c = \mu_t - c,$$

which suggests choosing $t = t(c)$ as the solution of the equation $\mu_t = c$, so that the rare event $\{X \geqslant c\}$ becomes a normal event if we compute probabilities using the density $g_{t(c)}$. A unique solution of the equation $\mu_t = c$ exists for all relevant values of c. In the cases that are of interest to us this is immediately obvious from the form of the exponentially tilted distributions, so we omit a formal proof.

Example 11.23 (exponential tilting for normal distribution). We illustrate the concept of exponential tilting in the simple case of a standard normal rv. Suppose that $X \sim N(0, 1)$ with density $\phi(x)$. Using exponential tilting we obtain the new density $g_t(x) = \mathrm{e}^{tx}\phi(x)/M_X(t)$. The moment-generating function of X is known to be $M_X(t) = \mathrm{e}^{t^2/2}$. Hence

$$g_t(x) = \frac{1}{\sqrt{2\pi}} \exp(tx - \tfrac{1}{2}(t^2 + x^2)) = \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}(x - t)^2),$$

so that, under the tilted distribution, $X \sim N(t, 1)$. Note that in this case exponential tilting corresponds to changing the mean of X.

An abstract view of importance sampling. To handle the more complex application to portfolio credit risk in the next section it helps to consider importance sampling from a slightly more general viewpoint. Given densities f and g as above, define probability measures P and Q by

$$P(A) = \int_A f(x)\,\mathrm{d}x \quad \text{and} \quad Q(A) = \int_A g(x)\,\mathrm{d}x, \quad A \subset \mathbb{R}.$$

With this notation, (11.41) becomes $\theta = E^P(h(X)) = E^Q(h(X)r(X))$, so that $r(X)$ equals $\mathrm{d}P/\mathrm{d}Q$, the (measure-theoretic) density of P with respect to Q. Using this more abstract view, exponential tilting can be applied in more general situations: given an rv X on $(\Omega, \mathcal{F}, P)$ such that $M_X(t) = E^P(\mathrm{e}^{tX}) < \infty$, define the measure Q_t on $(\Omega, \mathcal{F})$ by

$$\frac{\mathrm{d}Q_t}{\mathrm{d}P} = \frac{\mathrm{e}^{tX}}{M_X(t)}, \quad \text{i.e. } Q_t(A) = E^P\left(\frac{\mathrm{e}^{tX}}{M_X(t)}; A\right),$$

and note that $(\mathrm{d}Q_t/\mathrm{d}P)^{-1} = M_X(t)\mathrm{e}^{-tX} = r_t(X)$. The IS algorithm remains essentially unchanged: simulate independent realizations X_i under the measure Q_t and set $\hat{\theta}^{\mathrm{IS}} = (1/n)\sum_{i=1}^n X_i r_t(X_i)$ as before.

11.4.2 Application to Bernoulli Mixture Models

In this section we return to the subject of credit losses and consider a portfolio loss of the form $L = \sum_{i=1}^m e_i Y_i$, where the e_i are deterministic, positive exposures and the Y_i are default indicators with default probabilities p_i. We assume that $\boldsymbol{Y}$ follows a Bernoulli mixture model in the sense of Definition 11.5 with factor vector $\boldsymbol{\Psi}$ and conditional default probabilities $p_i(\boldsymbol{\Psi})$. We study the problem of estimating exceedance probabilities $\theta = P(L \geqslant c)$ for c substantially larger than $E(L)$ using importance sampling. This is useful for risk-management purposes, as, for $c \approx q_\alpha(L)$, a good IS distribution for the computation of $P(L \geqslant c)$ also yields a substantial variance reduction for computing expected shortfall or expected shortfall contributions.

We consider first the situation where the default indicators $Y_1, \ldots, Y_m$ are independent, and then we discuss the extension to the case of conditionally independent default indicators. Our exposition is based on Glasserman and Li (2005).

Independent default indicators. Here we use the more general IS approach outlined at the end of the previous section. Set $\Omega = \{0, 1\}^m$, the state space of $\boldsymbol{Y}$. The probability measure P is given by

$$P(\{\boldsymbol{y}\}) = \prod_{i=1}^m p_i^{y_i}(1 - p_i)^{1-y_i}, \quad \boldsymbol{y} \in \{0, 1\}^m.$$

We need to understand how this measure changes under exponential tilting using L. The moment-generating function of L is easily calculated to be

$$M_L(t) = E\left(\exp\left(t\sum_{i=1}^m e_i Y_i\right)\right) = \prod_{i=1}^m E(\mathrm{e}^{te_i Y_i}) = \prod_{i=1}^m (\mathrm{e}^{te_i} p_i + 1 - p_i).$$

The measure Q_t is given by $Q_t(\{\boldsymbol{y}\}) = E^P(\mathrm{e}^{tL}/M_L(t); \boldsymbol{Y} = \boldsymbol{y})$ and hence

$$Q_t(\{\boldsymbol{y}\}) = \frac{\exp(t\sum_{i=1}^m e_i y_i)}{M_L(t)} P(\{\boldsymbol{y}\}) = \prod_{i=1}^m \frac{\mathrm{e}^{te_i y_i}}{\mathrm{e}^{te_i} p_i + 1 - p_i} p_i^{y_i}(1 - p_i)^{1-y_i}.$$

Define new default probabilities by $\bar{q}_{t,i} := \mathrm{e}^{te_i} p_i/(\mathrm{e}^{te_i} p_i + 1 - p_i)$. It follows that $Q_t(\{\boldsymbol{y}\}) = \prod_{i=1}^m \bar{q}_{t,i}^{y_i}(1 - \bar{q}_{t,i})^{1-y_i}$, so that after exponential tilting the default indicators remain independent but with new default probability $\bar{q}_{t,i}$. Note that $\bar{q}_{t,i}$ tends

to 1 for $t \to \infty$ and to 0 for $t \to -\infty$, so that we can shift the mean of L to any point in $(0, \sum_{i=1}^m e_i)$.

In analogy with our previous discussion, for IS purposes, the optimal value of t is chosen such that $E^{Q_t}(L) = c$, leading to the equation $\sum_{i=1}^m e_i \bar{q}_{t,i} = c$.

Conditionally independent default indicators. The first step in the extension of the importance-sampling approach to conditionally independent defaults is obvious: given a realization $\boldsymbol{\psi}$ of the economic factors, the conditional exceedance probability $\theta(\boldsymbol{\psi}) := P(L \geqslant c \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$ is estimated using the approach for independent default indicators described above. We have the following algorithm.

Algorithm 11.24 (IS for conditional loss distribution).

(1) Given $\boldsymbol{\psi}$, calculate the conditional default probabilities $p_i(\boldsymbol{\psi})$ according to the particular model, and solve the equation

$$\sum_{i=1}^m e_i \frac{\mathrm{e}^{te_i} p_i(\boldsymbol{\psi})}{\mathrm{e}^{te_i} p_i(\boldsymbol{\psi}) + 1 - p_i(\boldsymbol{\psi})} = c;$$

the solution $t = t(c, \boldsymbol{\psi})$ gives the optimal degree of tilting.

(2) Generate n_1 conditional realizations of the default vector $\boldsymbol{Y} = (Y_1, \ldots, Y_m)'$. The defaults of the companies are simulated independently, with the default probability of the ith company given by

$$\frac{\exp(t(c, \boldsymbol{\psi})e_i)\, p_i(\boldsymbol{\psi})}{\exp(t(c, \boldsymbol{\psi})e_i)\, p_i(\boldsymbol{\psi}) + 1 - p_i(\boldsymbol{\psi})}.$$

(3) Denote by $M_L(t, \boldsymbol{\psi}) := \prod_{i=1}^m \{\mathrm{e}^{te_i} p_i(\boldsymbol{\psi}) + 1 - p_i(\boldsymbol{\psi})\}$ the conditional moment-generating function of L. From the simulated default data construct n_1 conditional realizations of $L = \sum_{i=1}^m e_i Y_i$ and label these $L^{(1)}, \ldots, L^{(n_1)}$. Determine the IS estimator for the conditional loss distribution:

$$\hat{\theta}_{n_1}^{\mathrm{IS},1}(\boldsymbol{\psi}) = M_L(t(c, \boldsymbol{\psi}), \boldsymbol{\psi}) \frac{1}{n_1} \sum_{j=1}^{n_1} I_{\{L^{(j)} \geqslant c\}} \exp(-t(c, \boldsymbol{\psi}) L^{(j)}).$$

In principle, the approach discussed above also applies in the more general situation where the loss given default is random; all we need to assume is that the L_i are conditionally independent given $\boldsymbol{\Psi}$, as in Assumption (A1) of Section 11.3. However, the actual implementation can become quite involved.

IS for the distribution of the factor variables. Suppose we now want to estimate the unconditional probability $\theta = P(L \geqslant c)$. A naive approach would be to generate realizations of the factor vector $\boldsymbol{\Psi}$ and to estimate θ by averaging the IS estimator of Algorithm 11.24 over these realizations. As is shown in Glasserman and Li (2005), this is not the best solution for large portfolios of dependent credit risks. Intuitively, this is due to the fact that for such portfolios most of the variation in L is caused by fluctuations of the economic factors, and we have not yet applied IS to the distribution

of $\boldsymbol{\Psi}$. For this reason we now discuss a full IS algorithm that combines IS for the economic factor variables with Algorithm 11.24.

We consider the important case of a Bernoulli mixture model with multivariate Gaussian factors and conditional default probabilities $p_i(\boldsymbol{\Psi})$ for $\boldsymbol{\Psi} \sim N_p(\mathbf{0}, \Omega)$, such as the probit-normal Bernoulli mixture model described in Example 11.11. In this context it is natural to choose an importance-sampling density such that $\boldsymbol{\Psi} \sim N_p(\boldsymbol{\mu}, \Omega)$ for a new mean vector $\boldsymbol{\mu} \in \mathbb{R}^p$, i.e. we take g as the density of $N_p(\boldsymbol{\mu}, \Omega)$. For a good choice of $\boldsymbol{\mu}$ we expect to generate realizations of $\boldsymbol{\Psi}$ leading to high conditional default probabilities more frequently. The corresponding likelihood ratio $r_{\boldsymbol{\mu}}(\boldsymbol{\Psi})$ is given by the ratio of the respective multivariate normal densities, so that

$$r_{\boldsymbol{\mu}}(\boldsymbol{\Psi}) = \frac{\exp(-\frac{1}{2}\boldsymbol{\Psi}'\Omega^{-1}\boldsymbol{\Psi})}{\exp(-\frac{1}{2}(\boldsymbol{\Psi}-\boldsymbol{\mu})'\Omega^{-1}(\boldsymbol{\Psi}-\boldsymbol{\mu}))} = \exp(-\boldsymbol{\mu}'\Omega^{-1}\boldsymbol{\Psi} + \tfrac{1}{2}\boldsymbol{\mu}'\Omega^{-1}\boldsymbol{\mu}).$$

Essentially, this is a multivariate analogue of the exponential tilting applied to a univariate normal distribution in Example 11.23.

Now we can describe the algorithm for full IS. At the outset we have to choose the overall number of simulation rounds, n, the number of repetitions of conditional IS per simulation round, n_1, and the mean of the IS distribution for the factors, $\boldsymbol{\mu}$. Whereas the value of n depends on the desired degree of precision and is best determined in a simulation study, n_1 should be taken to be fairly small. An approach to determine a sensible value of $\boldsymbol{\mu}$ is discussed below.

Algorithm 11.25 (full IS for mixture models with Gaussian factors).

(1) Generate $\boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_n \sim N(\boldsymbol{\mu}, I_p)$.

(2) For each $\boldsymbol{\Psi}_i$ calculate $\hat{\theta}_{n_1}^{\mathrm{IS},1}(\boldsymbol{\Psi}_i)$ as in Algorithm 11.24.

(3) Determine the full IS estimator:

$$\hat{\theta}_n^{\mathrm{IS}} = \frac{1}{n}\sum_{i=1}^{n} r_{\boldsymbol{\mu}}(\boldsymbol{\Psi}_i)\hat{\theta}_{n_1}^{\mathrm{IS},1}(\boldsymbol{\Psi}_i).$$

Choosing $\boldsymbol{\mu}$. A key point in the full IS approach is the determination of a value for $\boldsymbol{\mu}$ that gives a low variance for the IS estimator. Here we sketch the solution proposed by Glasserman and Li (2005). Since $\hat{\theta}_{n_1}^{\mathrm{IS},1}(\boldsymbol{\psi}) \approx P(L \geqslant c \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$, applying IS to the factors essentially amounts to finding a good IS density for the function $\boldsymbol{\psi} \to P(L \geqslant c \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$. Now recall from our discussion in the previous section that the optimal IS density g^* satisfies

$$g^*(\boldsymbol{\psi}) \propto P(L \geqslant c \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) \exp(-\tfrac{1}{2}\boldsymbol{\psi}'\Omega^{-1}\boldsymbol{\psi}), \tag{11.46}$$

where "$\propto$" stands for "proportional to". Sampling from that density is obviously not feasible, as the normalizing constant involves the exceedance probability $P(L \geqslant c)$ that we are interested in. In this situation the authors suggest using a multivariate normal density with the same mode as g^* as an approximation to the optimal IS

density. Since a normal density attains its mode at the mean $\boldsymbol{\mu}$, this amounts to choosing $\boldsymbol{\mu}$ as the solution to the optimization problem

$$\max_{\boldsymbol{\psi}} P(L \geqslant c \mid \boldsymbol{\Psi} = \boldsymbol{\psi}) \exp(-\tfrac{1}{2}\boldsymbol{\psi}' \Omega^{-1} \boldsymbol{\psi}). \tag{11.47}$$

An exact (numerical) solution of (11.47) is difficult because the function $P(L \geqslant c \mid \boldsymbol{\Psi} = \boldsymbol{\psi})$ is usually not available in closed form. Glasserman and Li (2005) discuss several approaches to overcoming this difficulty (see their paper for details).

Example 11.26. We give a very simple example of IS to show the gains that can be obtained by applying IS at both the level of the factor variables and the level of the conditional loss distribution.

Consider an exchangeable one-factor Bernoulli mixture model in which the factor Ψ is standard normally distributed and the conditional probability of default is given by

$$p_i(\psi) = P(Y_i = 1 \mid \Psi = \psi) = \Phi\left(\frac{\Phi^{-1}(p) + \sqrt{\rho}\psi}{\sqrt{1-\rho}}\right)$$

for all obligors i. Let the unconditional default probability be $p = 0.05$, let the asset correlation $\rho = 0.05$ and consider $m = 100$ obligors, each with an identical exposure $e_i = 1$. We are interested in the probability $\theta = P(L \geqslant 20)$, where $L = \sum_{i=1}^{m} Y_i$. In this set-up we can calculate, using numerical integration, that $\theta \approx 0.00112$, so $\{L \geqslant 20\}$ is a relatively rare event. In the first panel of Figure 11.3 we apply naive Monte Carlo estimation of θ and plot $\hat{\theta}_n^{\mathrm{MC}}$ against n; the true value is shown by a horizontal line.

In the second panel we apply importance sampling at the level of the factor Ψ using the value $\mu = -2.8$ for the mean of the distribution of Ψ under Q and plot the resulting estimate for different values of n, the number of random draws of the factor. In the third panel we apply IS at the level of the conditional loss distribution, using Algorithm 11.24 with $n_1 = 50$, but we apply naive Monte Carlo to the distribution of the factor.

In the final panel we apply full IS using Algorithm 11.25, and we plot $\hat{\theta}_n^{\mathrm{IS}}$ against n using $n_1 = 50$ as before. This is clearly the only estimate that appears to have converged to the true value by the time we have sampled $n = 10\,000$ values of the factor.

Notes and Comments

Our discussion of IS for credit portfolios follows Glasserman and Li (2005) closely. Theoretical results on the asymptotics of the IS estimator for large portfolios and numerical case studies contained in Glasserman and Li (2005) indicate that full IS is a very useful tool for dealing with large Bernoulli mixture models. Merino and Nyfeler (2004) and Kalkbrener, Lotter and Overbeck (2004) undertook related work—the latter paper gives an interesting alternative solution to finding a reasonable IS mean $\boldsymbol{\mu}$ for the factors.

For a general introduction to importance sampling we refer to the excellent textbook by Glasserman (2003) (see also Asmussen and Glynn 2007; Robert and Casella

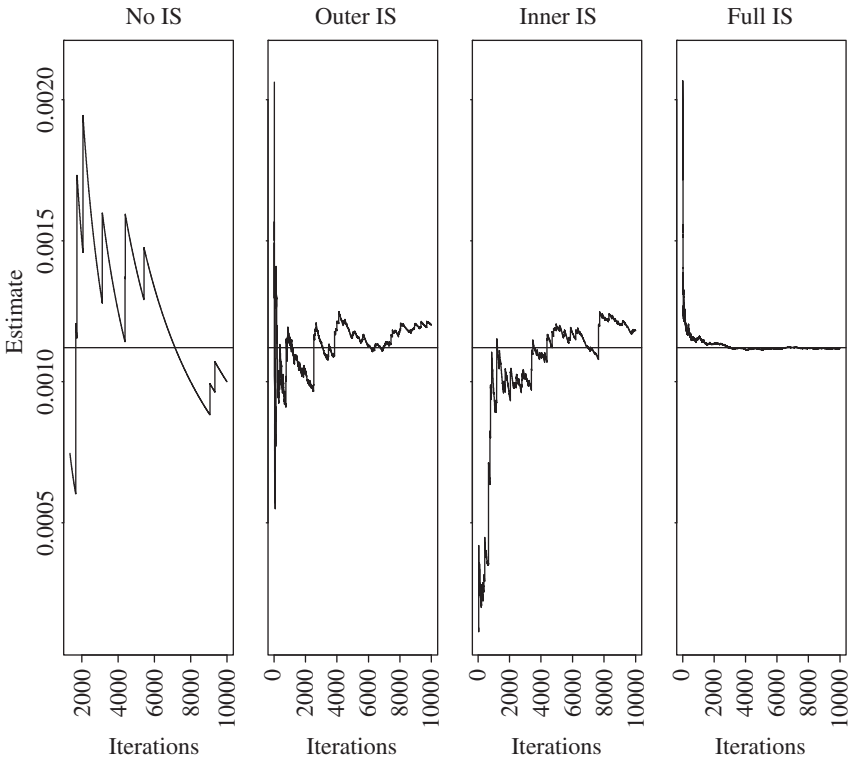

Figure 11.3. Illustration of the significant improvements that can be made in the estimation of rare-event probabilities for Bernoulli mixture models when importance sampling is applied at the level of both the factors and the conditional loss distribution given the factors (see Example 11.26 for details).

1999). For applications of importance sampling to heavy-tailed distributions, where exponential families cannot be applied directly, see Asmussen, Binswanger and Højgaard (2000) and Glasserman, Heidelberger and Shahabuddin (1999).

An alternative to simulation is the use of analytic approximations to the portfolio loss distribution. Applications of the saddle-point approximation (see Jensen 1995) are discussed in Martin, Thompson and Browne (2001) and Gordy (2002).

11.5 Statistical Inference in Portfolio Credit Models

In the remainder of this chapter we consider two different approaches to the estimation of portfolio credit risk models. In Section 11.5.1 we discuss the calibration of industry threshold models, such as the CreditMetrics model and the portfolio version of the Moody's public-firm EDF model; we focus on the estimation of the factor model describing the dependence structure of the critical variables using proxy data on equity or asset returns.

In Sections 11.5.2–11.5.4 we discuss the direct estimation of Bernoulli or Poisson mixture models from historical default data. This approach has been less widely applied in industry due to the relative scarcity of data on defaults, particularly for

higher-rated firms. However, it has become increasingly feasible with the availability of larger databases of historical defaults and rating migrations.

11.5.1 Factor Modelling in Industry Threshold Models

Recall from Section 11.1.3 that many industry models take the form of a Gaussian threshold model $(\boldsymbol{X}, \boldsymbol{d})$ with $\boldsymbol{X} \sim N_m(\mathbf{0}, P)$, where the random vector $\boldsymbol{X}$ contains the critical variables representing the credit quality or "ability-to-pay" of the obligors in the portfolio, the deterministic vector $\boldsymbol{d}$ contains the critical default thresholds, and P is the so-called asset correlation matrix, which is estimated with the help of a factor model for $\boldsymbol{X}$.

Industry models generally separate the calibration of the vector $\boldsymbol{d}$ (or the threshold matrix D in a multi-state model) and the calibration of the factor model for $\boldsymbol{X}$. As discussed in Section 11.1.3, in a default-only model the threshold d_i is usually set at $d_i = \Phi^{-1}(p_i)$, where p_i is an estimate of the default probability for obligor i for the time period in question (generally one year). Depending on the type of obligor, the default probability may be estimated in different ways: for larger corporates it may be estimated using credit ratings or using a firm-value approach, such as the Moody's public-firm EDF model; for retail obligors it may be estimated on the basis of credit scores.

We concentrate in this section on the estimation of the factor model for $\boldsymbol{X}$. We recall that this takes the form

$$X_i = \sqrt{\beta_i}\tilde{F}_i + \sqrt{1-\beta_i}\varepsilon_i, \quad i = 1, \ldots, m, \tag{11.48}$$

where $\tilde{F}_i$ and $\varepsilon_1, \ldots, \varepsilon_m$ are independent standard normal variables, and where $0 \leqslant \beta_i \leqslant 1$ for all i. The systematic variables $\tilde{F}_i$ are assumed to be of the form $\tilde{F}_i = \boldsymbol{a}_i'\boldsymbol{F}$, where $\boldsymbol{F}$ is a vector of common factors satisfying $\boldsymbol{F} \sim N_p(\mathbf{0}, \Omega)$ with $p < m$, and where Ω is a correlation matrix. The factors typically represent country and industry effects and the assumption that $\operatorname{var}(\tilde{F}_i) = 1$ imposes the constraint that $\boldsymbol{a}_i'\Omega\boldsymbol{a}_i = 1$ for all i.

Different industry models use different data for $\boldsymbol{X}$ to calibrate the factor model (11.48). The Moody's Analytics Global Correlation, or GCorr, model has sub-models for many different kinds of obligor, including public corporate firms, private firms, small and medium-sized enterprises (SMEs), retail customers and sovereigns (Huang et al. 2012). The sub-model for public firms (GCorr Corporate) is calibrated using data on weekly asset value returns, where asset values are determined as part of the public-firm EDF methodology described in Section 10.3.3. In the CreditMetrics framework, weekly equity returns are viewed as a proxy for asset returns and used to estimate the factor model (RiskMetrics Group 1997). In both cases the data contain information about changing credit quality.

We now provide a sketch of a generic procedure for estimating a factor model for corporates where the factors have country- and industry-sector interpretations. Specific industry models follow this procedure in outline but may differ in the details of the calculations at certain steps. We assume that we have a high-dimensional multivariate time series $(\boldsymbol{X}_t)_{1 \leqslant t \leqslant n}$ of asset returns (or other proxy data for changing

credit quality) over a period of time in which stationarity can be assumed; we also assume that each component time series has been scaled to have mean 0 and variance 1.

(1) We first fix the structure of the factor vector $\boldsymbol{F}$. For example, the first block of components might represent country factors and the second block of components might represent industry factors. We then assign vectors of factor weights $\boldsymbol{a}_i$ to each obligor based on our knowledge of the companies. The elements of $\boldsymbol{a}_i$ may simply consist of ones and zeros if the company can be clearly identified with a single country and industry, but may also consist of weights if the company has significant activity in more than one country or more than one industry sector. For example, a firm that does 60% of its business in one country and 40% in another would be coded with weights of 0.6 and 0.4 in the relevant positions of $\boldsymbol{a}_i$.

(2) We then use cross-sectional estimation techniques to estimate the factor values $\boldsymbol{F}_t$ at each time point t. Effectively, the factor estimates $\hat{\boldsymbol{F}}_t$ are constructed as weighted sums of the $X_{t,i}$ data for obligors i that are exposed to each factor. One way of achieving this is to construct a matrix A with rows $\boldsymbol{a}_i$ and then to estimate a fundamental factor model of the form $\boldsymbol{X}_t = A\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t$ at each time point t, as described in Section 6.4.4.

(3) The raw factor estimates form a multivariate time series of dimension p. We standardize each component series to have mean 0 and variance 1 to obtain $(\hat{\boldsymbol{F}}_t)_{1\leqslant t\leqslant n}$ and calculate the sample covariance matrix of the standardized factor estimates, which serves as our estimate of Ω.

(4) We then scale the vectors of factor weights $\boldsymbol{a}_i$ such that the conditions $\boldsymbol{a}_i'\hat{\Omega}\boldsymbol{a}_i = 1$ are met for each obligor.

(5) Time series of estimated systematic variables for each obligor are then constructed by calculating $\tilde{F}_{t,i} = \boldsymbol{a}_i'\hat{\boldsymbol{F}}_t$ for $t = 1, \ldots, n$.

(6) Finally, we estimate the β_i parameters by performing a time-series regression of $X_{t,i}$ on $\tilde{\tilde{F}}_{t,i}$ for each obligor.

Note that the accurate estimation of the β_i in the last step is particularly important. In Section 11.1.5 we showed that there is considerable model risk associated with the size of the specific risk component, particularly when the tail of a credit loss distribution is of central importance. The estimate of β_i is the so-called R-squared of the time-series regression model in step (6) and will be largest for the firms whose credit-quality changes are best explained by systematic factors.

11.5.2 Estimation of Bernoulli Mixture Models

We now turn our attention to the estimation of Bernoulli mixture models of portfolio credit risk from historical default data. The models we describe are motivated by the format of the data we consider, which can be described as *repeated cross-sectional data*. This kind of data, comprising observations of the default or non-default of groups of monitored companies in a number of time periods, can be readily extracted

from the rating-migration and default databases of rating agencies. Since the group of companies may differ from period to period, as new companies are rated and others default or cease to be rated, we have a cross-section of companies in each period, but the cross-section may change from period to period.

In this section we discuss the estimation of default probabilities and default correlations for homogeneous groups, e.g. groups with the same credit rating. In Sections 11.5.3 and 11.5.4 we consider more complicated one-factor models allowing more heterogeneity and make a link to the important class of generalized linear mixed models (GLMMs) used in many statistical applications.

Suppose that we observe historical default numbers over n periods of time for a homogeneous group; typically these might be yearly data. For $t = 1, \ldots, n$, let m_t denote the number of observed companies at the start of period t and let M_t denote the number that defaulted during the period; the former will be treated as fixed at the outset of the period and the latter as an rv. Suppose further that within a time period these defaults are generated by an exchangeable Bernoulli mixture model of the kind described in Section 11.2.2. In other words, assume that, given some mixing variable Q_t taking values in $(0, 1)$ and the cohort size m_t, the number of defaults M_t is conditionally binomially distributed and satisfies $M_t \mid Q_t = q \sim B(m_t, q)$. Further assume that the mixing variables $Q_1, \ldots, Q_n$ are identically distributed. We consider two methods for estimating the fundamental parameters of the mixing distribution $\pi = \pi_1, \pi_2$ and ρ_Y (default correlation); these are the method of moments and the maximum likelihood method.

A simple moment estimator. For $1 \leqslant t \leqslant n$, let $Y_{t,1}, \ldots, Y_{t,m_t}$ be default indicators for the m_t companies in the cohort. Suppose we define the rv

$$\binom{M_t}{k} := \sum_{\{i_1,\ldots,i_k\}\subset\{1,\ldots,m_t\}} Y_{t,i_1} \cdots Y_{t,i_k}; \tag{11.49}$$

this represents the number of possible subgroups of k obligors among the defaulting obligors in period t (and takes the value zero when $k > M_t$). By taking expectations in (11.49) we get

$$E\left(\binom{M_t}{k}\right) = \binom{m_t}{k}\pi_k$$

and hence

$$\pi_k = E\left(\binom{M_t}{k}\right) \Big/ \binom{m_t}{k}.$$

We estimate the unknown theoretical moment π_k by taking a natural empirical average (11.50) constructed from the n years of data:

$$\hat{\pi}_k = \frac{1}{n}\sum_{t=1}^{n} \frac{\binom{M_t}{k}}{\binom{m_t}{k}} = \frac{1}{n}\sum_{t=1}^{n} \frac{M_t(M_t-1)\cdots(M_t-k+1)}{m_t(m_t-1)\cdots(m_t-k+1)}. \tag{11.50}$$

For $k = 1$ we get the standard estimator of default probability

$$\hat{\pi} = \frac{1}{n}\sum_{t=1}^{n} \frac{M_t}{m_t},$$

and ρ_Y can obviously be estimated by taking $\hat{\rho}_Y = (\hat{\pi}_2 - \hat{\pi}^2)/(\hat{\pi} - \hat{\pi}^2)$. The estimator is unbiased for π_k and consistent as $n \to \infty$ (for more details see Frey and McNeil (2001)). Note that, for Q_t random, consistency requires observations for a large number of years; it is not sufficient to observe a large pool in a single year.

Maximum likelihood estimators. To implement a maximum likelihood (ML) procedure we assume a simple parametric form for the density of the Q_t (such as beta, logit-normal or probit-normal). The joint probability function of the default counts $M_1, \ldots, M_n$ given the cohort sizes $m_1, \ldots, m_n$ can then be calculated using (11.14), under the assumption that the Q_t variables in different years are independent. This expression is then maximized with respect to the natural parameters of the mixing distribution (i.e. a and b in the case of beta and μ and σ for the logit-normal and probit-normal). Of course, independence may be an unrealistic assumption for the mixing variables, due to the phenomenon of economic cycles, but the method could then be regarded as a quasi-maximum likelihood (QML) procedure, which misspecifies the serial dependence structure but correctly specifies the marginal distribution of defaults in each year and still gives reasonable parameter estimates.

In practice, it is easiest to use the beta mixing distribution, since, in this case, given the group size m_t in period t, the rv M_t has a beta-binomial distribution with probability function given in (11.16). The likelihood to be maximized therefore takes the form

$$L(a, b; \text{data}) = \prod_{t=1}^{n} \binom{m_t}{M_t} \frac{\beta(a + M_t, b + m_t - M_t)}{\beta(a, b)},$$

and maximization can be performed numerically with respect to a and b. For further information about the ML method consult Section A.3. The ML estimates of $\pi = \pi_1, \pi_2$ and ρ_Y are calculated by evaluating moments of the fitted distribution using (11.15); the formulas are given in Example 11.7.

A comparison of moment estimation and ML estimation. To compare these two approaches we conduct a simulation study summarized in Table 11.4. To generate data in the simulation study we consider the beta, probit-normal and logit-normal mixture models of Section 11.2.2. In any single experiment we generate 20 years of data using parameter values that roughly correspond to one of the Standard & Poor's credit ratings CCC, B or BB (see Table 11.3 for the parameter values). The number of firms m_t in each of the years is generated randomly using a binomial-beta model to give a spread of values typical of real data; the defaults are then generated using one of the Bernoulli mixture models, and estimates of π, π_2 and ρ_Y are calculated. The experiment is repeated 5000 times and a relative root mean square error (RRMSE) is estimated for each parameter and each method: that is, we take the square root of the estimated MSE and divide by the true parameter value. Methods are compared by calculating the percentage increase of the estimated RRMSE with respect to the better method (i.e. the RRMSE-minimizing method) for each parameter.

It may be concluded from Table 11.4 that the ML method is better in all but one experiment. Surprisingly, it is better even in the experiments when it is misspecified

Table 11.4. Each part of the table relates to a block of 5000 simulations using a particular exchangeable Bernoulli mixture model with parameter values roughly corresponding to a particular S&P rating class. For each parameter of interest, an estimated RRMSE is tabulated for both estimation methods: moment estimation using (11.50) and ML estimation based on the beta model. Methods can be compared by using Δ, the percentage increase of the estimated RRMSE with respect to the better method (i.e. the RRMSE-minimizing method) for each parameter. For each parameter the better method therefore has $\Delta = 0$. The table clearly shows that MLE is at least as good as the moment estimator in all but one case.

			Moment		MLE-beta	
Group	True model	Parameter	RRMSE	Δ	RRMSE	Δ
CCC	Beta	π	0.101	0	0.101	0
CCC	Beta	π_2	0.202	0	0.201	0
CCC	Beta	ρ_Y	0.332	5	0.317	0
CCC	Probit-normal	π	0.100	0	0.100	0
CCC	Probit-normal	π_2	0.205	1	0.204	0
CCC	Probit-normal	ρ_Y	0.347	11	0.314	0
CCC	Logit-normal	π	0.101	0	0.101	0
CCC	Logit-normal	π_2	0.209	1	0.208	0
CCC	Logit-normal	ρ_Y	0.357	11	0.320	0
B	Beta	π	0.130	0	0.130	0
B	Beta	π_2	0.270	0	0.269	0
B	Beta	ρ_Y	0.396	8	0.367	0
B	Probit-normal	π	0.130	0	0.130	0
B	Probit-normal	π_2	0.286	3	0.277	0
B	Probit-normal	ρ_Y	0.434	19	0.364	0
B	Logit-normal	π	0.131	0	0.132	0
B	Logit-normal	π_2	0.308	7	0.289	0
B	Logit-normal	ρ_Y	0.493	26	0.392	0
BB	Beta	π	0.199	0	0.199	0
BB	Beta	π_2	0.435	0	0.438	1
BB	Beta	ρ_Y	0.508	7	0.476	0
BB	Probit-normal	π	0.197	0	0.197	0
BB	Probit-normal	π_2	0.492	10	0.446	0
BB	Probit-normal	ρ_Y	0.607	27	0.480	0
BB	Logit-normal	π	0.196	0	0.196	0
BB	Logit-normal	π_2	0.572	24	0.462	0
BB	Logit-normal	ρ_Y	0.752	45	0.517	0

and the true mixing distribution is either probit-normal or logit-normal; in fact, in these cases, it offers more of an improvement than in the beta case. This can partly be explained by the fact that when we constrain well-behaved, unimodal mixing distributions with densities to have the same first and second moments, these distributions are very similar (see Figure 11.2). Finally, we observe that the ML method tends to outperform the moment method more as we increase the credit quality, so that defaults become rarer.

11.5.3 Mixture Models as GLMMs

A one-factor Bernoulli mixture model. Recall the simple one-factor model (11.17), which generalizes the exchangeable model in Section 11.2.2, and consider the case where $\sigma_i = \sigma$ for all obligors i. Rewriting slightly, this model has the form

$$p_i(\Psi) = h(\mu + \boldsymbol{\beta}' \boldsymbol{x}_i + \Psi), \tag{11.51}$$

where h is a link function, the vector $\boldsymbol{x}_i$ contains covariates for the ith firm, such as indicators for group membership or key balance sheet ratios, and $\boldsymbol{\beta}$ and μ are model parameters. Examples of link functions include the standard normal df $\Phi(x)$ and the logistic df $(1 + \mathrm{e}^{-x})^{-1}$. The scale parameter σ has been subsumed in the normally distributed random variable $\Psi \sim N(0, \sigma^2)$, representing a common or systematic factor.

This model can be turned into a multi-period model for default counts in different periods by assuming that a series of mixing variables $\Psi_1, \ldots, \Psi_n$ generates default dependence in each time period $t = 1, \ldots, n$. The default indicator $Y_{t,i}$ for the ith company in time period t is assumed to be Bernoulli with default probability $p_{t,i}(\Psi_t)$ depending on Ψ_t according to

$$p_{t,i}(\Psi_t) = h(\mu + \boldsymbol{x}_{t,i}' \boldsymbol{\beta} + \Psi_t), \tag{11.52}$$

where $\Psi_t \sim N(0, \sigma^2)$ and $\boldsymbol{x}_{t,i}$ are covariates for the ith company in time period t. Moreover, the default indicators $Y_{t,1}, \ldots, Y_{t,m_t}$ in period t are assumed to be conditionally independent given Ψ_t.

To complete the model we need to specify the joint distribution of $\Psi_1, \ldots, \Psi_n$, and it is easiest to assume that these are iid mixing variables. To capture possible economic cycle effects causing dependence between numbers of defaults in successive time periods, one could either enter covariates at the level of $\boldsymbol{x}_{ti}$ that are known to be good proxies for "the state of the economy", such as changes in GDP over the time period, or an index like the Chicago Fed National Activity Index (CFNAI) in the US, or one could consider a serially dependent time-series structure for the systematic factors (Ψ_t).

A one-factor Poisson mixture model. When considering higher-grade portfolios of companies with relatively low default risk, there may sometimes be advantages (particularly in the stability of fitting procedures) in formulating Poisson mixture models instead of Bernoulli mixture models. A multi-period mixture model based on Definition 11.14 can be constructed by assuming that the default count variable $\tilde{Y}_{t,i}$ for the ith company in time period t is conditionally Poisson with rate parameter $\lambda_{t,i}(\Psi_t)$ depending on Ψ_t according to

$$\lambda_{t,i}(\Psi_t) = \exp(\mu + \boldsymbol{x}_{t,i}' \boldsymbol{\beta} + \Psi_t), \tag{11.53}$$

with all other elements of the model as in (11.52). Again the variables $\tilde{Y}_{t,1}, \ldots, \tilde{Y}_{t,m_t}$ are assumed to be conditionally independent given Ψ_t.

GLMMs. Both the multi-period Bernoulli and Poisson mixture models in (11.52) and (11.53) belong to a family of widely used statistical models known as *generalized linear mixed models* (GLMMs). The three basic elements of such a model are as follows.

(1) The vector of *random effects*. In our examples this is the vector $(\Psi_1, \ldots, \Psi_n)$ containing the systematic factors for each time period.

(2) A distribution from the *exponential family* for the conditional distribution of the responses ($Y_{t,i}$ or $\tilde{Y}_{t,i}$) given the random effects. Responses are assumed to be conditionally independent given the random effects. The Bernoulli, binomial and Poisson distributions all belong to the exponential family (see, for example, McCullagh and Nelder 1989, p. 28).

(3) A *link function* relating $E(Y_{t,i} \mid \Psi_t)$, the mean response conditional on the random effects, to the so-called *linear predictor*. In our examples the linear predictor for $Y_{t,i}$ is

$$\eta_{t,i}(\Psi_t) = \mu + \boldsymbol{x}_{t,i}'\boldsymbol{\beta} + \Psi_t. \tag{11.54}$$

We have considered the so-called probit and logit link functions in the Bernoulli case and the log-link function in the Poisson case. (Note that it is usual in GLMMs to write the model as $g(E(Y_{t,i} \mid \Psi_t)) = \eta_{t,i}(\Psi_t)$ and to refer to g as the link function; hence the probit link function is the quantile function of the standard normal, and the link in the Poisson case (11.53) is referred to as "log" rather than "exponential".)

When no random effects are modelled in a GLMM, the model is simply known as a generalized linear model, or GLM. The role of the random effects in the GLMM is, in a sense, to capture patterns of variability in the responses that cannot be explained by the observed covariates alone, but which might be explained by additional unobserved factors. In our case, these unobserved factors are bundled into a time-period effect that we loosely describe as the state of the economy in that time period; alternatively, we refer to it as the systematic risk.

The GLMM framework allows models of much greater complexity. We can add further random effects to obtain multi-factor mixture models. For example, we might know the industry sector of each firm and wish to include random effects for sectors that are nested within the year effect; in this way we might capture additional variability associated with economic effects in different sectors over and above the global variability associated with the year effect. Such models can be considered in the GLMM framework by allowing the linear predictor in (11.54) to take the form $\eta_{t,i}(\Psi_t) = \mu + \boldsymbol{x}_{t,i}'\boldsymbol{\beta} + \boldsymbol{z}_{t,i}'\boldsymbol{\Psi}_t$ for some vector of random effects $\boldsymbol{\Psi}_t = (\Psi_{t,1}, \ldots, \Psi_{t,p})'$; the vector $\boldsymbol{z}_{t,i}$ is a known *design element* of the model that selects the random effects that are relevant to the response $Y_{t,i}$. We would then have a total of $p \times n$ random effects in the model. We may or may not want to model serial dependence in the time series $\boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_n$.

Inference for GLMMs. Full ML inference for a GLMM is an option for the simplest models. Consider the form of the likelihood for the one-factor models in (11.52) and (11.53). If we write $p_{Y_{t,i}|\Psi_t}(y \mid \psi)$ for the conditional probability mass function of the response $Y_{t,i}$ (or $\tilde{Y}_{t,i}$) given Ψ_t, we have, for data $\{Y_{t,i} : t = 1, \ldots, n,\ i = 1, \ldots, m_t\}$,

$$L(\boldsymbol{\beta}, \sigma; \text{data}) = \int \cdots \int \left(\prod_{t=1}^{n} \prod_{i=1}^{m_t} p_{Y_{t,i}|\Psi_t}(Y_{t,i} \mid \psi_t) \right) f(\psi_1, \ldots, \psi_n)\, \mathrm{d}\psi_1 \cdots \mathrm{d}\psi_n, \tag{11.55}$$

where f denotes the assumed joint density of the random effects. If we do not assume independent random effects from time period to time period, then we are faced with an n-dimensional integral (or an $(n \times p)$-dimensional integral in multi-factor models). Assuming iid Gaussian random effects with marginal Gaussian density f_Ψ, the likelihood (11.55) becomes

$$L(\boldsymbol{\beta}, \sigma; \text{data}) = \prod_{t=1}^{n} \left(\int \prod_{i=1}^{m_t} p_{Y_{t,i}|\Psi_t}(Y_{t,i} \mid \psi_t) f_\Psi(\psi_t)\, \mathrm{d}\psi_t \right), \tag{11.56}$$

so we have a product of one-dimensional integrals and this can be easily evaluated numerically and maximized over the unknown parameters. Alternatively, faster approximate likelihood methods, such as penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL), can be used (see Notes and Comments).

Another attractive possibility is to treat inference for these models from a Bayesian point of view and to use Markov chain Monte Carlo (MCMC) methods to make inferences about parameters (McNeil and Wendin 2006, 2007, see, for example,). The Bayesian approach has two main advantages. First, a Bayesian MCMC approach allows us to work with much more complex models than can be handled in the likelihood framework, such as a model with serially dependent random effects. Second, the Bayesian approach is ideal for handling the considerable parameter uncertainty in portfolio credit risk, particularly in models for higher-rated counterparties, where default data are scarce.

11.5.4 A One-Factor Model with Rating Effect

In this section we fit a Bernoulli mixture model to annual default count data from Standard & Poor's for the period 1981–2000; these data have been reconstructed from published default rates in Brand and Bahr (2001, Table 13, pp. 18–21). Standard & Poor's uses the ratings AAA, AA, A, BBB, BB, B, CCC, but because the observed one-year default rates for AAA-rated and AA-rated firms are mostly zero, we concentrate on the rating categories A–CCC.

In our model we assume a single yearly random effect representing the state of the economy and treat the rating category as an observed covariate for each firm in each time period. Our model is a particular instance of the one-factor Bernoulli mixture model in (11.52) and a multi-period extension of the model described in Example 11.8. We assume for simplicity that random effects in each year are iid normal, which allows us to use the likelihood (11.56).

Since we are able to pool companies into groups by year and rating category, we note that it is possible to reformulate the model as a binomial mixture model. Let $r = 1, \ldots, 5$ index the five rating categories in our study, and write $m_{t,r}$ for the number of followed companies in year t with rating r, and $M_{t,r}$ for the number of these that default. Our model assumption is that, conditional on Ψ_t (and the group sizes), the default counts $M_{t,1}, \ldots, M_{t,5}$ are independent and are distributed in such a way that $M_{t,r} \mid \Psi_t = \psi \sim B(m_{t,r}, p_r(\psi))$. Using the probit link, the conditional default probability of an r-rated company in year t is given by

$$p_r(\Psi_t) = \Phi(\mu_r + \Psi_t). \tag{11.57}$$

The model may be fitted under the assumption of iid random effects in each year by straightforward maximization of the likelihood in (11.56). The parameter estimates and obtained standard errors are given in Table 11.5, together with the estimated default probabilities $\hat{\pi}^{(r)}$ for each rating category and estimated default correlations $\hat{\rho}_Y^{(r_1,r_2)}$ implied by the parameter estimates. Writing Ψ for a generic random effect variable, the default probability for rating category r is given by

$$\hat{\pi}^{(r)} = E(\hat{p}_r(\Psi)) = \int_{-\infty}^{\infty} \Phi(\hat{\mu}_r + \hat{\sigma} z)\phi(z)\,\mathrm{d}z, \quad 1 \leqslant r \leqslant 5,$$

where ϕ is the standard normal density. The default correlation for two firms with ratings r_1 and r_2 in the same year is calculated easily from the joint default probability for these two firms, which is

$$\hat{\pi}_2^{(r_1,r_2)} = E(\hat{p}_{r_1}(\Psi)\hat{p}_{r_2}(\Psi)) = \int_{-\infty}^{\infty} \Phi(\hat{\mu}_{r_1} + \hat{\sigma} z)\Phi(\hat{\mu}_{r_2} + \hat{\sigma})\phi(z)\,\mathrm{d}z.$$

The default correlation is then

$$\hat{\rho}_Y^{(r_1,r_2)} = \frac{\hat{\pi}_2^{(r_1,r_2)} - \hat{\pi}^{(r_1)}\hat{\pi}^{(r_2)}}{\sqrt{(\hat{\pi}^{(r_1)} - (\hat{\pi}^{(r_1)})^2)(\hat{\pi}^{(r_2)} - (\hat{\pi}^{(r_2)})^2)}}.$$

Note that the default correlations are correlations between event indicators for very low probability events and are necessarily very small.

The model in (11.57) assumes that the variance of the systematic factor Ψ_t is the same for all firms in all years. When compared with the very general Bernoulli mixture model (11.23) we might be concerned that the simple model considered in this section does not allow for enough heterogeneity in the variance of the systematic risk. A simple extension of the model is to allow the variance to be different for different rating categories: that is, to fit a model where $p_r(\Psi_t) = \Phi(\mu_r + \sigma_r \Psi_t)$ and where Ψ_t is a standard normally distributed random effect. This increases the number of parameters in the model by four but is no more difficult to fit than the basic model. The maximized value of the log-likelihood in the model with heterogeneous scaling is -2557.4, and the value in the model with homogeneous scaling is -2557.7; a likelihood ratio test suggests that no significant improvement results from allowing heterogeneous scaling. If rating is the only categorical variable, the simple model seems adequate, but if we had more information on the industrial and geographical sectors to which the companies belonged, it would be natural to introduce further

Table 11.5. Maximum likelihood parameter estimates and standard errors (se) for a one-factor Bernoulli mixture model fitted to historical Standard & Poor's one-year default data, together with the implied estimates of default probabilities $\hat{\pi}^{(r)}$ and default correlations $\hat{\rho}_Y^{(r_1,r_2)}$. The MLE of the scaling parameter σ is 0.24 with standard error 0.05. Note that we have tabulated default correlation in absolute terms and not in percentage terms.

Parameter	A	BBB	BB	B	CCC	
μ_r	−3.43	−2.92	−2.40	−1.69	−0.84	
se (μ_r)	0.13	0.09	0.07	0.06	0.08	
$\pi^{(r)}$	0.000 4	0.002 3	0.009 7	0.050 3	0.207 8	
$\rho_Y^{(r_1,r_2)}$	0.000 40	0.000 77	0.001 30	0.002 19	0.003 04	A
	0.000 77	0.001 49	0.002 55	0.004 35	0.006 15	BBB
	0.001 30	0.002 55	0.004 40	0.007 63	0.010 81	BB
	0.002 19	0.004 35	0.007 63	0.013 28	0.019 06	B
	0.003 04	0.006 15	0.010 81	0.019 06	0.027 88	CCC

random effects for these sectors and to allow more heterogeneity in the model in this way.

The implied default probability and default correlation estimates in Table 11.5 can be a useful resource for calibrating simple credit models to homogeneous groups defined by rating. For example, to calibrate a Clayton copula to group BB we use the inputs $\pi^{(3)} = 0.0097$ and $\rho_Y^{(3,3)} = 0.0044$ to determine the parameter θ of the Clayton copula (see Example 11.13). Note also that we can now immediately use the scaling results of Section 11.3 to calculate approximate risk measures for large portfolios of companies that have been rated with the Standard & Poor's system (see Example 11.20).

Notes and Comments

The main references for our account of industry factor models are Huang et al. (2012) and RiskMetrics Group (1997).

The estimator (11.50) for joint default probabilities is also used in Lucas (1995) and Nagpal and Bahar (2001), although de Servigny and Renault (2002) suggest there may be problems with this estimator for groups with low default rates. A related moment-style estimator has been suggested by Gordy (2000) and appears to have a similar performance to (11.50) (see Frey and McNeil 2003). A further paper on default correlation estimation is Gordy and Heitfield (2002).

A good overview article on generalized linear mixed models is Clayton (1996). For generalized linear models a standard reference is McCullagh and Nelder (1989) (see also Fahrmeir and Tutz 1994).

The analysis of Section 11.5.4 is very similar to the analysis in Frey and McNeil (2003) (where heterogeneous variances for each rating category were assumed). The results reported in this book were obtained by full maximization of the likelihood using our own R code. Very similar results are obtained with the `glmer` function

in the `lme4` R package, which maximizes an adaptive Gauss–Hermite approximation to the log-likelihood. GLMMS may also be estimated using the approximate penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL) methods (see Breslow and Clayton 1993). For a Bayesian approach to fitting the model using Markov chain Monte Carlo techniques, see McNeil and Wendin (2007), who also incorporate an autoregressive time-series structure for the random effects.

Although we have only described default models it is also possible to analyse rating migrations in the generalized linear model framework (with or without random effects). A standard model is the ordered probit model, which is used without random effects in Nickell, Perraudin and Varotto (2000) to provide evidence of time variation in default rates attributable to macroeconomic factors; a similar message is found in Bangia et al. (2002). McNeil and Wendin (2006) show how random effects and unobserved factors may be included in such models and carry out Bayesian inference. See also Gagliardini and Gouriéroux (2005), in which a variety of rating-migration models with serially dependent unobserved factors are studied.

There is a large literature on models with latent structure designed to capture the dynamics of systematic risk, and there is quite a lot of variation in the types of model considered. Crowder, Davis and Giampieri (2005) use a two-state hidden Markov structure to capture periods of high and low default risk, Koopman, Lucas and Klaassen (2005) use an unobserved components time-series model to describe US company failure rates, Koopman, Lucas and Monteiro (2008) develop a latent factor intensity model for rating transitions, and Koopman, Lucas and Schwaab (2012) combine macroeconomic factors, unobserved frailties and industry effects in a model of US defaults through the crisis of 2008.

12
Portfolio Credit Derivatives

In this chapter we study portfolio credit derivatives such as collateralized debt obligations (CDOs) and related products. The primary use of portfolio credit derivatives is in the securitization of credit risk: that is, the transformation of credit risk into securities that may be bought and sold by investors. The market for portfolio credit derivatives peaked in the period leading up to the 2007–9 credit crisis (as discussed in Section 1.2.1) and has only partly recovered since.

In Section 12.1 we describe the most important portfolio credit derivatives and their properties. We also provide some more discussion of the role that these products played in the credit crisis. Section 12.2 introduces copula models for portfolio credit risk, which have become the market standard in pricing CDOs and related credit derivatives. In Section 12.3 we discuss pricing and model calibration in factor copula models. More advanced dynamic portfolio credit risk models are studied in Chapter 17.

This chapter makes extensive use of the analysis of single-name credit risk models in Sections 10.1–10.4 and of basic notions in copula theory. We restrict our attention to models for random default times with deterministic hazard functions without adding the extra complexity of doubly stochastic default times as in Sections 10.5 and 10.6. The simpler models are sufficient to understand the key features of portfolio credit derivative pricing.

12.1 Credit Portfolio Products

In this section we describe the pay-off and qualitative properties of certain important credit portfolio products such as CDOs. We begin by introducing the necessary notation.

We consider a portfolio of m firms with default times $\tau_1, \dots, \tau_m$. In keeping with the notation introduced in the earlier credit chapters, the random vector $\boldsymbol{Y}_t = (Y_{t,1}, \dots, Y_{t,m})'$ with $Y_{t,i} = I_{\{\tau_i \leqslant t\}}$ describes the default state of the portfolio at some point in time $t \geqslant 0$. Note that $Y_{t,i} = 1$ if firm i has defaulted by time t, and $Y_{t,i} = 0$ otherwise. We assume throughout that there are no simultaneous defaults, so we may define the *ordered default times* $T_0 < T_1 < \cdots < T_m$ by setting $T_0 = 0$ and recursively setting

$$T_n := \min\{\tau_1, \dots, \tau_m : \tau_i > T_{n-1}\}, \quad 1 \leqslant n \leqslant m.$$

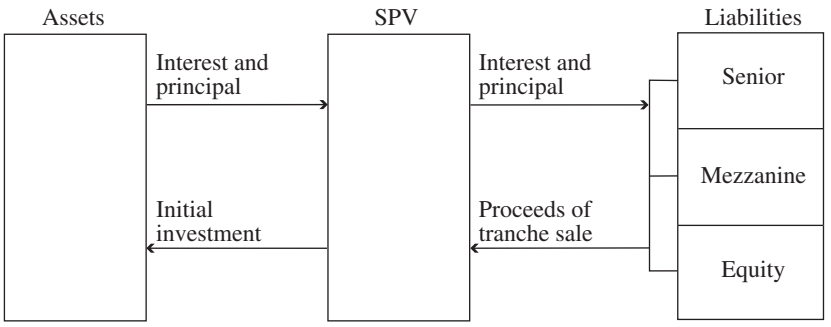

Figure 12.1. Schematic representation of the payments in a CDO structure.

As in Chapter 11, the exposure to firm (or so-called reference entity) i is denoted by e_i and the percentage loss given default (LGD) of firm i is denoted by $\delta_i \in [0, 1]$. The *cumulative loss* of the portfolio up to time t is therefore given by $L_t = \sum_{i=1}^{m} \delta_i e_i Y_{t,i}$. While δ_i and e_i may in principle be random, we mostly work with deterministic exposures and LGDs; further assumptions about these quantities are introduced as and when needed.

12.1.1 Collateralized Debt Obligations

Before the credit crisis of 2007–9 CDO markets were a fast-growing segment of the credit market. Although activity on CDO markets has slowed down since the crisis, CDOs and credit products with a similar structure remain an important asset class for risk managers to study.

A CDO is a financial instrument for the securitization of a portfolio of credit products such as bonds, loans or mortgages. This portfolio forms the so-called *asset pool* underlying the contract. The CDOs that are traded in practice come in many different varieties, but the basic structure is the same. The assets are sold to a *special-purpose vehicle* (SPV): a company that has been set up with the single purpose of carrying out the securitization deal. To finance the acquisition of the assets, the SPV issues securities in tranches of differing seniority, which form the *liability side* of the structure. The tranches of the liability side are called (in order of increasing seniority) equity, mezzanine and senior tranches (sometimes there are also super-senior tranches). The rules that determine the exact cash flow of the tranches are known as the *waterfall structure* of the CDO. These rules can be quite complex. Roughly speaking, the waterfall structure ensures that losses due to credit events on the asset side are borne first by the equity tranche; if the equity tranche is exhausted, losses are borne by the mezzanine tranches and only thereafter by the senior tranches. The credit quality of the more senior tranches is therefore usually higher than the average credit quality of the asset pool. The payments associated with a typical CDO are depicted schematically in Figure 12.1.

CDOs where the asset pool consists mainly of bonds are known as collateralized bond obligations (CBOs); if the asset side consists mainly of loans, a CDO is termed a collateralized loan obligation (CLO); CDOs for the securitization of mortgages are also known as mortgage-backed securities (MBSs) or asset-backed securities

(ABSs). There is also a liquid market for *synthetic* CDO tranches. In these contracts payments are triggered by default events in a pool of reference names (typically major corporations), but there are no actual assets underlying the contract. A precise pay-off description for synthetic CDO tranches is given in Section 12.1.2.

There are a number of economic motivations for arranging a CDO transaction. To begin with, in a typical CDO structure a large part of the asset pool is allocated to the senior tranches with a fairly high credit quality, even if the quality of the underlying assets is substantially lower. For instance, according to Hull and White (2010), for a typical ABS created from residential mortgages, about 75–80% of the underlying mortgage principal is allocated to senior tranches with a AAA rating. Many institutional investors prefer an investment in highly rated securities because of legal or institutional constraints. Securitization can therefore be a way to sell a large part of the underlying assets to investors who are unable to invest directly in the asset pool. Another incentive to set up a CDO transaction is related to capital adequacy; CDOs are often issued by banks who want to sell some of the credit-risky securities on their balance sheet in order to reduce their regulatory capital requirements.

Securitization via CDOs or ABSs is an important tool in credit markets. It allows lenders to reduce concentration risk and leverage and to refinance themselves more efficiently. Securitization can therefore increase the lending capacity of the financial sector. On the other hand, the credit crisis of 2007–9 clearly exposed a number of problems related to securitization and the use of asset-backed CDOs.

- Securitization can create *incentive problems*: if a mortgage originator knows that most of the mortgages he sells to homeowners will be securitized later on, he has little interest in evaluating the credit quality of the borrowers carefully. This can lead to a deterioration of lending standards. There is a lot of evidence that this actually happened in the years preceding the subprime crisis (see, for example, Crouhy, Jarrow and Turnbull 2008). This problem could be addressed by better aligning the interests of loan originators and of ABS investors. For instance, originators could be forced to keep a certain percentage of all the tranches they sell on the securitization market.
- The exact cash-flow structure of most asset-backed CDOs is very complicated. In fact, the legal documentation defining the payments of an asset-backed CDO can run to several hundred pages. This makes it difficult for investors to form an opinion of the value and the riskiness of any given tranche, thus contributing to the low trading volume in securitization markets during the credit crisis. At the height of the subprime crisis CDO products were on offer that crossed ethical and even legal boundaries. An example is the infamous ABACUS 2700-AC-1 CDO of Goldman Sachs (see Duffie (2010) for details).
- CDOs and ABSs were clearly misused by banks before the credit crisis to exploit *regulatory arbitrage*. They allowed loan-related credit risk to be transferred from the banking book to the trading book, where it enjoyed a more lenient capital treatment. By holding tranches that had been overoptimistically

rated as AAA in the trading book, banks were able to substantially lower their regulatory capital.

- The pay-off distribution associated with CDO tranches is very sensitive to the characteristics of the underlying asset pool, which makes it difficult to properly assess the risk associated with these instruments. In particular, a small increase in default correlation often leads to a substantial increase in the likelihood of losses for senior tranches. An intuitive explanation of this correlation sensitivity is given below.

Stylized CDOs and correlation sensitivity. In order to gain a better understanding of the main qualitative features of CDOs without getting bogged down in the details of the waterfall structure, we introduce a hypothetical contract that we label a *stylized CDO*. We consider a portfolio of m firms with cumulative loss $L_t = \sum_{i=1}^{m} \delta_i e_i Y_{t,i}$ and deterministic exposures. The stylized CDO has k tranches, indexed by $\kappa \in \{1, \dots, k\}$ and characterized by attachment points $0 = K_0 < K_1 < \cdots < K_k \leqslant \sum_{i=1}^{m} e_i$. The value of the *notional* corresponding to tranche κ can be described as follows. Initially, the notional is equal to $K_\kappa - K_{\kappa-1}$; it is reduced whenever there is a default event such that the cumulative loss falls in the layer $[K_{\kappa-1}, K_\kappa]$. In mathematical terms, $N_{t,\kappa}$, the notional of tranche κ at time t, is given by

$$N_{t,\kappa} = N_\kappa(L_t) \quad \text{with } N_\kappa(l) = \begin{cases} K_\kappa - K_{\kappa-1} & \text{for } l < K_{\kappa-1}, \\ K_\kappa - l & \text{for } l \in [K_{\kappa-1}, K_\kappa], \\ 0 & \text{for } l > K_\kappa. \end{cases} \tag{12.1}$$

Note that $N_\kappa(l) = (K_\kappa - l)^+ - (K_{\kappa-1} - l)^+$, so $N_{t,\kappa}$ is equal to the sum of a long position in a put option on L_t with strike price K_κ and a short position in a put on L_t with strike price $K_{\kappa-1}$. Such positions are also known as put spreads.

We assume that in a stylized CDO the pay-off of tranche κ is equal to $N_{T,\kappa}$, the value of the tranche notional at the maturity date T. In Figure 12.2 we have graphed the pay-off for a stylized CDO with maturity $T = 5$ years and three tranches (equity, mezzanine, senior) on a homogeneous portfolio of $m = 1000$ firms, each with exposure one unit and loss given default $\delta_i = 0.5$. The attachment points are $K_1 = 20$, $K_2 = 40$, $K_3 = 60$, corresponding to 2%, 4% and 6% of the overall exposure; tranches with higher attachment points are ignored. We have plotted two distributions for L_T: first, a loss distribution corresponding to a five-year default probability of 5% and a five-year default correlation of 2%; second, a loss distribution with a five-year default probability of 5% but with independent defaults. In both cases the expected loss is given by $E(L_T) = 25$. Figure 12.2 illustrates how the value of different CDO tranches depends on the extent of the dependence between default events.

- For independent defaults, L_T is typically close to its mean due to diversification effects within the portfolio. It is therefore quite unlikely that a tranche κ with lower attachment point $K_{\kappa-1}$ substantially larger than $E(L_T)$ (such

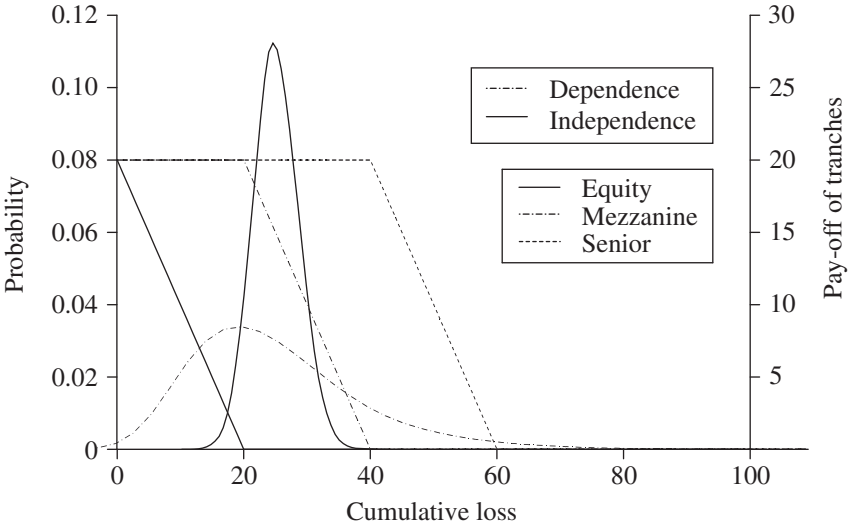

Figure 12.2. Pay-off of a stylized CDO contract and distribution of the five-year loss L_5 for a five-year default probability of 5% and different default correlations. Detailed explanations are given in the text.

as the senior tranche in Figure 12.2) suffers a loss, so the value of such a tranche is quite high. On the other hand, since the upper attachment point K_1 of the equity tranche is lower than $E(L_T) = 25$, it is quite unlikely that L_T is substantially smaller than K_1, and the value of the equity tranche is low.

- If defaults are dependent, diversification effects in the portfolio are less pronounced. Realizations of the loss L_T larger than the lower attachment point K_2 of the senior tranche are more likely, as are realizations of L_T smaller than the upper attachment point K_1 of the equity tranche. This reduces the value of tranches with high seniority and increases the value of the equity tranche compared with the case of independent defaults.
- The impact of changing default correlations on mezzanine tranches is unclear and cannot be predicted up front.

The relationship between default dependence and the value of CDO tranches carries over to the more complex structures that are actually traded, so dependence modelling is a key issue in any model for pricing CDO tranches (see also Section 12.3.2 below).

Pricing and the role of rating agencies. Before the financial crisis, rating agencies played a dominant role in the valuation of asset-backed CDOs. In fact, many CDO investors lacked the necessary sophistication and data to form an independent judgement of the riskiness of asset-backed CDOs, a problem that was compounded by the complex waterfall structure of most CDO issues. They therefore based their investment decision solely on the risk assessment of the rating agencies. This is particularly true for AAA-rated tranches, which appeared to be attractive investment

opportunities due to the relatively high offered yield (compared with the yield that could be earned on a standard AAA-rated bond, for example).

In relying on ratings, investors implicitly assumed that a high-quality rating such as AAA for a CDO tranche meant that the tranche had a similar risk profile to a AAA-rated bond. This perception is clearly wrong; since the loss distribution of an asset-backed CDO tranche is extremely sensitive with respect to the credit quality and the default correlation of the mortgages in the underlying asset pool, ratings for CDO tranches change rapidly with changes in these parameters. Moreover, while rating agencies have a lot of experience in rating corporate and sovereign debt, their experience with CDOs and other structured credit products was quite limited. As a result, ratings for CDO tranches turned out to be very unstable. In fact, at the onset of the crisis the rating of a large proportion of the traded ABS–CDOs (CDOs where the underlying asset pool consists of mortgage-based ABSs) was downgraded from investment grade to speculative grade, including default, within a very short period of time (Crouhy, Jarrow and Turnbull 2008). This massive rating change has sparked an intense debate about the appropriateness of rating methodologies, and about the role and the incentives of rating agencies more generally. We refer to Notes and Comments for further reading.

CDO-squared contracts. CDO-squared contracts are CDOs where the underlying asset pool itself consists of CDO tranches. These products are very complex and difficult, if not impossible, to value. For this reason they never became particularly popular on markets for synthetic CDOs. The situation is different in markets for asset-backed CDOs. Before the crisis there was intense trading activity in ABS–CDOs. The main reason for this was the fact that these products seemed to offer a way to create additional AAA-rated securities from the mezzanine tranches of the original ABSs, thus satisfying the high demand for AAA-rated securities. Investors in highly rated ABS–CDOs incurred severe losses during the credit crisis and, as shown by Hull and White (2010), the AAA rating carried by many of these structures was very hard to defend in retrospect. Many studies since the financial crisis have emphasized the need for simpler and more standardized financial products: see, for example, Crouhy, Jarrow and Turnbull (2008) and Hull (2009). ABS–CDOs are clearly a prime case in point.

12.1.2 Credit Indices and Index Derivatives

Credit index derivatives are standardized credit products whose pay-off is determined by the occurrence of credit events in a fixed pool of major firms that form the so-called credit index. A key requirement for the inclusion of a firm in a credit index is the existence of a liquid single-name CDS market in that firm. The availability of indices has helped to create a liquid market for certain credit index derivatives that has become a useful benchmark for model calibration and an important reference point for academic studies.

At present there are two major families of credit indices: the CDX family and the iTraxx family. CDX indices refer to American companies and iTraxx indices refer either to European firms or to Asian and Australian firms. Characteristics of

Table 12.1. Composition of main credit indices (taken from Couderc and Finger (2010)). The most important indices are the iTraxx Europe and CDX.NA.IG indices.

Name	Pool size	Region	Credit quality
CDX.NA.IG	125	North America	Investment grade
CDX.NA.IG.HVOL	30	North America	Low-quality investment grade
CDX.NA.HY	100	North America	Non-investment-grade
iTraxx Europe	125	Europe	Investment grade
iTraxx Europe	30	Europe	Low-quality investment grade

the main credit indices are given in Table 12.1. In order to reflect changes in the credit quality of the constituents, the composition of most credit indices changes every six months at the so-called roll dates (20 March and 20 September), and the pools corresponding to the roll dates are known as the different *series* of the index. Products on older series continue to trade but the market for products related to the current series is by far the most liquid.

Standardized index derivatives are credit index swaps and single-tranche CDOs with a standardized set of attachment points. The cash flow of these products bears some similarities to the cash flows of a single-name CDS as described in Sections 10.1.4 and 10.4.4. Each contract consists of a *premium payment leg* (payments made by the protection buyer) and a *default payment leg* (payments made by the protection seller). Premium payments are due at deterministic time points $0 < t_1 < \cdots < t_N = T$, where T is the maturity of the contract. Standardized index derivatives have quarterly premium payments, i.e. $t_n - t_{n-1} = 0.25$; the time to maturity at issuance is three, five, seven or ten years, with five-year products being the most liquid.

Next we describe the pay-offs of an index swap and a CDO tranche. We consider a fixed pool of m names ($m = 125$ for derivatives related to the iTraxx Europe and CDX.NA.IG indices) and we normalize the exposure of each firm to 1, so that the cumulative portfolio loss at time t equals $L_t = \sum_{i=1}^m \delta_i Y_{t,i}$.

Credit index swaps. At a default time $T_k \leqslant T$ there is a default payment of size δ_{ξ_k}, where $\xi_k \in \{1, \ldots, m\}$ is the identity of the name defaulting at T_k. The cumulative cash flows of the default payment leg up to time $t \leqslant T$ (received by the protection buyer) are therefore given by

$$\sum_{T_k \leqslant t} \delta_{\xi_k} = \sum_{\tau_i \leqslant t} \delta_i = \sum_{i=1}^m \delta_i Y_{t,i} = L_t.$$

Given an annualized swap spread x, the premium payment at time t_n (received by the protection seller) is given by

$$x(t_n - t_{n-1}) N_{t_n}^{\mathrm{Ind}},$$

where the *notional* N_t^{Ind} of the index swap is equal to the number of surviving firms at time t: that is, $N_t^{\mathrm{Ind}} = m - \sum_{i=1}^m Y_{t,i}$. This definition of the notional reflects the

Table 12.2. Standardized attachment points for single-tranche CDOs on the CDX.NA.IG and iTraxx Europe indices.

CDX.NA.IG	0–3%	3–7%	7–10%	10–15%	15–30%
iTraxx Europe	0–3%	3–6%	6–9%	9–12%	12–20%

fact that at a default time the defaulting entity is removed from the index. Moreover, at a default time $T_k \in (t_{n-1}, t_n]$, the protection buyer pays the protection seller the part of the premium that has accrued since the last regular premium payment date: that is, the quantity $x(T_k - t_{n-1})$. A credit event therefore has a double effect on the cash-flow structure of the index swap: it leads to a default payment and it reduces future premium payments.

Single-tranche CDOs. A single-tranche CDO on the reference portfolio is characterized by fixed lower and upper attachment points $0 \leqslant l < u \leqslant 1$, expressed as percentages of the overall notional m of the index pool. As in (12.1) we define the notional $N_t^{[l,u]}$ of the tranche by a put spread:

$$N_t^{[l,u]} = (um - L_t)^+ - (lm - L_t)^+. \tag{12.2}$$

In particular, for $L_0 = 0$ the initial notional of the tranche is equal to $m(u - l)$. The cumulative *tranche loss* up to time t is then given by

$$L_t^{[l,u]} := m(u - l) - N_t^{[l,u]} = (L_t - lm)^+ - (L_t - um)^+, \tag{12.3}$$

so the tranche loss can be viewed as a call spread on the cumulative portfolio loss. At a default time $T_k \leqslant T$ the protection seller makes a default payment of size

$$\Delta L_{T_k}^{[l,u]} := L_{T_k}^{[l,u]} - L_{T_k-}^{[l,u]}. \tag{12.4}$$

Again the premium payment leg consists of regular and accrued premium payments. Given an annualized tranche spread x, the regular premium payment at date t_n is given by $x(t_n - t_{n-1})N_{t_n}^{[l,u]}$. The accrued payment at a default time $T_k \in (t_{n-1}, t_n]$ equals $x(T_k - t_{n-1})\Delta L_{T_k}^{[l,u]}$. In order to simplify the exposition we will usually omit accrued premium payments below.

Single-tranche CDOs are sometimes called *synthetic*, as there is no physical transfer of credit-risky securities from the protection seller to the protection buyer, in contrast to asset-backed CDOs. There is a standardized set of attachment points for index tranches on the iTraxx Europe and CDX.NA.IG indices (see Table 12.2). In analogy with the terminology used for asset-backed CDOs, the tranche with lower attachment point $l = 0$ is known as the *equity tranche*; the equity tranche is clearly affected by the first losses in the underlying pool. The tranche with the highest attachment point is termed the *senior* tranche and the other tranches are known as *mezzanine* tranches of differing seniority. Tranches with non-standard maturity dates or attachment points and tranches on portfolios other than the constituents of a popular credit index are known as *bespoke* CDO tranches.

12.1.3 Basic Pricing Relationships for Index Swaps and CDOs

In this section we discuss some elementary model-independent pricing relationships for index swaps and CDOs that will be useful in this chapter. In our analysis we do not specify the underlying portfolio credit risk model in any detail; rather we take as given the joint distribution of the default times and of the LGDs calculated under some risk-neutral pricing measure Q. For simplicity we set the valuation date equal to $t = 0$. Moreover, we assume that the default-free interest rate $r(t)$ is deterministic, and we denote by

$$p_0(0,t) = \exp\left(-\int_0^t r(s)\,\mathrm{d}s\right), \quad t \geqslant 0,$$

the default-free discount factors or zero-coupon bond prices as seen from time $t = 0$. Deterministic interest rates are assumed in most of the literature on portfolio credit risk models, essentially because the additional complexity of stochastic interest rates is not warranted given the large amount of uncertainty surrounding the modelling of default dependence.

CDS index swaps. The market value V^{Def} of the default payments of a CDS index swap is given by the Q-expectation of the associated discounted cash-flow stream. The latter is given by $\sum_{T_k \leqslant T} p_0(0,T_k)\Delta L_{T_k}$, where $\Delta L_{T_k} = L_{T_k} - L_{T_k-} = \delta_{\xi_k}$. Since this sum may be written more succinctly as $\int_0^T p_0(0,t)\,\mathrm{d}L_t$ we get

$$V^{\mathrm{Def}} = E^Q\left(\int_0^T p_0(0,t)\,\mathrm{d}L_t\right) = \sum_{i=1}^m E^Q\left(\int_0^T p_0(0,t)\,\mathrm{d}L_{t,i}\right), \tag{12.5}$$

where $L_{t,i} = \delta_i Y_{t,i}$ is the cumulative loss process of firm i.

Given a generic spread x, the market value of the premium payments is given by $V^{\mathrm{Prem}}(x)$, where

$$\begin{aligned} V^{\mathrm{Prem}}(x) &= x\sum_{n=1}^N p_0(0,t_n)(t_n - t_{n-1})E^Q(N_t^{\mathrm{Ind}}) \\ &= x\sum_{i=1}^m\sum_{n=1}^N p_0(0,t_n)(t_n - t_{n-1})E^Q(1 - Y_{t_n,i}). \end{aligned} \tag{12.6}$$

Clearly, $V^{\mathrm{Prem}}(x) = xV^{\mathrm{Prem}}(1)$.

The market value at $t = 0$ of a protection buyer position in an index swap with given spread x is thus given by $V^{\mathrm{Def}} - V^{\mathrm{Prem}}(x)$. As in the case of single-name CDSs, the *fair index swap spread* x^{Ind} of the contract at a given point in time is set such that the market value of the contract at that date is 0. This leads to the formula

$$x^{\mathrm{Ind}} = \frac{V^{\mathrm{Def}}}{V^{\mathrm{Prem}}(1)}. \tag{12.7}$$

Next we consider the relationship between the index swap spread x^{Ind} and the fair CDS spread x^i for the single-name CDSs on the constituents of the index. It is tempting to conclude that the index swap spread is simply the arithmetic average of

the x^i, but this is wrong in general as the following analysis shows. For a single-name CDS on firm i with identical maturity and premium payment dates as the index swap, one has $V_i^{\mathrm{Def}} = V_i^{\mathrm{Prem}}(x^i)$ with

$$V_i^{\mathrm{Def}} = E^Q\bigg(\int_0^T p_0(0,t)\,\mathrm{d}L_{t,i}\bigg),$$
$$V_i^{\mathrm{Prem}}(x) = x\sum_{n=1}^N p_0(t,t_n)(t_n - t_{n-1})E^Q(1 - Y_{t_n,i}), \quad x > 0.$$

Comparing these relations with (12.5) and (12.6), we see that $V^{\mathrm{Def}} = \sum_{i=1}^m V_i^{\mathrm{Def}}$ and $V^{\mathrm{Prem}}(x) = \sum_{i=1}^m V_i^{\mathrm{Prem}}(x)$. For the fair index spread we therefore obtain

$$x^{\mathrm{Ind}} = \frac{V^{\mathrm{Def}}}{V^{\mathrm{Prem}}(1)} = \frac{\sum_{i=1}^m V_i^{\mathrm{Def}}}{\sum_{i=1}^m V_i^{\mathrm{Prem}}(1)} = \frac{\sum_{i=1}^m x^i V_i^{\mathrm{Prem}}(1)}{\sum_{i=1}^m V_i^{\mathrm{Prem}}(1)} =: \sum_{i=1}^m w_i x^i, \quad (12.8)$$

with weights given by $w_i := V_i^{\mathrm{Prem}}(1)/(\sum_{i=1}^m V_i^{\mathrm{Prem}}(1))$. The index spread is indeed therefore a weighted average of the single-name CDS spreads, but the weights are in general not equal to $1/m$. In fact, if firm i is of high credit quality and firm j of relatively low credit quality, one has

$$E^Q(1 - Y_{t,i}) = Q(\tau_i > t) > Q(\tau_j > t) = E^Q(1 - Y_{t,j}), \quad t > 0.$$

This implies that $V_i^{\mathrm{Prem}}(1) > V_j^{\mathrm{Prem}}(1)$ and hence that $w_i > w_j$, so that high-quality firms have a larger weight than low-quality firms. Of course, in the special case where all τ_i have the same distribution we get $w_1 = \cdots = w_m = 1/m$. An example is given by the simple model in which the default times are exponentially distributed with identical hazard rate γ^Q so that $Q(\tau_i > t) = \mathrm{e}^{-\gamma^Q t}$, and in which the LGD is deterministic and identical across firms. In that case we have $x^1 = \cdots = x^m = x^{\mathrm{ind}}$. Moreover, the parameter γ^Q can be calibrated from a market-observed index spread x^* using the same procedure as in the case of single-name CDS spreads (see Section 10.4.4). This setup is frequently employed by practitioners in the computation of implied correlations for single-tranche CDOs.

Single-tranche CDOs. Finally, we provide a more explicit description for the value of a single-tranche CDO. We begin with the premium payments. According to the definition of the tranche notional in (12.2), the market value of the regular premium payments for a generic CDO spread x is given by $V^{\mathrm{Prem}}(x)$, where

$$V^{\mathrm{Prem}}(x) = x\sum_{n=1}^N p_0(0,t_n)(t_n - t_{n-1})E^Q((um - L_{t_n})^+ - (lm - L_{t_n})^+). \quad (12.9)$$

Concerning the default payments we note from (12.4) that the discounted cash-flow stream of the default payment leg is given by

$$V^{\mathrm{Def}} = \sum_{T_k \leqslant T} p_0(0,T_k)\Delta L_{T_k}^{[l,u]} = \int_0^T p_0(0,t)\,\mathrm{d}L_t^{[l,u]}. \quad (12.10)$$

The integral on the right can be approximated by a Riemann sum, using the premium payment dates as gridpoints:

$$\int_0^T p_0(0,t)\,\mathrm{d}L_t^{[l,u]} \approx \sum_{n=1}^{N} p_0(0,t_n)(L_{t_n}^{[l,u]} - L_{t_{n-1}}^{[l,u]}). \tag{12.11}$$

In economic terms this approximation means that losses occurring during the period $[t_{n-1}, t_n]$ are paid only at time t_n (and are therefore discounted with a slightly higher factor). Since in practice premium payments typically occur quarterly, the error of this approximation is negligible. Recall that $L_t^{[l,u]}$ is a function of L_t, namely $L_t^{[l,u]} = v^{[l,u]}(L_t) := (L_t - lm)^+ - (L_t - um)^+$. Hence, with a slight abuse of notation we have

$$V^{\mathrm{Def}} = \sum_{n=1}^{N} p_0(0,t_n)(E^Q(v^{[l,u]}(L_{t_n})) - E^Q(v^{[l,u]}(L_{t_{n-1}}))). \tag{12.12}$$

Summarizing, we find that the evaluation of (12.12) and (12.9), and hence the determination of CDO spreads, reduces to computing call or put option prices on the cumulative loss process at the premium payment dates $t_1, \ldots, t_N$.

Notes and Comments

There are many contributions that discuss the pros and cons of securitization in the light of the subprime credit crisis of 2007 and 2008. Excellent descriptions of the events surrounding the crisis—including discussions of steps that should be implemented to prevent a repeat of it—are given by Crouhy, Jarrow and Turnbull (2008), Hull (2009) and, in an insurance context, Donnelly and Embrechts (2010) (see also Das, Embrechts and Fasen 2013). Hull and White (2010) test the ratings given to ABSs and ABS–CDOs before the crisis. They find that, whereas the AAA ratings assigned to ABSs were not unreasonable, the AAA ratings assigned to tranches of CDOs created from mezzanine tranches of ABSs cannot be justified by any proper quantitative analysis.

In his analysis of the credit crisis, Brunnermeier (2009) is particularly concerned with the various transmission mechanisms that caused losses in the relatively small American subprime mortgages market to be amplified in such a way that they created a global financial crisis. An interesting discussion of securitization in the light of the subprime crisis from a regulatory viewpoint is the well-known Turner Review (Lord Turner 2009). Incentive problems in the securitization of mortgages are discussed in Franke and Kahnen (2009). A more technical analysis of the value of securitization as a risk-management tool can be found in Frey and Seydel (2010). From the multitude of books we single out Dewatripont, Rochet and Tirole (2010) and Shin (2010). We also suggest that the interested reader look at the various documents published by the Bank for International Settlements and the Basel Committee on Banking Supervision (see www.bis.org/bcbs).

For alternative textbook treatments of CDOs and related index derivatives we refer to Bluhm and Overbeck (2007) and Brigo, Pallavicini and Torresetti (2010).

Both texts contain a wealth of institutional details on CDO markets. The book by Bluhm and Overbeck (2007) also discusses so-called *basket default swaps*, or, more technically, kth-to-default swaps. These products offer protection against the kth default in a portfolio with $m > k$ obligors (the basket). As in the case of an ordinary CDS, the premium payments on a kth-to-default swap take the form of a periodic payment stream, which stops at the kth default time T_k. The default payment is triggered if T_k is smaller than the maturity date of the swap. While first-to-default swaps are fairly common, higher-order default swaps are encountered only rarely.

12.2 Copula Models

Copula models are widely used in practice for the pricing of CDO tranches and basket credit derivatives. In this section we discuss this important class of models with a particular focus on models where the copula has a factor structure.

12.2.1 Definition and Properties

Definition 12.1 (copula model for default times). Let C be a copula and let $\gamma_i(t)$, $1 \leqslant i \leqslant m$, be nonnegative functions such that $\Gamma_i(t) := \int_0^t \gamma_i(s)\,\mathrm{d}s < \infty$ for all $t > 0$ and $\lim_{t\to\infty} \Gamma_i(t) = \infty$. The default times $\tau_1, \ldots, \tau_m$ then follow a copula model with *survival copula* C and *marginal hazard functions* $\gamma_i(t)$ if their joint survival function can be written in the form

$$\bar{F}(t_1, \ldots, t_m) = C(\mathrm{e}^{-\Gamma_1(t_1)}, \ldots, \mathrm{e}^{-\Gamma_m(t_m)}). \tag{12.13}$$

The marginal survival functions in a copula model for defaults are obviously given by $\bar{F}_i(t) = \mathrm{e}^{-\Gamma_i(t)}$. The marginal survival functions and marginal dfs are continuous, and it follows from Sklar's Theorem that both the copula and the survival copula C of the vector of default times $\boldsymbol{\tau} := (\tau_1, \ldots, \tau_m)'$ are unique. Of course, the joint distribution of the default times could also be described in terms of the copula of $\boldsymbol{\tau}$ and the marginal distribution functions $F_1, \ldots, F_m$. We focus on survival copulas as this ties in with a large part of the literature. If C in (12.13) is radially symmetric (see Definition 7.15), then C is also the copula of $\boldsymbol{\tau}$. This is true, in particular, if C is an elliptical copula, such as the Gauss copula or the t copula.

From a mathematical point of view it makes no difference whether we specify the survival copula and the marginal hazard functions separately or whether we specify the joint survival function $\bar{F}$ directly; every joint survival function $\bar{F}$ with absolutely continuous marginal distributions has a unique representation of the form (12.13). However, the representation (12.13) is convenient for the *calibration* of the model to prices of traded credit derivatives. The usual calibration process is carried out in two steps, as we now explain.

In the first step, the marginal hazard functions are calibrated to a given term structure of CDS spreads or spreads of defaultable bonds, as described in Section 10.4.4. In the second step, the parameters of the survival copula C are calibrated to the observed prices of traded portfolio credit derivatives, most notably CDO index tranches. The two-stage calibration is feasible because the second step has no effect

on the parameters calibrated in the first step. This is very advantageous from a numerical point of view, which is one of the reasons for the popularity of copula models among practitioners. We return to the calibration of copula models in Section 12.3.

In order to link our analysis to the discussion of copulas in threshold models in Section 11.1, we take a brief look at the one-period portfolio model that is implied by a copula model for the default times $\tau_1, \dots, \tau_m$. We fix a horizon T and set $Y_i := Y_{T,i}$. By definition, $Y_i = 1$ if and only if $\tau_i \leqslant T$, so the one-period model has the threshold representation $(\boldsymbol{\tau}, (T, \dots, T)')$. The dependence of the default events in the one-period model is governed by the dependence structure of the critical variables $\tau_1, \dots, \tau_m$. Note that in Section 11.1 this dependence was described in terms of the copula of the critical variables, whereas in this section we prefer to work with the survival copula of the default times.

Definition 12.1 immediately leads to an explicit construction of the random times $\tau_1, \dots, \tau_m$ via random thresholds; this in turn yields a generic simulation method for copula models. Consider a random vector $\boldsymbol{U} \sim C$ and define the default times by

$$\tau_i = \inf\{t \geqslant 0 \colon \mathrm{e}^{-\Gamma_i(t)} \leqslant U_i\}, \quad 1 \leqslant i \leqslant m, \tag{12.14}$$

so that firm i defaults at the first time point where the marginal survival function $\bar{F}_i(t)$ crosses the random threshold U_i. Equation (12.14) yields

$$\begin{aligned} P(\tau_1 > t_1, \dots, \tau_m > t_m) &= P(U_1 \leqslant \mathrm{e}^{-\Gamma_1(t_1)}, \dots, U_m \leqslant \mathrm{e}^{-\Gamma_m(t_m)}) \\ &= C(\mathrm{e}^{-\Gamma_1(t)}, \dots, \mathrm{e}^{-\Gamma_m(t)}), \end{aligned}$$

as required. To generate a realization of $\boldsymbol{\tau}$ we generate a realization of the random vector $\boldsymbol{U}$ and construct the τ_i according to (12.14). An alternative simulation algorithm for factor copula models is given at the end of this section.

Factor copula models. Most copula models used in practice have a factor structure. By this we mean that the threshold vector $\boldsymbol{U} \sim C$ used in (12.14) has a conditional independence structure in the sense of Definition 11.9, i.e. there is a p-dimensional random vector $\boldsymbol{V}$, $p < m$, such that the U_i are independent given $\boldsymbol{V}$.

The conditional independence of the U_i allows for an alternative representation of the joint survival function of the default times in terms of a mixture model. By conditioning on $\boldsymbol{V}$, it follows from (12.14) that we can write the joint survival function of $\boldsymbol{\tau}$ as

$$\begin{aligned} \bar{F}(t_1, \dots, t_m) &= E(P(U_1 \leqslant \bar{F}_1(t_1), \dots, U_m \leqslant \bar{F}_m(t_m) \mid \boldsymbol{V})) \\ &= E\biggl(\prod_{i=1}^{m} P(U_i \leqslant \bar{F}_i(t_i) \;\Big|\; \boldsymbol{V})\biggr). \end{aligned} \tag{12.15}$$

Moreover, by the definition of τ_i in (12.14) we have $P(\tau_i > t \mid \boldsymbol{V}) = P(U_i \leqslant \bar{F}_i(t) \mid \boldsymbol{V})$. The conditional survival function of τ_i given $\boldsymbol{V} = \boldsymbol{v}$ therefore satisfies

$$\bar{F}_{\tau_i|\boldsymbol{V}}(t \mid \boldsymbol{v}) = P(U_i \leqslant \bar{F}_i(t) \mid \boldsymbol{V} = \boldsymbol{v}),$$

and we may write $\bar{F}(t_1, \ldots, t_m) = E(\prod_{i=1}^m \bar{F}_{\tau_i|\boldsymbol{V}}(t_i \mid \boldsymbol{V}))$. Denoting the density or probability mass function of $\boldsymbol{V}$ by $g_{\boldsymbol{V}}$, we will usually write the joint survival function in the form

$$\bar{F}(t_1, \ldots, t_m) = \int_{\mathbb{R}^p} \prod_{i=1}^m \bar{F}_{\tau_i|\boldsymbol{V}}(t_i \mid \boldsymbol{v}) g_{\boldsymbol{V}}(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v}. \tag{12.16}$$

Note that the representation (12.16) is analogous to the representation of one-period threshold models with conditional independence structure as Bernoulli mixture models (see Section 11.2.4). In particular, (12.16) shows that for T fixed the default indicators follow a Bernoulli mixture model with factor vector $\boldsymbol{V}$ and conditional default probabilities $p_i(\boldsymbol{v}) = 1 - \bar{F}_{\tau_i|\boldsymbol{V}}(t \mid \boldsymbol{v})$.

The mixture-model representation of a factor copula model gives rise to the following generic algorithm for the simulation of default times.

Algorithm 12.2 (sampling from factor copula models).

(1) Generate a realization of $\boldsymbol{V}$.

(2) Generate independent rvs τ_i with df $1 - \bar{F}_{\tau_i|\boldsymbol{V}}(t \mid \boldsymbol{V})$, $1 \leqslant i \leqslant m$.

The importance-sampling techniques discussed in Section 11.4 in the context of one-period Bernoulli mixture models can be applied to improve the performance of Algorithm 12.2. These techniques are particularly useful if one is interested in rare events, such as in the pricing of CDO tranches with high attachment points.

12.2.2 Examples

In this section specific examples of factor copula models will be discussed. Any random vector with continuous marginal distributions and a p-dimensional conditional independence structure (see Definition 11.9) can be used to construct a factor copula model. Important examples in practice include factor copulas models based on the Gauss copula C_P^{Ga}, the LT-Archimedean copulas discussed in Section 7.4.2, and general one-factor copulas including the so-called double-t and double-NIG copulas.

Example 12.3 (one-factor Gauss copula). Factor copula models based on a Gauss copula C_P^{Ga} are frequently used in practice. We will compute the joint survival function for the default times in the one-factor case. Let

$$X_i = \sqrt{\rho_i} V + \sqrt{1 - \rho_i} \varepsilon_i, \tag{12.17}$$

where $\rho_i \in (0, 1)$ and where V and $(\varepsilon_i)_{1 \leqslant i \leqslant m}$ are iid standard normal rvs. The random vector $\boldsymbol{X} = (X_1, \ldots, X_m)'$ satisfies $\boldsymbol{X} \sim N_m(\boldsymbol{0}, P)$, where the (i, j)th element of P is given by $\rho_{ij} = \sqrt{\rho_i \rho_j}$. We set $U_i = \Phi(X_i)$ so that $\boldsymbol{U} = (U_1, \ldots, U_m)' \sim C_P^{\mathrm{Ga}}$. Both $\boldsymbol{X}$ and $\boldsymbol{U}$ have a one-factor conditional independence structure.

The conditional survival function of τ_i is easy to compute. Writing $d_i(t) := \Phi^{-1}(\bar{F}_i(t))$, we have that

$$\bar{F}_{\tau_i|V}(t \mid v) = P(U_i \leqslant \bar{F}_i(t) \mid V = v) = P\left(\varepsilon_i \leqslant \frac{d_i(t) - \sqrt{\rho_i} V}{\sqrt{1 - \rho_i}} \;\middle|\; V = v\right),$$

leading to $\bar{F}_{\tau_i|V}(t \mid v) = \Phi((d_i(t) - \sqrt{\rho_i}v)/(\sqrt{1-\rho_i}))$. Hence

$$\bar{F}(t_1, \ldots, t_m) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \prod_{i=1}^{m} \Phi\left(\frac{d_i(t_i) - \sqrt{\rho_i}v}{\sqrt{1-\rho_i}}\right) \mathrm{e}^{-v^2/2}\, \mathrm{d}v, \tag{12.18}$$

which is easily computed using one-dimensional numerical integration.

In applications of a one-factor Gauss copula model to the pricing of portfolio credit derivatives it is frequently assumed that $\rho_i = \rho$ for all i (the equicorrelation case). This model is known as the *exchangeable Gauss copula model*. In this case $\rho = \mathrm{corr}(X_i, X_j)$, so ρ is readily interpreted in terms of asset correlation. This makes the exchangeable Gauss copula model popular with practitioners. In fact, it is common practice on CDO markets to quote prices for tranches of synthetic CDOs in terms of *implied asset correlations* computed in an exchangeable Gauss copula model, as will be discussed in detail in Section 12.3.2.

Example 12.4 (general one-factor copula). This class of examples generalizes the construction of the one-factor Gauss copula in Example 12.3; some of the resulting models are useful in explaining prices of CDO tranches on credit indices. As in (12.17), one starts with random variables $X_i = \sqrt{\rho_i}V + \sqrt{1-\rho_i}\varepsilon_i$ for $\rho_i \in [0, 1]$ and *independent* rvs $V, \varepsilon_1, \ldots, \varepsilon_m$, but now V and the ε_i can have arbitrary continuous distributions (not necessary normal). The distribution F_{X_i} is then given by the convolution of the distributions of $\sqrt{\rho_i}V$ and of $\sqrt{1-\rho_i}\varepsilon_i$. The corresponding copula is the df of the random vector $\boldsymbol{U}$ with $U_i = F_{X_i}(X_i)$. A similar calculation to that in the case of the one-factor Gauss copula gives the following result for the conditional survival function:

$$\bar{F}_{\tau_i|V}(t \mid v) = F_\varepsilon\left(\frac{F_{X_i}^{-1}(\bar{F}_i(t)) - \sqrt{\rho_i}v}{\sqrt{1-\rho_i}}\right).$$

Popular examples include the so-called *double-t copula* (the case where V and ε_i follow a univariate t_ν distribution) and the *double-GH copula* (the case where V and ε_i both follow a univariate GH distribution as introduced in Section 6.2.3). We note that the double-t copula is not the same as the usual t copula since the rvs $V, \varepsilon_1, \ldots, \varepsilon_m$ do not have a multivariate t distribution (recall that uncorrelated bivariate t-distributed rvs are *not* independent, whereas V and ε_i are independent); the situation is similar for the double-GH copula.

The main computational challenge in working with the double-t and double-GH copulas is the determination of the convolution of V and ε_i and the corresponding quantile function $F_{X_i}^{-1}$. There exist a number of good solutions to this problem; references are given in Notes and Comments.

Example 12.5 (LT-Archimedean copulas). Consider a positive rv V with df G_V such that $G_V(0) = 0$ and denote by $\hat{G}_V$ the Laplace–Stieltjes transform of G_V. According to Definition 7.52 the associated LT-Archimedean copula is given by the formula

$$C(u_1, \ldots, u_m) = E\left(\exp\left(-V \sum_{i=1}^{m} \hat{G}_V^{-1}(u_i)\right)\right).$$

As usual, denote by $\bar{F}_i(t_i)$ the marginal survival function of τ_i. Using (12.13) the joint survival function of $\boldsymbol{\tau}$ can be calculated to be

$$\bar{F}(t_1, \ldots, t_m) = E\left(\prod_{i=1}^{m} \exp\{-V \hat{G}_V^{-1}(\bar{F}_i(t_i))\} \right), \tag{12.19}$$

which is obviously of the general form (12.16). Recall that in the special case of the Clayton copula with parameter θ we have $V \sim \mathrm{Ga}(1/\theta, 1)$; explicit formulas for $\hat{G}_V$ and $\hat{G}_V^{-1}$ in that case are given in Algorithm 7.53 for the simulation of LT-Archimedean copulas. Note that LT-Archimedean copulas are in general not radially symmetric, so the one-period version of the copula model for default times with survival function (12.19) is *not* the same as the threshold model with Archimedean copula presented in Example 11.4; it corresponds instead to a threshold model with Archimedean survival copula.

Notes and Comments

The first copula model for portfolio credit risk was given by Li (2000); his model is based on the Gauss copula. General copula models were introduced for the first time in Schönbucher and Schubert (2001). One of the first papers on factor copula models was Laurent and Gregory (2005). These models have received a lot of attention, mostly in connection with implied correlation skews (see also Section 12.3 below). The double-t copula was proposed by Hull and White (2004); numerical aspects of this copula model are discussed in Vrins (2009). Double-GH copulas are considered by Kalemanova, Schmid and Werner (2005), Guégan and Houdain (2005) and Eberlein, Frey and von Hammerstein (2008), among others.

12.3 Pricing of Index Derivatives in Factor Copula Models

In this section we are interested in the pricing of index derivatives (index swaps and CDOs) in factor copula models. These models are the market standard for dealing with CDO tranches. We begin with analytical and numerical methods for the valuation of index derivatives. In Section 12.3.2 we turn to qualitative properties of observed CDO spreads and discuss the well-known correlation skews. In Section 12.3.3 we consider an alternative factor model, known as an implied copula model, that has been developed in order to explain correlation skews.

12.3.1 Analytics

Throughout we consider a factor copula model with factor-dependent recovery risk. More precisely, we assume that under the risk-neutral pricing measure Q the survival function of the default times follows a factor copula model with mixture representation (12.16). Moreover, the loss given default of firm i is state dependent, so, given a realization of the factor $\boldsymbol{V} = \boldsymbol{v}$, the loss given default is of the form $\delta_i(\boldsymbol{v})$ for some function $\delta_i : \mathbb{R}^p \to (0, 1)$. A state-dependent loss given default provides extra flexibility for calibrating the model. Consider, for example, the case where V is a one-dimensional rv and where the conditional default probabilities are increasing in

V. By taking increasing functions $\delta_i(v)$ it is possible to model negative correlation between default probabilities and recovery rates. This extension is reasonable from an empirical viewpoint, as was discussed in Section 11.2.3.

Index swaps. We begin by computing the value of the default and the premium payment leg for index swaps. Recall that $V^{\mathrm{Def}}(x) = \sum_{i=1}^m V_i^{\mathrm{Def}}$ and $V^{\mathrm{Prem}}(x) = \sum_{i=1}^m V_i^{\mathrm{Prem}}(x)$, where V_i^{Def} and $V_i^{\mathrm{Prem}}(x)$ represent the value of the default and of the premium payment leg of the single-name CDS on obligor i. Moreover, V_i^{Def} and $V_i^{\mathrm{Prem}}(x)$ depend only on the loss process of firm i, which is given by $L_{t,i} = \delta_i Y_{t,i}$. The value of an index swap does not therefore actually depend on the copula of the default times; this is in contrast to the single-tranche CDO analysed below.

We now turn to the actual computations. If the recovery rates are constant, V_i^{Def} and $V_i^{\mathrm{Prem}}(x)$ can be computed using the results on the pricing of single-name CDSs in hazard rate models discussed in Section 10.4.4. In the general case one may proceed as follows. First, similar reasoning to that applied in (12.11) gives

$$V^{\mathrm{Def}} = E^Q\bigg(\int_0^T p_0(0,t)\,\mathrm{d}L_t\bigg) \approx \sum_{n=1}^N p_0(0,t_n)(E^Q(L_{t_n}) - E^Q(L_{t_{n-1}})). \tag{12.20}$$

Since $L_{t_n} = \sum_{i=1}^m \delta_i Y_{t_n,i}$, we find, by conditioning on $\boldsymbol{V}$, that

$$E^Q(L_{t_n}) = \int_{\mathbb{R}^p} \sum_{i=1}^m \delta_i(\boldsymbol{v}) E^Q(Y_{t_n,i} \mid \boldsymbol{V} = \boldsymbol{v}) g_{\boldsymbol{V}}(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}.$$

Now $Y_{t_n,i}$ is Bernoulli with $Q(Y_{t_n,i} = 1 \mid \boldsymbol{V} = \boldsymbol{v}) = 1 - \bar{F}_{\tau_i|\boldsymbol{V}}(t_n \mid \boldsymbol{v})$. Hence,

$$E^Q(L_{t_n}) = \int_{\mathbb{R}^p} \sum_{i=1}^m \delta_i(\boldsymbol{v})(1 - \bar{F}_{\tau_i|\boldsymbol{V}}(t_n \mid \boldsymbol{v})) g_{\boldsymbol{V}}(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v},$$

and substitution of this expression into (12.20) gives the value of the default payment leg. Similar reasoning allows us to conclude that $V^{\mathrm{Prem}}(x)$, the value of the premium payments for a generic spread x, is equal to

$$V^{\mathrm{Prem}}(x) = x \sum_{n=1}^N p_0(0,t_n) \int_{\mathbb{R}^p} \sum_{i=1}^m \bar{F}_{\tau_i|\boldsymbol{V}}(t_n \mid \boldsymbol{v}) g_{\boldsymbol{V}}(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}. \tag{12.21}$$

In both cases the integration over $\boldsymbol{v}$ can be carried out with numerical quadrature methods. This is straightforward in the cases that are most relevant in practice, where V is a one-dimensional rv.

CDO tranches. Recall from Section 12.1.3 that the essential task in computing CDO spreads is to determine the price of call or put options on L_{t_n}. By a conditioning argument we obtain, for a generic function $f\colon \mathbb{R}^+ \to \mathbb{R}$ such as the pay-off of a call or put option, the formula

$$E^Q(f(L_{t_n})) = \int_{\mathbb{R}^p} E^Q\bigg(f\bigg(\sum_{i=1}^m \delta_i(\boldsymbol{v}) Y_{t_n,i} \;\bigg|\; \boldsymbol{V} = \boldsymbol{v}\bigg)\bigg) g_{\boldsymbol{V}}(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}.$$

Table 12.3. Fair tranche spreads computed in an exchangeable Gauss copula model using the LPA and the simple normal approximation. The exact spread value is computed by an extensive Monte Carlo approximation. The row labelled "asset correlation" reports the value of the correlation parameter that is used in the pricing of the tranche. These correlations are so-called implied tranche correlations (see Table 12.4). The normal approximation is a substantial improvement over the LPA for the tranches where asset correlation is low, such as the [3, 6] and [6, 9] tranches; for the other tranches both approximation methods perform reasonably well. The reason for this is the fact that for high asset correlation the loss distribution is essentially determined by the realization of V, whereas for low asset correlation the form of the conditional distribution of L_t matters. Numerical values are taken from Frey, Popp and Weber (2008).

Tranche	[0, 3]	[3, 6]	[6, 9]	[9, 12]	[12, 22]
Asset correlation	$\rho = 0.219$	$\rho = 0.042$	$\rho = 0.148$	$\rho = 0.223$	$\rho = 0.305$
LPA (%)	30.66	0.79	0.53	0.36	0.18
Normal approximation (%)	29.38	1.51	0.66	0.42	0.20
Exact spread (%)	28.38	1.55	0.68	0.42	0.20

To discuss this formula we fix t_n and introduce the simpler notation $q_i(\boldsymbol{v}) := Q(Y_{t_n,i} = 1 \mid \boldsymbol{V} = \boldsymbol{v})$. Since the rvs $Y_{t_n,i}$, $1 \leqslant i \leqslant m$, are conditionally independent given $\boldsymbol{V}$, computing the inner conditional expectation essentially amounts to finding the distribution of a sum of independent Bernoulli rvs. This is fairly straightforward for a homogeneous portfolio, where the LGD functions and the conditional default probabilities for all firms are equal, so that $q_i(\boldsymbol{v}) \equiv q(\boldsymbol{v})$ and $\delta_i(\boldsymbol{v}) \equiv \delta(\boldsymbol{v})$. In that case, $L_{t_n} = \delta(V)M_{t_n}$, where $M_t = \sum_{i=1}^m Y_{t,i}$ gives the number of firms that have defaulted up to time t. Given $\boldsymbol{V} = \boldsymbol{v}$, M_{t_n} is the sum of independent and (due to the homogeneity) identically distributed Bernoulli rvs, so M_{t_n} has a binomial distribution with parameters m and $q(\boldsymbol{v})$.

In a heterogeneous portfolio, on the other hand, the evaluation of the conditional loss distribution is not straightforward, essentially because one is dealing with sums of independent Bernoulli rvs that are *not* identically distributed. We discuss several approaches to this problem in the remainder of this section.

Large-portfolio approximation. In the so-called large-portfolio approximation (LPA), the conditional distribution of L_{t_n} given $\boldsymbol{V} = \boldsymbol{v}$ is replaced by a point mass at the conditional mean $\ell(\boldsymbol{v}) = \sum_{i=1}^m \delta_i(\boldsymbol{v})q_i(\boldsymbol{v})$. This leads to the approximation

$$E^Q(f(L_{t_n})) \approx \int_{\mathbb{R}^p} f(\ell(\boldsymbol{v}))g_V(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}. \tag{12.22}$$

The method is motivated by the asymptotic results of Section 11.3. In particular, in that section we have shown that, in the case of one-factor models, quantiles of the portfolio loss distribution can be well approximated by quantiles of the rv $\ell(V)$, provided that the portfolio is sufficiently large and not too inhomogeneous. One would therefore expect that for typical credit portfolios the approximation (12.22) performs reasonably well. Numerical results on the performance of the LPA are given in Table 12.3.

Normal and Poisson approximation. The LPA completely ignores the impact of the fluctuations of L_{t_n} around its conditional mean. One way of capturing these fluctuations is to replace the unknown conditional distribution of L_{t_n} with a known distribution with similar moments. In a simple normal approximation the conditional distribution of L_{t_n} is approximated by a normal distribution with mean $\ell(\boldsymbol{v})$ and variance

$$\sigma^2(\boldsymbol{v}) := \operatorname{var}(L_{t_n} \mid \boldsymbol{V} = \boldsymbol{v}) = \sum_{i=1}^{m} \delta_i^2(\boldsymbol{v}) q_i(\boldsymbol{v})(1 - q_i(\boldsymbol{v}))$$

(the choice of the normal distribution is of course motivated by the central limit theorem). Under the normal approximation one has

$$E^Q(f(L_{t_n})) \approx \int_{\mathbb{R}^p} \frac{1}{\sqrt{2\pi\sigma^2(\boldsymbol{v})}} \int_{\mathbb{R}} f(l) \exp\left(-\frac{(l - \ell(\boldsymbol{v}))^2}{2\sigma^2(\boldsymbol{v})}\right) \mathrm{d}l \, g_V(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v}.$$

The performances of the normal approximation and the LPA are illustrated in Table 12.3 for the case of an exchangeable Gauss copula model.

In a similar spirit, in a Poisson approximation the conditional loss distribution is approximated by a Poisson distribution with parameter $\lambda = \ell(\boldsymbol{v})$. These simple normal and Poisson approximation methods can be refined substantially (see El Karoui, Kurtz and Jiao (2008) for details). There are a number of other techniques for evaluating the conditional loss distribution. On the one hand, one may apply Fourier or Laplace inversion to the problem; on the other hand, there are a number of recursive techniques that can be used. References for both approaches can be found in Notes and Comments.

12.3.2 Correlation Skews

Investors typically express CDO spreads in terms of implied correlations computed in a simple homogeneous Gauss copula model known as the *benchmark model*. In this way tranche spreads can be compared across attachment points and maturities. This practice is similar to the use of implied volatilities as a common yardstick for comparing prices of equity and currency options. In this section we explain this quoting convention in more detail and discuss qualitative properties of market-observed implied correlations.

The benchmark model is an exchangeable Gauss copula model, as in Example 12.3. It is assumed that all default times are exponentially distributed with $Q(\tau_i > t) = \mathrm{e}^{-\gamma^Q t}$ for all i and that the LGDs are deterministic and equal to 60% for all firms. The hazard rate γ^Q is used to calibrate the model to the spread of the index swap with the same maturity as the tranche under consideration; see the discussion of calibration of a homogeneous model to index swaps in Section 12.1.3. The only free parameter in the benchmark model is the "asset correlation" ρ. This parameter can be used to calibrate the model to tranche spreads observed in the market, which leads to various notions of implied correlation.

Implied correlations. The market uses two notions of implied correlation in order to describe market-observed tranche spreads. *Implied tranche correlation*, also known as *compound correlation*, is the lowest number $\rho \in [0, 1]$ such that the spread computed in the benchmark model equals the spread observed in the market. For mezzanine tranches the notion of implied tranche correlation suffers from two problems. First, a correlation number ρ such that the model spread matches the market spread does not always exist; this problem was encountered frequently for the spreads that were quoted in the period 2008–9 (during the financial crisis). Second, for certain values of the market spread there is more than one matching value of $\rho \in [0, 1]$, and the convention to quote the lowest such number is arbitrary.

If market spreads are available for a complete set of tranches, as is the case for the standardized tranches on iTraxx and CDX, the so-called *base correlation* is frequently used as an alternative means of expressing tranche spreads. To compute base correlation we recursively compute implied correlations for a nested set of equity tranches so that they are consistent with the observed tranche spreads. The base correlation methodology is popular since it mitigates the calibration problems arising in the computation of implied tranche correlation. On the other hand, base correlations are more difficult to interpret and there may be theoretical inconsistencies, as we explain below. In the following algorithm the computation of base correlation is described in more detail.

Algorithm 12.6. Suppose that we observe market spreads x_κ^*, $\kappa = 1, \dots, K$, for a set of CDO tranches with attachment points (l_κ, u_κ), $\kappa = 1, \dots, k$, with $l_1 = 0$ and $l_{\kappa+1} = u_\kappa$. Denote by $V^{\mathrm{Prem}}(u, \rho, x)$ the value of the premium payment leg of an equity tranche with attachment point u and spread $x > 0$ in the benchmark model with correlation parameter $\rho \in [0, 1]$. Denote by $V^{\mathrm{Def}}(u, \rho)$ the value of the corresponding default payment leg. We carry out the following steps.

(1) Compute the base correlation ρ_1 as the implied tranche correlation of the $[0, u_1]$ tranche using the observed spread x_1. Set $\kappa = 1$.

(2) Given the base correlation ρ_κ and the spread $x_{\kappa+1}^*$ (the market spread of the $[l_{\kappa+1}, u_{\kappa+1}]$ tranche), compute the base correlation $\rho_{\kappa+1}$ as the solution of the following equation (an explanation is given after the algorithm):

$$V^{\mathrm{Def}}(u_{\kappa+1}, \rho_{\kappa+1}) - V^{\mathrm{Def}}(u_\kappa, \rho_\kappa) = V^{\mathrm{Prem}}(u_{\kappa+1}, \rho_{\kappa+1}, x_{\kappa+1}^*) - V^{\mathrm{Prem}}(u_\kappa, \rho_\kappa, x_{\kappa+1}^*). \quad (12.23)$$

(3) Replace κ with $\kappa + 1$ and return to step (2).

Equation (12.23) is derived from the following considerations. It is the basic premise of the base correlation approach that for all $1 \leqslant \kappa \leqslant K$ and for any spread x, the correct values of the default and premium legs of an equity tranche with upper attachment point u_κ are given by $V^{\mathrm{Def}}(u_\kappa, \rho_\kappa)$ and $V^{\mathrm{Prem}}(u_\kappa, \rho_\kappa, x)$, respectively. Under this assumption, the left-hand side of (12.23) gives the value of the default payment leg of the $[l_{\kappa+1}, u_{\kappa+1}]$ tranche and the right-hand side gives the value of the corresponding premium payment leg at the spread $x_{\kappa+1}^*$ observed in the market.

For the observed spread $x^*_{\kappa+1}$, the value of the default and of the premium payment leg of the $[l_{\kappa+1}, u_{\kappa+1}]$ tranche have to be equal, which leads to equation (12.23) for $\rho_{\kappa+1}$.

Note that, under the base-correlation methodology, two different models are used to derive the model value of the $[l_\kappa, u_\kappa]$ tranche: the benchmark model with correlation parameter ρ_κ is used for the upper attachment point and the benchmark model with correlation parameter $\rho_{\kappa-1}$ is used for the lower attachment point. As shown in Brigo, Pallavicini and Torresetti (2010), this may lead to inconsistencies if the base correlation approach is employed in the pricing of tranches with non-standard attachment points.

Empirical properties of implied correlations. In Table 12.4 we give market spreads and implied tranche correlations for iTraxx tranches for three different days in 2004, 2006 and 2008. These numbers are representative of three different periods in the credit index market: the early days of the market; before the credit crisis; and during the credit crisis. We see that implied correlations are quite unstable. In particular, a standard Gauss copula with a fixed correlation parameter ρ cannot explain all tranche spreads simultaneously (otherwise the implied correlation curves would be flat).

Before the financial crisis, implied correlations exhibited a typical form that became known as an *implied correlation skew*: implied tranche correlations showed a V-shaped relationship to attachment point; implied base correlations were strictly increasing; and implied correlations for senior tranches were comparatively high. The 2004 and 2006 rows of the table are typical of the behaviour of implied tranche correlations (see also Figure 12.3).

With the onset of the financial crisis, tranche spreads and implied tranche correlations became quite irregular. In particular, for mezzanine tranches it was often impossible to find a correlation number that would reproduce the spread that was observed on that day. For a more detailed description of the behaviour of implied correlation we refer the reader to the book by Brigo, Pallavicini and Torresetti (2010).

It can be argued that correlation skews reflect deficiencies of the benchmark model. To begin with, the Gauss copula is an ad hoc choice that is motivated mostly by analytical convenience and not by thorough data analysis. Moreover, it is highly unlikely that the dependence structure of the default times in a portfolio can be characterized by a single correlation number. Also, the assumption of constant recovery rates is at odds with reality.

Explaining observed CDO spreads in factor copula models. The fact that the benchmark model cannot reproduce observed CDO spreads for several tranches creates problems for the pricing of so-called *bespoke* tranches with non-standard maturities or attachment points and for the risk management of a book of tranches. In these applications one would like to take all available price information into account. There is a need for portfolio credit risk models that can be made consistent with observed spreads for several tranches simultaneously.

Table 12.4. Market quotes and implied tranche correlations for five-year tranches on the iTraxx Europe index on different dates. Note that the spread for the equity tranche corresponds to an upfront payment quoted as a percentage of the notional; the quarterly spread for the equity tranche is set to 5% by market convention.

Type of data	Year	Index	[0,3]	[3,6]	[6,9]	[9,12]	[12,22]
Market spread	2004	42 bp	27.6%	168 bp	70 bp	43 bp	20 bp
Tranche correlation	2004		22.4%	5.0%	15.3%	22.6%	30.6%
Market spread	2006	26 bp	14.5%	62 bp	18 bp	7 bp	3 bp
Tranche correlation	2006		18.6%	7.9%	14.1%	17.25%	23.54%
Market spread	2008	150 bp	46.5%	5.7%	3.7%	2.3%	1.45%
Tranche correlation	2008		51.1%	85.7%	95.3%	3.0%	17.6%

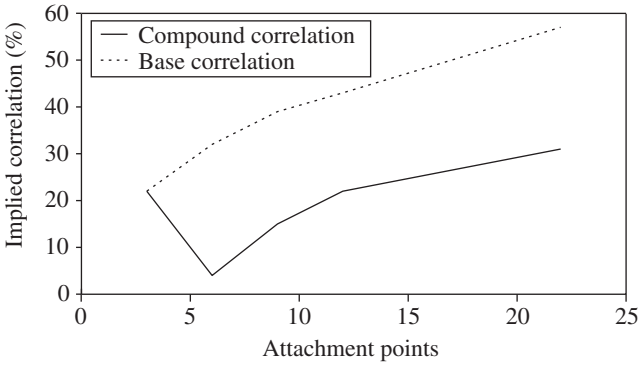

Figure 12.3. Compound correlation and base correlation corresponding to the CDO spreads in the 2004 row of Table 12.4. The spreads exhibit a typical correlation skew. (Data from Hull and White (2004).)

Quite naturally, many researchers have addressed this problem by considering more general factor copula models where some of the unrealistic assumptions made in the benchmark model are relaxed. To begin with, several authors have shown that by introducing state-dependent recovery rates that are negatively correlated with default probabilities, the correlation skew can be mitigated (but not eliminated completely). More importantly, alternative copula models have been employed, mostly models from the class of general one-factor copulas introduced in Example 12.4. It turns out that the double-t copula and, in particular, the double-GH copula model can be calibrated reasonably well to market-observed CDO tranche spreads (see, for example, Eberlein, Frey and von Hammerstein 2008). However, the empirical performance of these models worsened substantially with the onset of the credit crisis in 2007. For further details we refer to the references given in Notes and Comments.

12.3.3 The Implied Copula Approach

The class of implied copula models can be viewed as a generalization of the one-factor copula models considered in Section 12.2.2. In an implied copula model the factor variable V is modelled as a discrete rv whose probability mass function

is determined by calibration to market data such as observed index and tranche spreads. Since the support of V is usually taken as a fairly large set, this creates some additional flexibility that helps to explain observed market spreads for CDO tranches.

Definition and key properties. The structure of the joint survival function of $\tau_1, \dots, \tau_m$ in an implied copula model is similar to the mixture representation (12.16) in a factor copula model. It is assumed that the τ_i are conditionally independent given a mixing variable V that takes the discrete values $v_1, \dots, v_K$ (the *states* of the system); the conditional survival probabilities are of the form

$$Q(\tau_i > t \mid V = v_k) = \exp(-\lambda_i(v_k)t), \quad 1 \leqslant i \leqslant m,\ t \geqslant 0, \tag{12.24}$$

for functions $\lambda_i : \{v_1, \dots, v_K\} \to (0, \infty)$. The probability mass function of V is denoted by $\boldsymbol{\pi} = (\pi_1, \dots, \pi_k)$ with $\pi_k = Q(V = v_k), k = 1, \dots, K$. It follows that the joint survival function of $\tau_1, \dots, \tau_m$ is given by

$$\bar{F}(t_1, \dots, t_m) = \sum_{k=1}^{K} \pi_k \left\{ \prod_{i=1}^{m} \exp(-\lambda_i(v_k)t_i) \right\}. \tag{12.25}$$

It is possible to make the functions $\lambda_i(\cdot)$ in (12.24) time dependent, and this is in fact necessary if the model is to be calibrated simultaneously to tranche spreads of different maturities. We will describe the time-independent case for ease of exposition.

Note that the name "implied copula model" is somewhat misleading, since both the dependence structure and the marginal distributions of the τ_i change if the distribution of V is changed. For instance, for $\boldsymbol{\pi} = (1, 0, \dots, 0)$ we have $Q(\tau_i > t) = \exp(-\lambda_i(v_1)t)$, whereas for $\boldsymbol{\pi} = (0, \dots, 0, 1)$ we have $Q(\tau_i > t) = \exp(-\lambda_i(v_K)t)$. A better name for (12.25) would therefore be "a model with an implied factor distribution", but the label "implied copula model" has become standard, essentially because of the influential paper by Hull and White (2006).

In order to compute spreads of CDSs, index swaps or single-tranche CDOs in an implied copula model, we proceed in a similar fashion to the factor copula models discussed in the previous section. First we compute the conditional values of the default payment leg and the premium payment leg given $V = v_k$, denoted by $V^{\mathrm{Def}}(v_k)$ and $V^{\mathrm{Prem}}(v_k, x)$, where x represents a generic spread level; the methods described in Section 12.3.1 can be used for this purpose. The unconditional values of the default payment and the premium payment legs are then given by averaging over the different states: that is,

$$V^{\mathrm{Def}} = \sum_{k=1}^{K} \pi_k V^{\mathrm{Def}}(v_k) \quad \text{and} \quad V^{\mathrm{Prem}}(x) = \sum_{k=1}^{K} \pi_k V^{\mathrm{Prem}}(v_k, x).$$

The fair spread of the structure is equal to $x^* = V^{\mathrm{Def}} / V^{\mathrm{Prem}}(1)$.

Implied copula models are mostly used for the pricing of CDO tranches with non-standard attachment points or maturities. It is also possible to value tranches where

the underlying pool is different from a standard credit index, but the valuation of such products is somewhat subjective; we refer to the references given in Notes and Comments. In Section 17.4.3 we explain how to embed an implied copula model into a dynamic portfolio credit risk model. In that extension of the model it is possible to price options on standard index products such as CDS index swaps.

Example 12.7 (implied copula models for a homogeneous portfolio). In a homogeneous portfolio the default intensities of all firms are identical, $\lambda_i(\cdot) = \lambda(\cdot)$, and we may parametrize the model directly in terms of the values $\lambda_1, \dots, \lambda_K$ of the default intensity in the different states. Moreover, the states can be ordered so that $\lambda_1 < \lambda_2 < \cdots < \lambda_K$. In this way, state 1 can be viewed as the best state of the economy (the one with the lowest default intensity) and state K corresponds to the worst state. In our calibration example below we use a model with $K = 9$ states and we assume that the default intensity takes values in the set $\{0.01\%, 0.3\%, 0.6\%, 1.2\%, 2.5\%, 4.0\%, 8.0\%, 20\%, 70\%\}$.

For heterogeneous portfolios we require more sophisticated parametrizations. An example can be found in Rosen and Saunders (2009); further references are given in Notes and Comments.

In practical implementations of the model the set of states $\{v_1, \dots, v_K\}$ is typically chosen at the outset of the analysis and is kept fixed during model calibration. Moreover, in order to obtain pricing results for bespoke index products that are robust with respect to the precise values of the v_k it is advisable to work on a fine grid, and hence with a fairly large number of states K. Of course, this means that the determination of the probability mass function $\boldsymbol{\pi}$ becomes a high-dimensional problem, but the calibration of $\boldsymbol{\pi}$ is comparatively easy, as we now explain.

Calibration of $\boldsymbol{\pi}$. In the implied copula framework we need to determine the probability mass function $\boldsymbol{\pi}$ from observed market data. As is usual in model calibration, $\boldsymbol{\pi}$ is found by minimizing some distance between market prices and model prices. This is substantially facilitated by the observation that model prices are linear in $\boldsymbol{\pi}$. The set of all probability mass functions consistent with the price information at a given point in time t can therefore be described in terms of a set of linear inequalities. We now explain this for the case of a corporate bond and a single-name CDS; similar arguments apply for index swaps and CDO tranches.

- Consider a zero-coupon bond issued by firm i with maturity T, and denote by $p_i(v_k) = \exp(-\lambda_i(v_k)(T-t))$ the value of the bond in state v_k (we assume a recovery rate equal to zero for simplicity). Suppose that we observe bid and ask quotes $\underline{p} \leqslant \overline{p}$ for the bond. In order to be consistent with this information, a probability mass function $\boldsymbol{\pi}$ needs to satisfy the following linear inequalities:

$$\underline{p} \leqslant \sum_{k=1}^{K} \pi_k p_i(v_k) \leqslant \overline{p}.$$

- Consider next a CDS contract on firm i; this is a simple example of a contract where two cash-flow streams are exchanged. Suppose that at time t we

observe bid and ask spreads $\underline{x} \leqslant \overline{x}$ for the contract. Denote by $V_i^{\mathrm{Def}}(v_k)$ and $V_i^{\mathrm{Def}}(v_k, x)$ the values of the premium payment and default payment legs of the contract in state k for a generic CDS spread x. Then $\boldsymbol{\pi}$ must satisfy the following two inequalities:

$$\sum_{k=1}^{K} \pi_k (V_i^{\mathrm{prem}}(v_k, \underline{x}) - V_i^{\mathrm{def}}(v_k)) \leqslant 0,$$
$$\sum_{k=1}^{K} \pi_k (V_i^{\mathrm{prem}}(v_k, \overline{x}) - V_i^{\mathrm{def}}(k)) \geqslant 0.$$

Moreover, $\boldsymbol{\pi}$ needs to satisfy the obvious linear constraints $\pi_k \geqslant 0$ for all k and $\sum_{k=1}^{K} \pi_k = 1$.

It follows from the above discussion that the constraints on $\boldsymbol{\pi}$ at time t can be written in the generic form

$$A\boldsymbol{\pi} \leqslant b$$

for some matrix $A \in \mathbb{R}^{N \times K}$, some vector $b \in \mathbb{R}^N$ and some $N \in \mathbb{N}$. In order to find a vector $\boldsymbol{\pi}$ that satisfies this system of linear inequalities, fix a vector $\boldsymbol{c} = (c_1, \ldots, c_K)$ of weights and consider the linear programming problem

$$\min_{\boldsymbol{\pi} \in \mathbb{R}^K} \sum_{i=1}^{K} c_k \pi_k \quad \text{subject to } A\boldsymbol{\pi} \leqslant b. \tag{12.26}$$

Clearly, every solution of (12.26) is a probability mass function that is consistent with the given price information in the sense that it solves the system $A\boldsymbol{\pi} \leqslant b$. Note that problem (12.26) can be solved with standard linear programming software.

If the number of states K is large compared with the number of constraints, there will typically be more than one probability mass function $\boldsymbol{\pi}$ that solves the system $A\boldsymbol{\pi} \leqslant b$. An easy way to check this is to vary the weight vector $\boldsymbol{c}$ in (12.26), since different weight vectors usually correspond to different solutions of the system $A\boldsymbol{\pi} \leqslant b$. In that case a unique solution $\boldsymbol{\pi}^*$ of the calibration problem can be determined by a suitable *regularization procedure*. For instance, one could choose $\boldsymbol{\pi}^*$ by minimizing the *relative entropy* to the uniform distribution on the set $\{v_1, \ldots, v_K\}$. This leads to the convex optimization problem

$$\min_{\boldsymbol{\pi} \in \mathbb{R}^K} \sum_{i=1}^{K} \pi_k \ln \pi_k \quad \text{subject to } A\boldsymbol{\pi} \leqslant b, \tag{12.27}$$

which can be addressed with standard convex programming algorithms (see, for example, Bertsimas and Tsitsiklis 1997). Model calibration via entropy minimization has a number of attractive features, as is explained, for example, in Avellaneda (1998). In particular, the prices of bespoke tranches (the model output) depend smoothly on the spreads observed on the market (the model input).

We close this section on implied copula models by presenting results from a calibration exercise of Frey and Schmidt (2011). In that paper the homogeneous

Table 12.5. Results of the calibration of a homogeneous implied copula model to iTraxx spread data (index and tranches) for different data sets from several years; the components of $\boldsymbol{\pi}^*$ are expressed as percentages. The numerical results are from Frey and Schmidt (2011).

	λ (in %)								
	0.01	0.3	0.6	1.2	2.5	4.0	8.0	20	70
$\boldsymbol{\pi}^*$, data from 2004	12.6	22.9	42.0	17.6	2.5	1.45	0.54	0.13	0.03
$\boldsymbol{\pi}^*$, data from 2006	22.2	29.9	39.0	7.6	1.2	0.16	0.03	0.03	0.05
$\boldsymbol{\pi}^*$, data from 2008	1.1	7.9	57.6	10.8	11.7	4.9	1.26	1.79	2.60
$\boldsymbol{\pi}^*$, data from 2009	0.0	13.6	6.35	42.2	22.3	12.5	0.0	0.00	3.06

model of Example 12.7 was calibrated to iTraxx tranche and index spread data for the years 2004, 2006, 2008 and 2009; all contracts had a maturity of five years. The data from 2004 and 2006 are typical for tranche and index spreads before the credit crisis; the data from 2008 and 2009 represent the state of the market during the crisis. Entropy minimization was used in order to determine a solution $\boldsymbol{\pi}^*$ of the calibration problem. The resulting values of $\boldsymbol{\pi}^*$ are given in Table 12.5. We clearly see that with the emergence of the credit crisis the calibration procedure puts more mass on states where the default intensity is high; in particular, the extreme state where $\lambda = 70\%$ gets a probability of around 3%. This reflects the increased awareness of future defaults and the increasing risk aversion in the market after the beginning of the crisis.

The implied copula model can usually be calibrated very well to tranche and index spreads with a single maturity; calibrating tranche spreads for several maturities simultaneously is more involved (see, for example, Brigo, Pallavicini and Torresetti (2010) for details).

Notes and Comments

Semianalytic approaches for the pricing of synthetic CDOs in factor copula models have been developed by Laurent and Gregory (2005), Hull and White (2004), Gibson (2004) and Andersen and Sidenius (2004), among others. Laurent and Gregory exploit the conditional independence structure of factor copula models and develop methods based on Fourier analysis; Andersen and Sidenius, Gibson and Hull and White propose recursive methods. The LPA is originally due to Vasicek (1997). Frey, Popp and Weber (2008) propose the normal approximation as a simple yet efficient alternative to the LPA. A very deep study of normal and Poisson approximations for the sum of independent random variables with applications to CDO pricing is El Karoui, Kurtz and Jiao (2008).

There is a rich literature on general one-factor copula models in relation to implied correlation skews: the double-t copula has been studied in Hull and White (2004) and Vrins (2009); double-GH copulas have been analysed by Eberlein, Frey and von Hammerstein (2008), Guégan and Houdain (2005) and Kalemanova, Schmid and Werner (2005), among others. The idea of using state-dependent recovery rates to improve the fit of CDO pricing models is explored, for instance, in Hull and

White (2006). The base-correlation approach is used in a number of (fairly dubious) extrapolation procedures for the pricing of non-standard tranches. An in-depth discussion of implied correlation skews in credit risk can be found in Brigo, Pallavicini and Torresetti (2010).

The implied copula model is due to Hull and White (2006); similar ideas can be found in Rosen and Saunders (2009) and in Frey and Schmidt (2012). Hull and White (2006) and Rosen and Saunders (2009) discuss the pricing of bespoke CDO tranches in the context of the model. The calibration of implied copula models to inhomogeneous portfolios is discussed in detail in Rosen and Saunders (2009) and in Frey and Schmidt (2012). An empirical assessment of the calibration properties of implied copula models is also given in Brigo, Pallavicini and Torresetti (2010). Calibration methods based on entropy minimization are discussed by Avellaneda (1998).

13
Operational Risk and Insurance Analytics

We have so far concentrated on the modelling of market and credit risk, which reflects the historical development of quantitative risk management in the banking context. Some of the techniques we have discussed are also relevant in operational risk modelling, particularly the statistical models of extreme value theory (EVT) in Chapter 5 and the aggregation methodology of Chapter 8. But we also need other techniques tailored specifically to operational risk, and we believe that actuarial models used in non-life insurance are particularly relevant.

In the first half of this chapter (Section 13.1) we examine the Basel requirements for the quantitative modelling of operational risk in banks, discussing various potential approaches. On the basis of industry data we highlight the challenges involved in implementing a so-called advanced measurement (AM) approach based on modelling loss distributions, also known as the loss distribution approach (LDA).

In operational risk there is no consensus concerning the best modelling approach. In contrast to market and credit risk, the data sources are more limited and the overall statistical properties of available data show a high degree of non-homogeneity and non-stationarity, defying straightforward applications of statistical tools. Rather than offering specific models, the current chapter aims to provide a set of tools that can be used to learn more about this important but difficult-to-model risk category.

In Section 13.2 we summarize the techniques from actuarial modelling that are relevant to operational risk, under the heading of *insurance analytics*. Our discussion in that section, though motivated by quantitative modelling of operational risk, has much wider applicability in quantitative risk management. For example, some techniques have implicitly been used in the credit risk chapters. The Notes and Comments section at the end of the chapter gives an overview of further techniques from insurance mathematics that have potential applications in the broader field of quantitative risk management.

13.1 Operational Risk in Perspective

13.1.1 An Important Risk Class

In our overview of the development of the Basel regulatory framework in Section 1.2.2 we explained how *operational risk* was introduced under Basel II as a new risk class for which financial institutions were bound to set aside regulatory capital. It has also been incorporated into the Solvency II framework for insurers, although we will concentrate on the banking treatment in this chapter.

We first recall the Basel definition as it appears in the final comprehensive version of the Basel II document (Basel Committee on Banking Supervision 2006).

> Operational risk is defined as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk.

Examples of losses due to operational risk. Examples of losses falling within this category are, for instance, fraud (internal as well as external), losses due to IT failures, errors in settlements of transactions, litigation and losses due to external events like flooding, fire, earthquake or terrorism. Losses due to unfortunate management decisions, such as many of the mergers and acquisitions in the two decades leading up to the 2007–9 financial crisis, are definitely not included. However, the fact that legal risk is part of the definition has had a considerable impact on the financial industry in the aftermath of the crisis.

An early case that touched upon almost all aspects of the above definition was that of Barings (see also Section 1.2.2). From insufficient internal checks and balances (processes), to fraud (human risk), to external events (the Kobe earthquake), many operational risk factors contributed to the downfall of this once-renowned merchant bank. Further examples include the $691 million *rogue trading* loss at Allfirst Financial, the $484 million settlement due to misleading sales practices at Household Finance, and the estimated $140 million loss for the Bank of New York stemming from the September 11 attacks.

More recent examples in which legal risk has been involved include LIBOR rigging, for which the European Commission fined eight large financial institutions a total of $2.3 billion, and the possible rigging of rates in the foreign-exchange market, a market with an estimated $5.35 trillion daily turnover. Following the 2007–9 financial crisis several financial institutions faced fines for misselling of financial products—particularly securitized credit products—on the basis of inaccurate or misleading information about their risks. In the latter category, the Bank of America was fined the record amount of $16.65 billion on 21 August 2014.

Increasing use of algorithmic and high-frequency trading has resulted in operational losses. Examples include the 2010 Flash Crash and an estimated $440 million loss from a computer-trading glitch at Knight Capital Group. A number of prominent cases involving large losses attributable to rogue traders have also been widely reported in the press, including Fabrice Tourre at Goldman Sachs, Kweku Adoboli at UBS, Jérôme Kerviel at Société Générale and Bruno Iksil (also known as the London Whale) at JPMorgan Chase.

All the examples cited above, and the seriousness with which they have been taken by regulators worldwide, offer clear proof of the fundamental importance of operational risk as a risk class. Many banks have been forced to increase the capital they hold for operational risk (in some cases by up to 50%). An example of a regulatory document that probes more deeply into a trading loss at UBS is FINMA (2012).

Distinctiveness of operational risk. An essential difference between operational risk, on the one hand, and market and credit risk, on the other, is that operational risk has no upside for a bank. It comes about through the malfunctioning of parts of daily business and hence is as much a question of quality control as anything else. Although banks try as hard as possible to reduce operational risk, operational losses continue to occur.

Despite their continuing occurrence, a lack of publicly available, high-quality operational loss data has been a major issue in the development of operational risk models. This is similar to the problem faced by underwriters of catastrophe insurance. The insurance industry's answer to the problem has involved data pooling across industry participants, and similar developments are now taking place in the banking industry. Existing sources of pooled data include the Quantitative Impact Studies (QISs) of the Basel Committee, the database compiled by the Federal Reserve Bank of Boston, and subscription-based services for members like that of the ORX (Operational Riskdata eXchange Association). Increasingly, private companies are also providing data. However, the lack of more widely accessible data at the individual bank level remains a major problem. As data availability improves, many of the methods discussed in this book (such as the extreme value models of Chapter 5 and the insurance analytics of Section 13.2) will become increasingly useful.

Elementary versus advanced measurement approaches. In Section 13.1.3 we discuss the *advanced measurement* (AM) approach, which is typically adopted by larger banks that have access to high-quality operational loss data. This approach is often referred to as the *loss distribution approach* (LDA), since a series of distributional models or stochastic processes are typically fitted to operational loss data that have been categorized into different types of loss.

First, however, we discuss the so-called *elementary approaches* to operational risk modelling. In these approaches, aimed at smaller banks, the detailed modelling of loss distributions for different loss types is not required; a fairly simple volume-based capital charge is proposed.

We note that, as in the case of credit risk, the approaches proposed in the Basel framework for the calculation of regulatory capital represent a gradation in complexity. Recall that, for credit risk, banks must implement either the standardized approach or the internal-ratings-based (IRB) approach, as discussed in Section 1.3.1. The field of regulation is in a constant state of flux and the detail of the methods we describe below may change over time but the underlying principles are likely to continue to hold.

13.1.2 The Elementary Approaches

There are two elementary approaches to operational risk measurement. Under the *basic-indicator* (BI) approach, banks must hold capital for operational risk equal to the average over the previous three years of a fixed percentage (denoted by α) of positive annual gross income (GI). Figures for any year in which annual gross income is negative or zero should be excluded from both the numerator and denominator

when calculating the average. The risk capital under the BI approach for operational risk in year t is therefore given by

$$\mathrm{RC}^t_{\mathrm{BI}}(\mathrm{OR}) = \frac{1}{Z_t} \sum_{i=1}^{3} \alpha \max(\mathrm{GI}^{t-i}, 0), \tag{13.1}$$

where $Z_t = \sum_{i=1}^{3} I_{\{\mathrm{GI}^{t-i}>0\}}$ and GI^{t-i} stands for gross income in year $t-i$. Note that an operational risk capital charge is calculated on a yearly basis. The BI approach gives a fairly straightforward, volume-based, one-size-fits-all capital charge. Based on the various QISs, the Basel Committee suggests a value of α of 15%.

Under the *standardized* approach, banks' activities are divided into eight *business lines*: corporate finance; trading & sales; retail banking; commercial banking; payment & settlement; agency services; asset management; and retail brokerage. Precise definitions of these business lines are to be found in the final Basel II document (Basel Committee on Banking Supervision 2006). Within each business line, gross income is a broad indicator that serves as a proxy for the scale of business operations and thus the likely scale of operational risk exposure. The capital charge for each business line is calculated by multiplying gross income by a factor (denoted by β) assigned to that business line. As in (13.1), the total capital charge is calculated as a three-year average over positive GIs, resulting in the following capital charge formula:

$$\mathrm{RC}^t_{\mathrm{S}}(\mathrm{OR}) = \frac{1}{3} \sum_{i=1}^{3} \max\left[\sum_{j=1}^{8} \beta_j \mathrm{GI}^{t-i}_j, 0 \right]. \tag{13.2}$$

It may be noted that in formula (13.2), in any given year $t-i$, negative capital charges (resulting from negative gross income) in some business line j can offset positive capital charges in other business lines (albeit at the discretion of the national supervisor). This kind of "netting" should induce banks to go from the basic-indicator approach to the standardized approach; the word "netting" is of course to be used with care here. Based on the QISs, the Basel Committee has set the beta coefficients as in Table 13.1. Moscadelli (2004) gives a critical analysis of these beta factors, based on the full database of more than 47 000 operational losses of the second QIS of the summer of 2002 (see also Section 13.1.4). Concerning the use of GI in (13.1) and (13.2), see Notes and Comments.

13.1.3 Advanced Measurement Approaches

Under an AM approach, the regulatory capital is determined by a bank's own internal risk-measurement system according to a number of quantitative and qualitative criteria set forth in the regulatory documentation (Basel Committee on Banking Supervision 2006). We will not detail every relevant step in the procedure that leads to the acceptance of an AM approach for an internationally active bank and its subsidiaries; the Basel Committee's documents give a clear and readable account of this. We focus instead on the methodological aspects of a full quantitative approach to operational risk measurement. It should be stated, however, that, as in the case of

Table 13.1. Beta factors for the standardized approach.

Business line (j)	Beta factors (β_j)
$j = 1$, corporate finance	18%
$j = 2$, trading & sales	18%
$j = 3$, retail banking	12%
$j = 4$, commercial banking	15%
$j = 5$, payment & settlement	18%
$j = 6$, agency services	15%
$j = 7$, asset management	12%
$j = 8$, retail brokerage	12%

market and credit risk, the adoption of an AM approach to operational risk is subject to approval and continuing quality checking by the national supervisor.

While the BI and standardized approaches prescribe the explicit formulas (13.1) and (13.2), the AM approach lays down general guidelines. In the words of the Basel Committee (see Basel Committee on Banking Supervision 2006, paragraph 667):

> Given the continuing evolution of analytical approaches for operational risk, the Committee is not specifying the approach or distributional assumptions used to generate the operational risk measure for regulatory capital purposes. However, a bank must be able to demonstrate that its approach captures potentially severe "tail" loss events. Whatever approach is used, a bank must demonstrate that its operational risk measure meets a soundness standard comparable to that of the internal ratings-based approach for credit risk (comparable to a one year holding period and the 99.9 percent confidence interval).

In the usual LDA interpretation of an AM approach, operational losses are typically categorized according to the eight business lines mentioned in Section 13.1.2 as well as the following seven *loss-event types*: internal fraud; external fraud; employment practices & workplace safety; clients, products & business practices; damage to physical assets; business disruption & system failures; and execution, delivery & process management. While the categorization of losses in terms of eight business lines and seven loss-event types is standard, banks may deviate from this format if appropriate.

Banks are expected to gather internal data on repetitive, high-frequency losses (three to five years of data), as well as relevant external data on non-repetitive low-frequency losses. Moreover, they must add stress scenarios both at the level of loss severity (parameter shocks to model parameters) and correlation between loss types. In the absence of detailed joint models for different loss types, risk measures for the aggregate loss should be calculated by summing across the different loss categories. In general, both so-called *expected* and *unexpected* losses should be taken into account (i.e. risk-measure estimates cannot be reduced by subtraction of an expected loss amount).

We now describe a skeletal version of a typical AM solution for the calculation of an operational risk charge for year t. We assume that historical loss data from previous years have been collected in a data warehouse with the structure

$$\{X_k^{t-i,b,\ell} : i = 1, \dots, T;\ b = 1, \dots, 8;\ \ell = 1, \dots, 7;\ k = 1, \dots, N^{t-i,b,\ell}\}, \tag{13.3}$$

where $X_k^{t-i,b,\ell}$ stands for the kth loss of type ℓ for business line b in year $t-i$; $N^{t-i,b,\ell}$ is the number of such losses and $T \geqslant 5$ years, say. Note that thresholds may be imposed for each (i, b, ℓ) category, and small losses less than the threshold may be neglected; a threshold is typically of the order of €10 000. The total historical loss amount for business line b in year $t-i$ is obviously

$$L^{t-i,b} = \sum_{\ell=1}^{7} \sum_{k=1}^{N^{t-i,b,\ell}} X_k^{t-i,b,\ell}, \tag{13.4}$$

and the total loss amount for year $t-i$ is

$$L^{t-i} = \sum_{b=1}^{8} L^{t-i,b}. \tag{13.5}$$

The problem in the AM approach is to use the loss data to estimate the distribution of L_t for year t and to calculate risk measures such as VaR or expected shortfall (see Section 2.3) for the estimated distribution. Writing ϱ_α for the risk measure at a confidence level α, the regulatory capital is determined by

$$\mathrm{RC}_{\mathrm{AM}}^t(\mathrm{OR}) = \varrho_\alpha(L^t), \tag{13.6}$$

where α would typically take a value in the range 0.99–0.999 imposed by the local regulator. Because the joint distributional structure of the losses in (13.4) and (13.5) for any given year is generally unknown, we would typically resort to simple aggregation of risk measures across loss categories to obtain a formula of the form

$$\mathrm{RC}_{\mathrm{AM}}^t(\mathrm{OR}) = \sum_{b=1}^{8} \varrho_\alpha(L^{t,b}). \tag{13.7}$$

In view of our discussions in Chapter 8, the choice of an additive rule in (13.7) can be understood. Indeed, for any coherent risk measure ϱ_α, the right-hand side of (13.7) yields an upper bound for the total risk $\varrho_\alpha(L^t)$. In the important case of VaR, the right-hand side of (13.7) corresponds to the comonotonic scenario (see Proposition 7.20). The optimization results of Section 8.4.4 can be used to calculate bounds for $\varrho_\alpha(L^t)$ under different dependence scenarios for the business lines (see, in particular, Example 8.40).

Reduced to its most stylized form in the case when $\varrho_\alpha = \mathrm{VaR}_\alpha$ and $\alpha = 0.999$, a capital charge under the AM approach requires the calculation of a quantity of the type

$$\mathrm{VaR}_{0.999}\left(\sum_{k=1}^{N} X_k\right), \tag{13.8}$$

where (X_k) is some sequence of loss *severities* and N is an rv describing the *frequency* with which operational losses occur. Random variables of the type (13.8) are one of the prime examples of the actuarial models that we treat in Section 13.2.2. Before we move on to those models in the next section, we highlight some "stylized facts" concerning operational loss data.

13.1.4 Operational Loss Data

In order to reliably estimate (13.6), (13.7) or, in a stylized version, a quantity like (13.8), we need extensive data. The data situation for operational risk is much worse than that for credit risk, and is clearly an order of magnitude worse than for market risk, where vast quantities of data are publicly available. As discussed in Section 13.1.1, banks have not been gathering data for long and pooling initiatives are still in their infancy. As far as we know, no reliable *publicly* available data source on operational risk exists.

Our discussion below is based on industry data that we have been able to analyse as well as on the findings in Moscadelli (2004) for the QIS database and the results of the 2004 loss-data collection exercise by the Federal Reserve Bank of Boston (see Federal Reserve Bank of Boston 2005). An excellent overview of some of the data characteristics is to be found in the Basel Committee's report (Basel Committee on Banking Supervision 2003). As the latter report states:

> Despite this progress, inferences based on the data should still be made with caution.... In addition, the most recent data collection exercise provides data for only one year and, even under the best of circumstances, a one-year collection window will provide an incomplete picture of the full range of potential operational risk events, especially of rare but significant "tail events".

Further information is to be found in the increasing number of papers on the topic, and particularly in the various papers coming out of the ORX Consortium (see Notes and Comments).

In Figure 13.1 we have plotted operational loss data obtained from several sources; parts (a)–(c) show losses for three business lines for the period 1992–2001. It is less important for the reader to know the exact loss type—it is sufficient to accept that the data are typical for (b, ℓ) categories in (13.3). In part (d) the data from the three previous figures have been pooled.

Exploratory data analysis reveals the following stylized facts (confirmed in several other studies):

- loss severities have a heavy-tailed distribution;
- losses occur randomly in time; and
- loss frequency may vary substantially over time.

The third observation is partly explained by the fact that banks have gathered an increasing amount of operational loss data since the Basel II rules were first announced. There is therefore a considerable amount of *reporting bias*, resulting

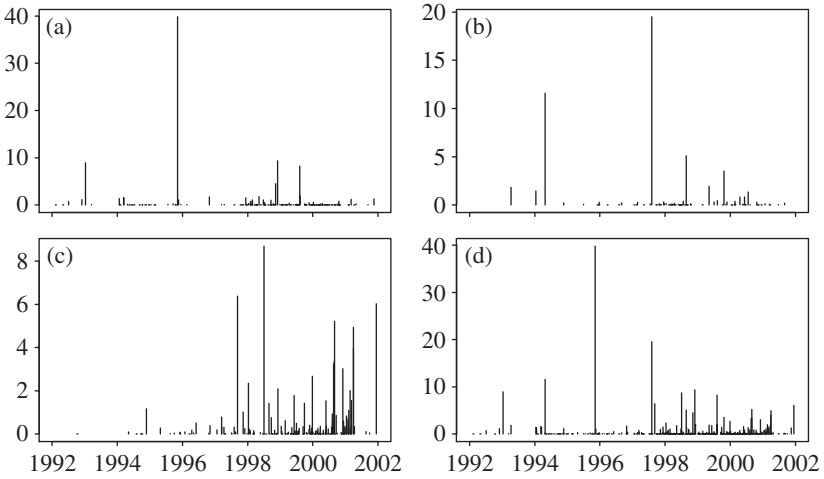

Figure 13.1. Operational risk losses: (a) type 1, $n = 162$; (b) type 2, $n = 80$; (c) type 3, $n = 175$; and (d) pooled losses $n = 417$.

in fewer losses in the first half of the 1990s and more losses afterwards. Moreover, several classes of loss may have a considerable cyclical component and/or may depend on changing economic covariables. For instance, back-office errors may depend on volume traded and fraud may be linked to the overall level of the economy (depressions versus boom cycles). Moreover, there may be a rise in legal losses in the aftermath of a severe crisis, as has been observed for the 2007–9 credit crisis. This clear inhomogeneity in the loss frequency makes an immediate application of statistical methodology difficult. However, it may be reasonable to at least assume that the (inflation-adjusted) loss sizes have a common severity distribution, which would allow, for instance, the application of methods from Chapter 5.

In Figure 13.2 we have plotted the sample mean excess functions (5.16) for the data in Figure 13.1. This figure clearly indicates the first stylized fact of heavy-tailed loss severities. The mean excess plots in (a) and (b) are clearly increasing in an approximately linear fashion, pointing to Pareto-type behaviour. This contrasts with (c), where the plot appears to level off from a threshold of 1. This hints at a loss distribution with finite upper limit, but this can only be substantiated by more detailed knowledge of the type of loss concerned. Pooling the data in (d) masks the different kinds of behaviour, and perhaps illustrates the dangers of naive statistical analyses that do not consider the data-generating mechanism.

Moscadelli (2004) performed a detailed EVT analysis (including a first attempt to solve the frequency problem) of the full QIS data set of more than 47 000 operational losses and concluded that the loss dfs are well fitted by generalized Pareto distributions (GPDs) in the upper-tail area (see Section 5.2.2 for the necessary statistical background). The estimated tail parameters (ξ in (5.14)) for the different business lines range from 0.85 for asset management to 1.39 for commercial banking. Six of the business lines have an estimate of ξ greater than 1, corresponding to an *infinite-mean model*! Based on these QIS data, the estimated risk capital/GI ratios (the β in

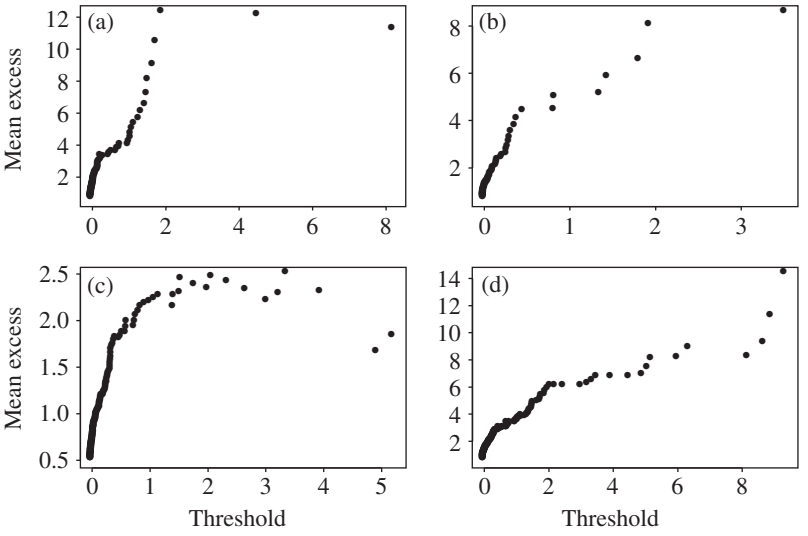

Figure 13.2. Corresponding sample mean excess plots for the data in Figure 13.1: (a) type 1; (b) type 2; (c) type 3; and (d) pooled.

Table 13.1) range from 8.3% for retail banking to 33.3% for payment & settlement, with an overall alpha value (see (13.1)) of 13.3%, slightly below the Basel II value of 15% used in the BI approach. Note the much wider range of values of β that emerge from the analysis of the QIS data compared with the prescribed range of 12–18% for the standardized approach in Table 13.1.

Notes and Comments

Several textbooks on operational risk have been published: see, for example, Cruz (2002, 2004), King (2001), the Risk Books publication edited by Jenkins and Roberts (2004), and chapters in Ong (2003) and Crouhy, Galai and Mark (2001). In particular, Chapter 4 of Cruz (2004), written by Carolyn Currie, gives an excellent overview of the regulatory issues surrounding operational risk. Further textbooks include Shevchenko (2011), Cruz, Peters and Shevchenko (2015) and Panjer (2006). The *Journal of Operational Risk* publishes relevant research from academia and practice, including informative papers coming out of the ORX Consortium: see, for example, Cope and Antonini (2008) and Cope and Labbi (2008).

Papers on implementing operational risk models in practice include Ebnöther et al. (2003); Frachot, Georges and Roncalli (2001), which discusses the loss distribution approach to operational risk; Doebeli, Leippold and Vanini (2003), which shows that a good operational risk framework may lead to an overall improvement in the quality of business operations; and Aue and Kalkbrener (2006), which provides a comprehensive description of an approach based on loss distributions that was developed at Deutsche Bank. Excellent data-analytic papers using published operational risk losses are de Fontnouvelle et al. (2003) and Moscadelli (2004). Rosenberg and Schuermann (2006) address the aggregation of market, credit and operational risk measures. A comprehensive study, combining internal and external

data, together with expert opinion, in a scenario-generation context is de Jongh et al. (2014); this paper also contains an excellent list of references.

Figure 13.1 is taken from Embrechts, Kaufmann and Samorodnitsky (2004). This paper also stresses the important difference between so-called repetitive and non-repetitive losses. For the former (to some extent less important) losses, statistical modelling can be very useful. For non-repetitive, low-probability, high-severity losses, much more care has to be taken before a statistical analysis can be performed (see Pézier 2003a,b).

EVT methods for operational risk quantification have been used by numerous authors (see, for example, Coleman 2002, 2003; Medova 2000; Medova and Kyriacou 2000). Because of the non-stationarity of operational loss data over several years, more refined EVT models are called for: see, for example, Chavez-Demoulin and Embrechts (2004) and Chavez-Demoulin, Embrechts and Hofert (2014) for some examples of such models. For a critical article on the use of EVT for the calculation of an operational risk capital charge, see Embrechts, Furrer and Kaufmann (2003), which contains a simulation study of the number of data needed to come up with a reasonable estimate of a high quantile. The use of statistical methods other than EVT are discussed in the textbooks referred to above. These methods include linear predictive models, Bayesian belief networks, discriminant analysis, and tools and techniques from reliability theory.

We have noted that severity models for operational risk are typically (extremely) heavy tailed. Several publications report infinite-mean models. For an early discussion of this issue, see Nešlehová, Embrechts and Chavez-Demoulin (2006). The reader interested in a more philosophical discussion of the economic sense of such models should carry out an internet search for "the dismal theorem" and read some of the material that has been written on this topic. The term "dismal theorem" was coined by Martin Weitzman, who introduced it in relation to the economics of catastrophic climate change; an interesting paper on the topic is Weitzman (2011).

It should be clear from what we learned in earlier chapters about risk measures, their aggregation and their statistical estimation that calculating a 99.9% yearly VaR for operational risk is, to put it mildly, a daunting task. As a consequence, Ames, Schuermann and Scott (2014) suggest several regulatory policy changes leading to simpler, more standardized, more stable and more robust methodologies, at least until our understanding of operational risk has increased. In a recent publication, the Basel Committee on Banking Supervision (2014) proposes a change from the gross income (GI) indicator in (13.1) and (13.2) to a new business indicator (BI).

13.2 Elements of Insurance Analytics

13.2.1 The Case for Actuarial Methodology

Actuarial tools and techniques for the modelling, pricing and reserving of insurance products in the traditional fields of life insurance, non-life insurance and reinsurance have a long history going back more than a century. More recently, the border between financial and insurance products has become blurred, examples of this

process being equity-linked life products and the transfer of insurance risks to the capital markets via securitization (see Chapter 1, particularly Section 1.5.3 and the Notes and Comments).

Whereas some of the combined bank-assurance products have not met with the success that was originally hoped for, it remains true that there exists an increasing need for financial and actuarial professionals who can close the methodological gaps between the two fields. In the sections that follow we discuss insurance-analytical tools that we believe the more traditional finance-oriented risk manager ought to be aware of; the story behind the name insurance analytics can be found in Embrechts (2002).

It is not only the occasional instance of joint product development between the banking and insurance worlds that prompts us to make a case for actuarial methodology in QRM, but also the observation that many of the concepts and techniques of QRM described in the preceding chapters are in fact borrowed from the actuarial literature.

- Risk measures like expected shortfall (Definition 2.12) have been studied in a systematic way in the insurance literature. Expected shortfall is also the standard risk measure to be used under the Solvency II guidelines.
- Many of the dependence modelling tools presented in Chapter 7 saw their first applications in the realm of insurance. Moreover, notions like comonotonicity of risk factors have their origins in actuarial questions.
- In Section 2.3.5 we discussed the axiomatization of financial risk measures and mentioned the parallel development of insurance premium principles (often with very similar goals and results).
- The statistical modelling of extremal events has been a bread-and-butter subject for actuaries since the start of insurance. Many of the tools presented in Chapter 5 are therefore well known to actuaries.
- Within the world of credit risk management, the industry model CreditRisk$^+$ (Section 11.2.5) is known as an actuarial model.
- The actuarial approach to the modelling of operational risk is apparent in the AM approach of Section 13.1.3.

In the sections that follow we give a brief discussion of relevant actuarial techniques. The material presented should enable the reader to transfer actuarial concepts to QRM in finance more easily. We do not strive for a full treatment of the relevant tools as that would fill a separate (voluminous) textbook (see, for example, Denuit and Charpentier (2004), Mikosch (2004) and Partrat and Besson (2004) for excellent accounts of many of the relevant techniques).

13.2.2 The Total Loss Amount

Reconsider formula (13.8), where a random number N of random losses or severities X_k occurring in a given time period are summed. To apply a risk measure like

VaR we need to make assumptions about the (X_k) and N, which leads us to one of the fundamental concepts of (non-life) insurance mathematics.

Definition 13.1 (total loss amount and distribution). Denote by $N(t)$ the (random) number of losses over a fixed time period $[0, t]$ and write $X_1, X_2, \ldots$ for the individual losses. The *total loss amount* (or *aggregate loss*) is defined as

$$S_{N(t)} = \sum_{k=1}^{N(t)} X_k, \tag{13.9}$$

with df $F_{S_{N(t)}}(x) = P(S_{N(t)} \leqslant x)$, the *total* (or *aggregate*) *loss df*. Whenever t is fixed, $t = 1$ say, we may drop the time index from the notation and simply write S_N and F_{S_N}.

Remark 13.2. The definition of (13.9) as an rv is to be understood as $S_{N(t)}(\omega) = \sum_{k=1}^{N(t)(\omega)} X_k(\omega)$, $\omega \in \Omega$, and is referred to as a random (or randomly indexed) sum.

A prime goal of this section will be the analytical and numerical calculation of F_{S_N}, which requires further assumptions about the (X_k) and N.

Assumption 13.3 (independence, compound sums). *We assume that the rvs (X_k) are iid with common df G, $G(0) = 0$. We further assume that the rvs N and (X_k) are independent; in that case we refer to (13.9) as a compound sum. The probability mass function of N is denoted by $p_N(k) = P(N = k)$, $k = 0, 1, 2, \ldots$. The rv N is referred to as the compounding rv.*

Proposition 13.4 (compound distribution). *Let S_N be a compound sum and suppose that Assumption 13.3 holds. Then, for all $x \geqslant 0$,*

$$F_{S_N}(x) = P(S_N \leqslant x) = \sum_{k=0}^{\infty} p_N(k) G^{(k)}(x), \tag{13.10}$$

where $G^{(k)}(x) = P(S_k \leqslant x)$, the kth convolution of G. Note that $G^{(0)}(x) = 1$ for $x \geqslant 0$, and $G^{(0)}(x) = 0$ for $x < 0$.

Proof. Suppose that $x \geqslant 0$. Then

$$F_{S_N}(x) = \sum_{k=0}^{\infty} P(S_N \leqslant x \mid N = k) P(N = k) = \sum_{k=0}^{\infty} p_N(k) G^{(k)}(x).$$

□

Although formula (13.10) is explicit, its actual calculation in specific cases is difficult because the convolution powers $G^{(k)}$ of a df G are not generally available in closed form. One therefore resorts to (numerical) approximation methods. A first class of these uses the fact that the Laplace–Stieltjes transform of a convolution is the product of the Laplace–Stieltjes transforms. Using the usual notation

$\hat{F}(s) = \int_0^\infty \mathrm{e}^{-sx}\,\mathrm{d}F(x)$, where $s \geqslant 0$ for Laplace–Stieltjes transforms, we have that $\widehat{G^{(k)}}(s) = (\hat{G}(s))^k$. It follows from Proposition 13.4 that

$$\hat{F}_{S_N}(s) = \sum_{k=0}^{\infty} p_N(k)\hat{G}^k(s) = \Pi_N(\hat{G}(s)), \quad s \geqslant 0, \tag{13.11}$$

where Π_N denotes the *probability-generating function* of N, defined by $\Pi_N(s) = \sum_{k=1}^{\infty} p_N(k)s^k$.

Example 13.5 (the compound Poisson df). Suppose that N has a Poisson df with intensity parameter $\lambda > 0$, denoted by $N \sim \mathrm{Poi}(\lambda)$. In that case, $p_N(k) = \mathrm{e}^{-\lambda}\lambda^k/k!$, $k \geqslant 0$, and, for $s > 0$,

$$\Pi_N(s) = \sum_{k=0}^{\infty} \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}s^k = \mathrm{e}^{-\lambda(1-s)}.$$

From (13.11) it therefore follows that, for $s \geqslant 0$,

$$\hat{F}_{S_N}(s) = \exp(-\lambda(1 - \hat{G}(s))).$$

In this case, the df of S_N is referred to as the *compound Poisson df* and we write $S_N \sim \mathrm{CPoi}(\lambda, G)$.

Example 13.6 (the compound negative binomial df). Suppose that N has a negative binomial df with parameters $\alpha > 0$ and $0 < p < 1$, denoted by $N \sim \mathrm{NB}(\alpha, p)$. The probability mass function is given by (A.18) and we get, for $0 < s < (1 - p)^{-1}$,

$$\Pi_N(s) = \sum_{k=0}^{\infty} \binom{\alpha + k - 1}{k} p^\alpha (1 - p)^k s^k = \left(\frac{p}{1 - s(1 - p)}\right)^\alpha.$$

From (13.11) it therefore follows that, for $s \geqslant 0$,

$$\hat{F}_{S_N}(s) = \left(\frac{p}{1 - \hat{G}(s)(1 - p)}\right)^\alpha.$$

In this case, the df of S_N is referred to as the *compound negative binomial df* and we write $S_N \sim \mathrm{CNB}(\alpha, p, G)$.

Formula (13.11) facilitates the calculation of moments of S_N and lends itself to numerical evaluation through Fourier inversion, using a technique known as the fast Fourier transform (FFT) (see Notes and Comments for references on the latter). For the calculation of moments, note that, under the assumption of the existence of sufficiently high moments and hence differentiability of $\hat{G}$ and Π_N, we obtain

$$\frac{\mathrm{d}^k}{\mathrm{d}s^k}\Pi_N(s)\bigg|_{s=1} = E(N(N-1)\cdots(N-k+1))$$

and

$$(-1)^k\frac{\mathrm{d}^k}{\mathrm{d}s^k}\hat{G}(s)\bigg|_{s=0} = E(X_1^k) = \mu_k.$$

Example 13.7 (continuation of Example 13.5). In the case of the compound Poisson df one obtains

$$E(S_N) = (-1)\frac{\mathrm{d}}{\mathrm{d}s}\hat{F}_{S_N}(s)\bigg|_{s=0} = \exp(-\lambda(1-\hat{G}(0)))\lambda(-\hat{G}'(0))$$
$$= \lambda\mu_1 = E(N)E(X_1).$$

Similar calculations yield $\operatorname{var}(S_N) = E(S_N^2) - (E(S_N))^2 = \lambda\mu_2$.

For the general compound case one obtains the following useful result.

Proposition 13.8 (moments of compound dfs). *Under Assumption 13.3 and assuming that $E(N) < \infty$, $\mu_2 < \infty$, we have that*

$$E(S_N) = E(N)E(X_1) \quad \textit{and} \quad \operatorname{var}(S_N) = \operatorname{var}(N)(E(X_1))^2 + E(N)\operatorname{var}(X_1). \tag{13.12}$$

Proof. This follows readily from (13.11), differentiating with respect to s. The following direct proof avoids the use of transforms. Conditioning on N and using Assumption 13.3 one obtains

$$E(S_N) = E(E(S_N \mid N)) = E\bigg(E\bigg(\sum_{k=1}^{N} X_k \;\bigg|\; N\bigg)\bigg)$$
$$= E\bigg(\sum_{k=1}^{N} E(X_k)\bigg) = E(N)E(X_1)$$

and, similarly,

$$E(S_N^2) = E\bigg(E\bigg(\bigg(\sum_{k=1}^{N} X_k\bigg)^2 \;\bigg|\; N\bigg)\bigg) = E\bigg(E\bigg(\sum_{k=1}^{N}\sum_{\ell=1}^{N} X_k X_\ell \;\bigg|\; N\bigg)\bigg)$$
$$= E(N\mu_2 + N(N-1)\mu_1^2) = E(N)\mu_2 + (E(N^2) - E(N))\mu_1^2$$
$$= E(N)\operatorname{var}(X_1) + E(N^2)(E(X_1))^2,$$

so $\operatorname{var}(S_N) = E(S_N^2) - (E(S_N))^2 = E(N)\operatorname{var}(X_1) + \operatorname{var}(N)(E(X_1))^2$. □

Remark 13.9. Formula (13.12) elegantly combines the randomness of the frequency ($\operatorname{var}(N)$) with that of the severity ($\operatorname{var}(X_1)$). In the compound Poisson case it reduces to the formula $\operatorname{var}(S_N) = \lambda E(X_1^2) = \lambda\mu_2$, as in Example 13.7. In the deterministic-sum case, when $P(N = n) = 1$, say, we find the well-known results $E(S_N) = n\mu_1$ and $\operatorname{var}(S_N) = n\operatorname{var}(X_1)$; indeed, in this degenerate case, $\operatorname{var}(N) = 0$.

The compound Poisson model is a basic model for aggregate financial or insurance risk losses. The ubiquitousness of the Poisson distribution in insurance can be understood as follows. Consider a time interval [0, 1] and let N denote the total number of losses in that interval. Suppose further that we have a number of potential loss generators (transactions, credit positions, insurance policies, etc.) that can produce, with probability p_n, one loss or, with probability $1 - p_n$, no loss in each

small subinterval $((k-1)/n, k/n]$ for $k = 1, \ldots, n$. Moreover, suppose that the occurrence or non-occurrence of a loss in any particular subinterval is not influenced by the occurrence of losses in other intervals. The number N_n of losses then has a binomial df with parameters n and p_n, so

$$P(N_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k}, \quad k = 0, \ldots, n.$$

Combined with a loss-severity distribution this frequency distribution gives rise, in (13.10), to the so-called *binomial loss model*. Next suppose that $n \to \infty$ in such a way that $\lim_{n\to\infty} np_n = \lambda > 0$. It follows from *Poisson's theorem of rare events* (see also Section 5.3.1) that

$$\lim_{n\to\infty} P(N_n = k) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots,$$

i.e. $N_\infty \sim \mathrm{Poi}(\lambda)$, explaining why the Poisson model assumption is very natural as a frequency distribution and why the compound Poisson model is a common aggregate loss model. The compound Poisson model has several nice properties, one of which concerns aggregation and is useful in the operational risk context in situations such as (13.5).

Proposition 13.10 (sums of compound Poisson rvs). *Suppose that the compound sums $S_{N_i} \sim \mathrm{CPoi}(\lambda_i, G_i)$, $i = 1, \ldots, d$, and that these rvs are independent. Then $S_N = \sum_{i=1}^d S_{N_i} \sim \mathrm{CPoi}(\lambda, G)$, where $\lambda = \sum_{i=1}^d \lambda_i$ and $G = \sum_{i=1}^d (\lambda_i/\lambda) G_i$.*

Proof. (For $d = 2$, the general case being similar.) Because of independence and Example 13.5 we have, for the Laplace–Stieltjes transform of S_N,

$$\begin{aligned}
\hat{F}_{S_N}(s) &= \hat{F}_{S_{N_1}}(s) \hat{F}_{S_{N_2}}(s) \\
&= \exp\left(-(\lambda_1+\lambda_2)\left(1 - \frac{1}{\lambda_1+\lambda_2}(\lambda_1 \hat{G}_1(s) + \lambda_2 \hat{G}_2(s))\right)\right) \\
&= \exp(-\lambda(1-\hat{G}(s))),
\end{aligned}$$

where $\lambda = \lambda_1 + \lambda_2$ and

$$G = \frac{\lambda_1}{\lambda_1+\lambda_2} G_1 + \frac{\lambda_2}{\lambda_1+\lambda_2} G_2.$$

The result follows since the Laplace–Stieltjes transform uniquely determines the underlying df. □

The new intensity λ is just the sum of the old ones, whereas the new severity df G is a *discrete mixture* of the loss dfs G_i with weights λ_i/λ, $i = 1, \ldots, d$. We can easily simulate losses from such a model through a two-stage procedure: first draw i ($i = 1, \ldots, d$) with probability λ_i/λ, and then draw a loss with df G_i.

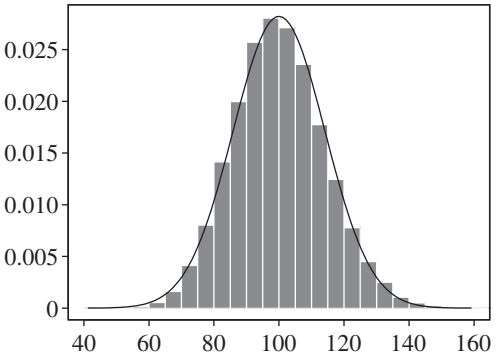

Figure 13.3. Histogram of simulated compound loss data ($n = 100\,000$) for $S_N \sim \text{CPoi}(100, \text{Exp}(1))$ together with normal approximation (13.14).

Beyond the Poisson model. The Poisson model can serve as a stylized representation of the loss-generating mechanism from which more realistic models can be derived. For instance, we may wish to introduce a time parameter in N to capture different occurrence patterns over time (see Section 13.2.6). Also, the intensity parameter λ may be assumed to be random (see Example 13.21). Indeed, a further step is to turn λ into a stochastic process, which gives rise to such models as doubly stochastic (or Cox) processes (see Section 10.5.1) or self-exciting processes, as encountered in Section 16.2.1. Furthermore, various forms of dependence among the X_k rvs or between N and (X_k) could be modelled. Finally, multiline portfolios require multivariate models for vectors of the type $(S_{N_1}, \ldots, S_{N_d})$. An ultimate goal of the AM approach to operational risk would be to model such random vectors where, for instance, d might stand for seven risk types, eight business lines, or a total of 56 loss category cells.

13.2.3 Approximations and Panjer Recursion

As mentioned in Section 13.2.2, the analytic calculation of F_{S_N} is not possible for the majority of reasonable models, which has led actuaries to come up with several numerical approximations. Below we review some of these approximations and illustrate their use for several choices of the severity df G. The basic example we look at is the compound Poisson case, $S_N \sim \text{CPoi}(\lambda, G)$, though most of the approximations discussed can be adjusted to deal with other distributions for N. Given λ and G we can easily simulate F_{S_N} and, by repeating this many times, we can get an empirical estimate that is close to the true df. Figure 13.3 contains a simulation of $n = 100\,000$ realizations of $S_N \sim \text{CPoi}(100, \text{Exp}(1))$. Although the histogram exhibits mild skewness (which can easily be shown theoretically (see (13.15))), a clear central limit effect takes place. This is used in the first approximation below.

Normal approximation. As the loss rvs X_i are iid (with finite second moment, say) and S_N is a (random) sum of the X_i variables, one can apply Theorem 2.5.16 from Embrechts, Klüppelberg and Mikosch (1997) and Proposition 13.8 to obtain

the following approximation, for general N:

$$F_{S_N}(x) \approx \Phi\left(\frac{x - E(N)E(X_1)}{\sqrt{\mathrm{var}(N)(E(X_1))^2 + E(N)\,\mathrm{var}(X_1)}}\right). \tag{13.13}$$

Here, and in the approximations below, "$\approx$" has no specific mathematical interpretation beyond "there exists a limit result justifying the right-hand side to be used as approximation of the left-hand side". In particular, for the compound Poisson case above, (13.13) reduces to

$$F_{S_N}(x) \approx \Phi\left(\frac{x - 100}{\sqrt{200}}\right), \tag{13.14}$$

where Φ is the standard normal df, as usual. It is this normal approximation that is superimposed on the histogram in Figure 13.3. Clearly, there are conditions that must be satisfied in order to obtain the approximation (13.13): for example, claims should not be too heavy tailed (see Theorem 13.22).

For CPoi(λ, G) it is not difficult to show that the skewness parameter satisfies

$$\frac{E((S_N - E(S_N))^3)}{(\mathrm{var}(S_N))^{3/2}} = \frac{E(X_1^3)}{\sqrt{\lambda(E(X_1^2))^3}} > 0 \tag{13.15}$$

(note that $X_1 \geqslant 0$ almost surely), so an approximation by a df with positive skewness may improve the approximation (13.14), especially in the tail area. This is indeed the case and leads to the next approximation.

Translated-gamma approximation. We approximate S_N by $k + Y$, where k is a translation parameter and $Y \sim \mathrm{Ga}(\alpha, \beta)$ has a gamma distribution (see Section A.2.4). The parameters (k, α, β) are found by matching the mean, the variance and the skewness of $k + Y$ and S_N. It is not difficult to check that the following equations result:

$$k + \frac{\alpha}{\beta} = \lambda E(X_1), \qquad \frac{\alpha}{\beta^2} = \lambda E(X_1^2), \qquad \frac{2}{\sqrt{\alpha}} = \frac{E(X_1^3)}{\sqrt{\lambda(E(X_1^2))^3}}.$$

In our case, where $\lambda = 100$ and X_1 has a standard exponential distribution, these yield the equations $k + \alpha/\beta = 100$, $\alpha/\beta^2 = 200$ and $2/\sqrt{\alpha} = 0.2121$ with solution $\alpha = 88.89$, $\beta = 0.67$, $k = -32.72$.

Commentary on these approximations. Both approximations work reasonably well for the bulk of the data. However, for risk-management purposes we are mainly interested in upper tail risk; in Figure 13.4 we have therefore plotted both approximations for $x \geqslant 120$ on a log–log scale. This corresponds to the tail area beyond the 90% quantile of F_{S_N}. Similar plots were routinely used in Chapter 5 on EVT (see, for example, Figure 5.6). It becomes clear that, as can be expected, the gamma approximation works better in this upper tail area where the normal approximation underestimates the loss potential.

Of course, for loss data with heavier tails than exponential (lognormal or Pareto, say), even the translated-gamma approximation will be insufficient, and other

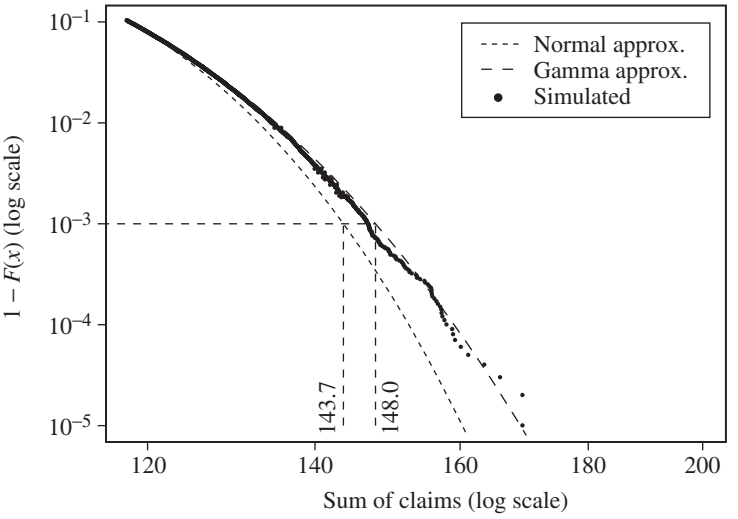

Figure 13.4. Simulated CPoi(100, Exp(1)) data together with normal and translated-gamma approximations (log–log scale). The 99.9% quantile estimates are also given.

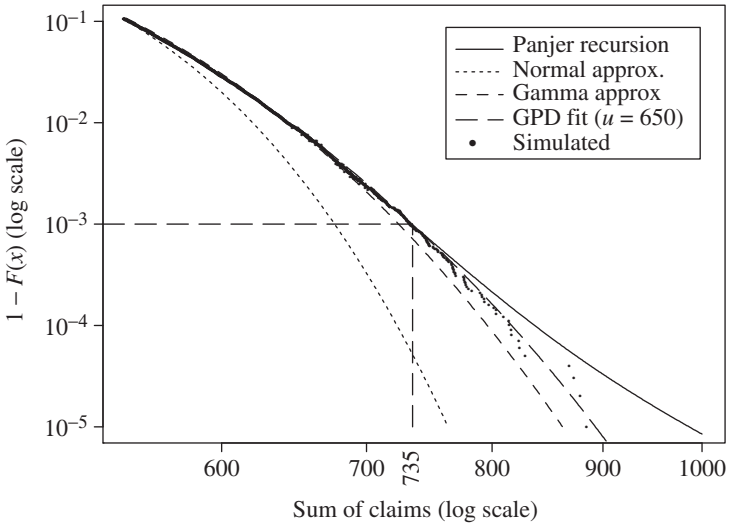

Figure 13.5. Simulated CPoi(100, LN(1, 1)) data ($n = 100\,000$) with normal, translated-gamma, GPD and Panjer recursion (see Example 13.18) approximations (log–log scale).

approximations can be devised based on heavier-tailed distributions, such as translated F, inverse gamma or generalized Pareto.

Another approach could be based on Monte Carlo simulation of aggregate losses S_N to which an appropriate heavy-tailed loss distribution could then be fitted. One possible approach would be to model the tail of these simulated compound losses with the GPD using the methodology of Section 5.2.2. This is what has been done in Figures 13.5 and 13.6, where we have plotted various approximations

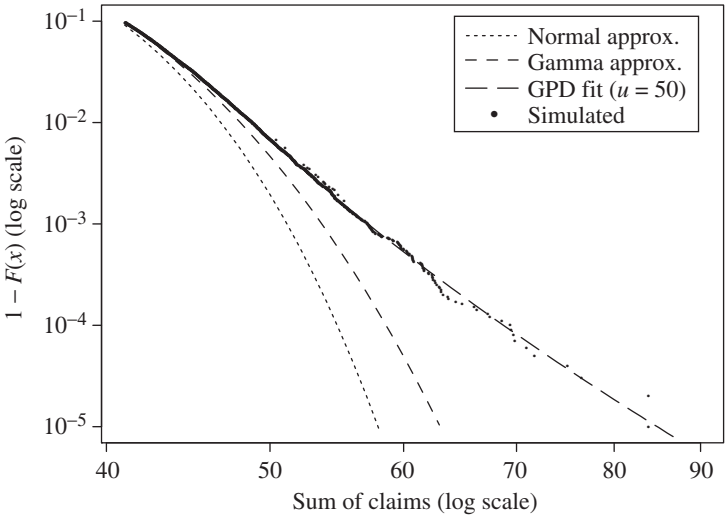

Figure 13.6. Simulated CPoi(100, Pa(4, 1)) data ($n = 100\,000$) with normal, translated-gamma, and GPD approximations (log–log scale).

for CPoi(100, LN(1, 1)) and CPoi(100, Pa(4, 1)). The former corresponds to a standard industry model for operational risk (see Frachot 2004). The latter corresponds to a class of operational risk models used in Moscadelli (2004). The message of these figures is clear: if the data satisfy the compound Poisson assumption, then the GPD yields a superior fit for high quantiles.

We now turn to an important class of approximations based on recursive methods. In the case where the loss sizes (X_i) are discrete and the distribution function of N satisfies a specific condition (see Definition 13.11 below) a reliable recursive method can be worked out.

Suppose that X_1 has a discrete distribution so that $P(X_1 \in \mathbb{N}_0) = 1$ with $g_k = P(X_1 = k)$, $p_k = P(N = k)$ (for notational convenience we write p_k for $p_N(k)$) and $s_k = P(S_N = k)$. For simplicity assume that $g_0 = 0$ and let

$$g_k^{(n)} = P(X_1 + \cdots + X_n = k),$$

the discrete convolution of the probability mass function g_k. Note that, by definition, $g_k^{(n+1)} = \sum_{i=1}^{k-1} g_i^{(n)} g_{k-i}$. We immediately obtain the following identities:

$$\left.\begin{aligned} s_0 &= P(S_N = 0) = P(N = 0) = p_0, \\ s_n &= P(S_N = n) = \sum_{k=1}^{\infty} p_k g_n^{(k)}, \quad n \geqslant 1, \end{aligned}\right\} \tag{13.16}$$

where the latter formula corresponds to Proposition 13.4 but now in the discrete case. As in Proposition 13.4 we note that (13.16) is difficult to calculate, mainly due to the convolutions $g_n^{(k)}$. However, for an important class of counting variables N, (13.16) can be reduced to a simple recursion. For this we introduce the so-called *Panjer classes*.

Definition 13.11 (Panjer class). The probability mass function (p_k) of N belongs to the Panjer(a, b) class for some $a, b \in \mathbb{R}$ if the following relationship holds for $r \geqslant 1$: $p_r = (a + (b/r))p_{r-1}$.

Example 13.12 (binomial). If $N \sim \mathrm{B}(n, p)$, then its probability mass function is $p_r = \binom{n}{r} p^r (1-p)^{n-r}$ for $0 \leqslant r \leqslant n$ and it can be easily checked that

$$\frac{p_r}{p_{r-1}} = -\frac{p}{1-p} + \frac{(n+1)p}{r(1-p)},$$

showing that N belongs to the Panjer(a, b) class with $a = -p/(1-p)$ and $b = (n+1)p/(1-p)$.

Example 13.13 (Poisson). If $N \sim \mathrm{Poi}(\lambda)$, then its probability mass function $p_r = \mathrm{e}^{-\lambda}\lambda^r/r!$ satisfies $p_r/p_{r-1} = \lambda/r$, so N belongs to the Panjer(a, b) class with $a = 0$ and $b = \lambda$.

Example 13.14 (negative binomial). If N has a negative binomial distribution, denoted by $N \sim \mathrm{NB}(\alpha, p)$, then its probability mass function is

$$p_r = \binom{\alpha + r - 1}{r} p^\alpha (1-p)^r, \quad r \geqslant 0,\ \alpha > 0,\ 0 < p < 1$$

(see Section A.2.7 for further details). We can easily check that

$$\frac{p_r}{p_{r-1}} = 1 - p + \frac{(\alpha - 1)(1-p)}{r}.$$

Hence N belongs to the Panjer(a, b) class with $a = 1 - p$ and $b = (\alpha - 1)(1 - p)$. In Proposition 13.21 we will show that the negative binomial model follows very naturally from the Poisson model when one randomizes the intensity parameter of the latter using a gamma distribution.

Remark 13.15. One can show that, neglecting degenerate models for (p_k), the above three examples are the *only* counting distributions satisfying Definition 13.11. This result goes back to Johnson and Kotz (1969) and was formulated explicitly in the actuarial literature in Sundt and Jewell (1982).

Theorem 13.16 (Panjer recursion). *Suppose that N satisfies the* Panjer(a, b) *class condition and that $g_0 = P(X_1 = 0) = 0$, then $s_0 = p_0$ and, for $r \geqslant 1$, $s_r = \sum_{i=1}^r (a + (bi/r)) g_i s_{r-i}$.*

Proof. We already know that $s_0 = p_0$ from (13.16), so suppose that $r \geqslant 1$. Noting that $X_1, \ldots, X_n$ are iid, we require the following well-known identity for exchangeable rvs:

$$\begin{aligned} E\left(X_1 \,\middle|\, \sum_{i=1}^n X_i = r\right) &= \frac{1}{n}\sum_{j=1}^n E\left(X_j \,\middle|\, \sum_{i=1}^n X_i = r\right) \\ &= \frac{1}{n} E\left(\sum_{j=1}^n X_j \,\middle|\, \sum_{i=1}^n X_i = r\right) = \frac{r}{n}. \end{aligned} \tag{13.17}$$

Moreover, using the fact that $g_0^{(n-1)} = 0$ for $n \geqslant 2$, we make the preliminary calculation that

$$\begin{aligned}
p_{n-1} \sum_{i=1}^{r-1} \left(a + \frac{bi}{r}\right) g_i g_{r-i}^{(n-1)} &= p_{n-1} \sum_{i=1}^{r} \left(a + \frac{bi}{r}\right) g_i g_{r-i}^{(n-1)} \\
&= p_{n-1} \sum_{i=1}^{r} \left(a + \frac{bi}{r}\right) P\left(X_1 = i,\ \sum_{j=2}^{n} X_j = r - i\right) \\
&= p_{n-1} \sum_{i=1}^{r} \left(a + \frac{bi}{r}\right) P\left(X_1 = i,\ \sum_{j=1}^{n} X_j = r\right) \\
&= p_{n-1} \sum_{i=1}^{r} \left(a + \frac{bi}{r}\right) P\left(X_1 = i \,\middle|\, \sum_{j=1}^{n} X_j = r\right) g_r^{(n)} \\
&= p_{n-1} E\left(a + \frac{bX_1}{r} \,\middle|\, \sum_{j=1}^{n} X_j = r\right) g_r^{(n)} \\
&= p_{n-1} \left(a + \frac{b}{n}\right) g_r^{(n)} = p_n g_r^{(n)},
\end{aligned}$$

where (13.17) is used in the final step. Therefore, the identity (13.16) yields

$$\begin{aligned}
s_r &= \sum_{n=1}^{\infty} p_n g_r^{(n)} = p_1 g_r + \sum_{n=2}^{\infty} p_n g_r^{(n)} \\
&= (a+b) p_0 g_r + \sum_{n=2}^{\infty} \sum_{i=1}^{r-1} \left(a + \frac{bi}{r}\right) g_i p_{n-1} g_{r-i}^{(n-1)} \\
&= (a+b) s_0 g_r + \sum_{i=1}^{r-1} \left(a + \frac{bi}{r}\right) g_i \sum_{n=2}^{\infty} p_{n-1} g_{r-i}^{(n-1)} \\
&= (a+b) g_r s_0 + \sum_{i=1}^{r-1} \left(a + \frac{bi}{r}\right) g_i s_{r-i} \\
&= \sum_{i=1}^{r} \left(a + \frac{bi}{r}\right) g_i s_{r-i}.
\end{aligned}$$

□

Remark 13.17. In the case of both the FFT method and the Panjer recursion, an initial discretization of the loss df G generally has to be made, which introduces an approximation error. An in-depth discussion of discretization errors for the computation of compound distributions is to be found in Grübel and Hermesmeier (1999, 2000) (see also references therein for a comparison of these approaches). A slight correction to Theorem 13.16 has to be made if $g_0 = P(X_1 = 0) > 0$. One obtains $s_0 = \sum_{k=0}^{\infty} p_k g_0^k$ and, for $r \geqslant 1$, $s_r = (1 - a g_0)^{-1} \sum_{i=1}^{r} (a + bi/r) g_i s_{r-i}$ (see Mikosch 2004, Theorem 3.3.10). We give further references in Notes and Comments.

Example 13.18 (Panjer recursion for the CPoi(100, LN(1, 1)) case). In Figure 13.5 we have included the Panjer approximation for the CPoi(100, LN(1, 1)) case. In order to apply Theorem 13.16 we first have to discretize the lognormal df. An equispaced discretization of about 0.5 yields the Panjer approximation in Figure 13.5, which is excellent for quantile values around 0.999, relevant for applications. The 99.9% quantile estimate based on the Panjer recursion is 735, a value very close to the GPD estimate. Far out in the tail, beyond 0.999, say, rounding errors become important (the tail drifts off) and one has to be more careful; we give some references in Notes and Comments on how to improve recursive methods far out in the tail.

13.2.4 Poisson Mixtures

Poisson mixture models have been used in both credit and operational risk modelling; for an example in the latter case see Cruz (2002, Section 5.2.2) as well as that book's jacket, which features a negative binomial distribution (a particular Poisson mixture model). Poisson mixtures have been used by actuaries for a long time; the negative binomial made its first appearance in the actuarial literature as the distribution of the number of repeated accidents suffered by an individual in a given time span (see Seal 1969).

In Example 13.5 we introduced the compound Poisson model CPoi(λ, G), where $N \sim \text{Poi}(\lambda)$ counts the number of losses and G is the loss severity df. One disadvantage of the Poisson frequency distribution is that $\text{var}(N) = \lambda = E(N)$, whereas count data often exhibit so-called *overdispersion*, meaning that they indicate a model where $\text{var}(N) > E(N)$. A standard way to achieve this is by mixing the intensity λ over some df $F_\Lambda(\lambda)$, i.e. assume that $\lambda > 0$ is a realization of a positive rv Λ with this df so that, by definition,

$$\begin{aligned} p_N(k) = P(N = k) &= \int_0^\infty P(N = k \mid \Lambda = \lambda)\, \mathrm{d}F_\Lambda(\lambda) \\ &= \int_0^\infty \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}\, \mathrm{d}F_\Lambda(\lambda). \end{aligned} \tag{13.18}$$

Definition 13.19 (the mixed Poisson distribution). The rv N with df (13.18) is called a *mixed Poisson* rv with *structure* (or *mixing*) distribution F_Λ.

A consequence of the next result is that mixing leads to overdispersion.

Proposition 13.20. *Suppose that N is mixed Poisson with structure df F_Λ. Then $E(N) = E(\Lambda)$ and $\text{var}(N) = E(\Lambda) + \text{var}(\Lambda)$, i.e. for Λ non-degenerate, N is overdispersed.*

Proof. One immediately obtains

$$E(N) = \sum_{k=0}^\infty k p_N(k) = \int_0^\infty \sum_{k=0}^\infty k \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}\, \mathrm{d}F_\Lambda(\lambda) = \int_0^\infty \lambda\, \mathrm{d}F_\Lambda(\lambda) = E(\Lambda).$$

And, similarly,

$$E(N^2) = \sum_{k=0}^{\infty} k^2 p_N(k) = E(\Lambda) + E(\Lambda^2),$$

so the result follows. □

We now give a concrete example of a mixed Poisson distribution, which is particularly important in both operational risk and credit risk modelling. Indeed we have already used the following result when describing the industry credit risk model CreditRisk$^+$ in Section 11.2.5.

Proposition 13.21 (negative binomial as Poisson mixture). *Suppose that the rv N has a mixed Poisson distribution with a gamma-distributed mixing variable $\Lambda \sim \mathrm{Ga}(\alpha, \beta)$. Then N has a negative binomial distribution $N \sim \mathrm{NB}(\alpha, \beta/(\beta+1))$.*

Proof. Using the definition of a gamma distribution in Section A.2.4 we have

$$P(N=k) = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} \lambda^{\alpha-1} \mathrm{e}^{-\beta\lambda}\, \mathrm{d}\lambda = \frac{\beta^\alpha}{k!\Gamma(\alpha)} \int_0^\infty \lambda^{\alpha+k-1} \mathrm{e}^{-(\beta+1)\lambda}\, \mathrm{d}\lambda.$$

Substituting $u = (\beta+1)\lambda$, the integral can be evaluated to be

$$\int_0^\infty (\beta+1)^{-(\alpha+k)} u^{\alpha+k-1} \mathrm{e}^{-u}\, \mathrm{d}u = \frac{\Gamma(\alpha+k)}{(\beta+1)^{\alpha+k}}.$$

This yields

$$P(N=k) = \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^k \frac{\Gamma(\alpha+k)}{k!\Gamma(\alpha)}.$$

Using the relation $\Gamma(\alpha+k) = (\alpha+k-1)\cdots\alpha\Gamma(\alpha)$, we see that this is equal to the probability mass function of a negative binomial rv with $p := \beta/(\beta+1)$ (see Section A.2.7). □

Recall the definition of compound sums from Section 13.2.2 (Assumption 13.3 and Proposition 13.4). In the special case of mixed Poisson rvs, compounding leads to so-called *compound mixed Poisson distributions*, such as the compound negative binomial distribution of Example 13.6. There is much literature on dfs of this type (see Notes and Comments).

13.2.5 Tails of Aggregate Loss Distributions

In Section 5.1.2 we defined the class of rvs with regularly varying or power tails. If the (claim size) df G is regularly varying with index $\alpha > 0$, then there exists a slowly varying function L (Definition 5.7) such that $\bar{G}(x) = 1 - G(x) = x^{-\alpha} L(x)$. The next result shows that, for a wide class of counting dfs $(p_N(k))$, the df of the compound sum S_N, F_{S_N}, inherits the power-like behaviour of G.

Theorem 13.22 (power-like behaviour of compound-sum distribution). *Suppose that S_N is a compound sum with $E(N) = \lambda$ and suppose that there exists an $\varepsilon > 0$*

such that $\sum_{k=0}^{\infty}(1+\varepsilon)^k p_N(k) < \infty$. *If* $\bar{G}(x) = x^{-\alpha}L(x)$ *with* $\alpha \geqslant 0$ *and* L *slowly varying, then*

$$\lim_{x\to\infty} \frac{\bar{F}_{S_N}(x)}{\bar{G}(x)} = \lambda,$$

so $\bar{F}_{S_N}$ *inherits the power-like behaviour of* $\bar{G}$.

Proof. This result holds more generally for *subexponential* dfs; a proof together with further discussion can be found in Embrechts, Klüppelberg and Mikosch (1997, Section 1.3.3). □

Example 13.23 (negative binomial). It is not difficult to show that the negative binomial case satisfies the condition on N in Theorem 13.22. The kind of argument that is required is to be found in Embrechts, Klüppelberg and Mikosch (1997, Example 1.3.11). Hence, if $\bar{G}(x) = x^{-\alpha}L(x)$, the tail of the compound-sum df behaves like the tail of G, i.e.

$$\bar{F}_{S_N}(x) \sim \frac{\alpha}{\beta}\bar{G}(x), \quad \text{as } x \to \infty.$$

(For details, see Embrechts, Klüppelberg and Mikosch (1997, Section 1.3.3).)

Under the conditions of Theorem 13.22 the asymptotic behaviour of $\bar{F}_{S_N}(x)$ in the case of a Pareto loss df is again Pareto with the same index. This is clearly seen in Figure 13.6 in the linear behaviour of the simulated losses as well as the fitted GPD. In the case of Figure 13.5, one can show that $\bar{F}_{S_N}(x)$ decays like a lognormal tail; see the reference given in the proof of Theorem 13.22 for details. Note that the GPD is able to pick up the features of the tail in both cases.

13.2.6 The Homogeneous Poisson Process

In the previous sections we looked at counting rvs N over a fixed time interval $[0, 1]$, say. Without any additional difficulty we could have looked at $N(t)$, counting the number of events in $[0, t]$ for $t \geqslant 0$. In the Poisson case this would correspond to $N(t) \sim \text{Poi}(\lambda t)$; hence, for fixed t and on replacing λ by λt, all of the previous results concerning $\text{Poi}(\lambda)$ rvs can be suitably adapted.

In this section we want to integrate the rvs $N(t)$, $t \geqslant 0$, into a stochastic process framework. The less mathematically trained reader should realize that there is a big difference between a family of rvs indexed by time, for which we only specify the one-dimensional dfs (which is what we have done so far), and a stochastic process with a specific structure in which these rvs are embedded. This difference is akin to the difference between marginal and joint distributions, a topic we have highlighted as very important in Chapter 7 through the notion of copulas; of course, in the stochastic process case, there also has to be some probabilistic consistency across time. In a certain sense, the finite-dimensional problem of Chapter 7 becomes an infinite-dimensional problem.

After these words of warning on the difference between rvs and stochastic processes, we now take some methodological shortcuts to arrive at our goal. The interested reader wanting to learn more will have to delve deeper into the mathematical

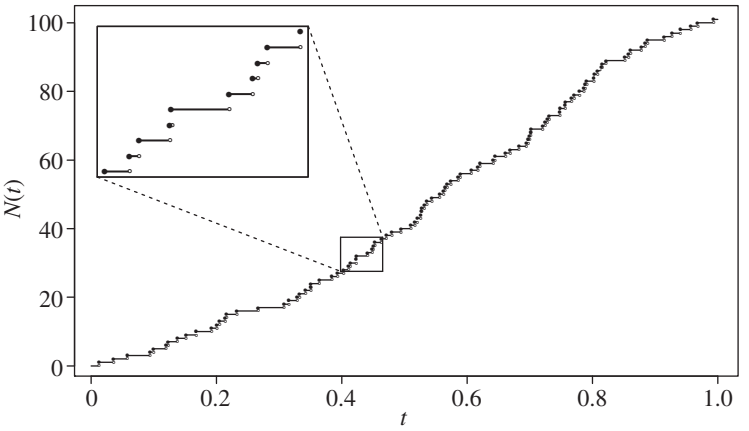

Figure 13.7. A sample path of a counting process.

background of stochastic processes in general and counting processes in particular. The Notes and Comments section contains some references.

Definition 13.24 (counting processes). A stochastic process $N = (N(t))_{t \geqslant 0}$ is a *counting process* if its sample paths are right continuous with left limits existing and if there exists a sequence of rvs $T_0 = 0, T_1, T_2, \ldots$ tending almost surely to ∞ such that $N(t) = \sum_{k=1}^{\infty} I_{\{T_k \leqslant t\}}$.

A typical realization of such a process is given in Figure 13.7. We now define the homogeneous Poisson process as a special counting process.

Definition 13.25 (homogeneous Poisson process). A stochastic process $N = (N(t))_{t \geqslant 0}$ is a *homogeneous Poisson process* with intensity (rate) $\lambda > 0$ if the following properties hold:

(i) N is a counting process;

(ii) $N(0) = 0$, almost surely;

(iii) N has stationary and independent increments; and

(iv) for each $t > 0$, $N(t) \sim \text{Poi}(\lambda t)$.

Remark 13.26. Note that conditions (iii) and (iv) imply that, for $0 < u < v < t$, the rvs $N(v) - N(u)$ and $N(t) - N(v)$ are independent and that, for $k \geqslant 0$,

$$\begin{aligned} P(N(v) - N(u) = k) &= P(N(v-u) = k) \\ &= \mathrm{e}^{-\lambda(v-u)} \frac{(\lambda(v-u))^k}{k!}. \end{aligned}$$

The rv $N(v) - N(u)$ counts the number of events (claims, losses) in the interval $(u, v]$; by stationarity, it has the same df as $N(v-u)$. In Figure 13.8 we have generated ten realizations of a homogeneous Poisson process on $[0, 1]$ with $\lambda = 100$. Note the rather narrow band within which the various sample paths fall.

For practical purposes, the following result contains the main properties of the homogeneous Poisson process.

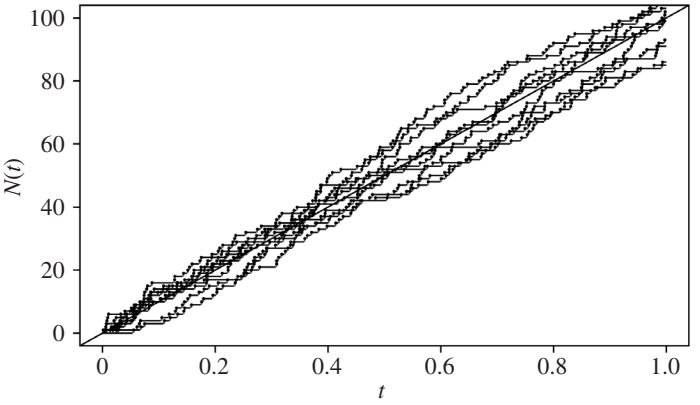

Figure 13.8. Ten realizations of a homogeneous Poisson process with $\lambda = 100$.

Theorem 13.27 (characterizations of the homogeneous Poisson process). *Suppose that N is a counting process. The following statements are then equivalent:*

(1) *N is a homogeneous Poisson process with rate $\lambda > 0$;*

(2) *N has stationary and independent increments and*

$$P(N(t) = 1) = \lambda t + o(t), \quad \text{as } t \downarrow 0,$$
$$P(N(t) \geqslant 2) = o(t), \quad \text{as } t \downarrow 0;$$

(3) *the inter-event-times $(\Delta_k = T_k - T_{k-1})_{k \geqslant 1}$ are iid with distribution* $\mathrm{Exp}(\lambda)$*; and*

(4) *for all $t > 0$, $N(t) \sim \mathrm{Poi}(\lambda t)$ and, given that $N(t) = k$, the occurrence times $T_1, T_2, \ldots, T_k$ have the same distribution as the ordered sample from k independent rvs, uniformly distributed on $[0, t]$; as a consequence, we can write the conditional joint density as*

$$f_{T_1,\ldots,T_k|N(t)=k}(t_1, \ldots, t_k) = \frac{k!}{t^k} I_{\{0<t_1<\cdots<t_k<t\}}.$$

Proof. Many standard textbooks on stochastic processes contain proofs of this important theorem (see, for example, Mikosch 2004; Resnick 1992). □

Discussion. Statement (2) in Theorem 13.27 implies that λ can indeed be interpreted as a rate or intensity: $\lambda = \lim_{t \downarrow 0}(1/t)P(N(t) = 1)$. Moreover, the same statement implies that a homogeneous Poisson process does not allow for clustering of events: $\lim_{t \downarrow 0} P(N(t) \geqslant 2) = 0$. Statement (3) gives an event-time definition of a homogeneous Poisson process. It follows immediately that the first event-time has an $\mathrm{Exp}(\lambda)$ df: $P(T_1 > t) = P(N(t) = 0) = \mathrm{e}^{-\lambda t}$, $t \geqslant 0$. Statement (3), however, goes well beyond this by stating that the inter-event-times Δ_k are iid with $\Delta_k \sim \mathrm{Exp}(\lambda)$. This leads to a straightforward way of simulating a stream of loss events from a homogeneous Poisson process with rate λ. Moreover, this equivalent definition immediately yields a generalization by assuming that the Δ_k are still iid

but that $\Delta_k \sim F_\Delta$, a general df. The resulting process is a so-called *renewal process* (note that the only Markovian renewal process is the homogeneous Poisson process).

Finally, statement (4) yields an easy algorithm to generate the occurrences of homogeneous Poisson times over the interval $[0, t]$ given that we have a total of k events up to t—we simply generate k uniform rvs on $[0, t]$ and order them.

Multivariate Poisson processes. In many applications we want to model the frequencies of different loss types with a number of Poisson processes while considering possible dependence between loss frequencies for different loss types. More generally, we might want to construct a number of compound Poisson processes where loss severities for the different business lines were also dependent. A natural approach to modelling this dependence is to assume that all losses can be related to a series of underlying and independent Poisson *shock* processes. In insurance these shocks might be natural catastrophes; in credit risk modelling they might be a variety of economic events, such as local or global recessions; in operational risk modelling they might be the failure of various IT systems. When a shock occurs this may cause losses of several different types; the common shock causes the numbers of losses of each type to be dependent. See Lindskog and McNeil (2003), Pfeifer and Nešlehová (2004) and Chavez-Demoulin and Embrechts (2004) for models of this kind.

13.2.7 Processes Related to the Poisson Process

Using the fundamental building block of the homogeneous Poisson process, one can construct more general counting processes that are useful for loss-event modelling in finance and insurance. Such generalizations include the following.

Renewal processes (mentioned above). The exponential waiting time distribution is replaced by a general df F_Δ.

Inhomogeneous Poisson processes. The constant intensity λ is replaced by a deterministic function $\lambda(\cdot)$.

Mixed Poisson processes. The deterministic constant intensity λ is replaced by an rv Λ.

Doubly stochastic or Cox processes. λ is replaced by a stochastic process $\{\lambda_t : t \geqslant 0\}$ in accordance with notation used in Chapter 10 (see, for example, Definition 10.15).

Self-exciting or Hawkes processes. λ is replaced by a stochastic process depending only on previous event-times. See Section 16.2.1 for a concrete example.

Below, we highlight some features of some of these processes.

Inhomogeneous Poisson processes.

Definition 13.28 (inhomogeneous Poisson). A counting process N is an *inhomogeneous Poisson process* if, for some deterministic function $\lambda(s) \geqslant 0$, the following conditions hold:

(i) $N(0) = 0$, almost surely;

(ii) N has independent increments; and

(iii) for all $t \geqslant 0$,

$$P(N(t+h) - N(t) = 1) = \lambda(t)h + o(h), \quad h \downarrow 0,$$
$$P(N(t+h) - N(t) \geqslant 2) = o(h), \quad h \downarrow 0.$$

The function $\lambda(\cdot)$ is referred to as the *intensity* or *rate function*. The integral $\Lambda(t) = \int_0^t \lambda(s)\,\mathrm{d}s$ is referred to as the *intensity measure* (or *cumulative intensity function*).

Remark 13.29. A characterization theorem, similar to Theorem 13.27, can be derived. In particular, we find that, for $0 < s < t$, $N(t) - N(s) \sim \mathrm{Poi}(\Lambda(t) - \Lambda(s))$.

The inhomogeneous Poisson process is a useful tool in loss modelling whenever a deterministic trend or seasonality component is to be modelled in the loss frequency. The next example also shows that this process naturally emerges as a counting process for record losses.

Example 13.30 (records). The world of finance and insurance abounds with statements on *record events*: the largest single-day drop in the dollar/yen, the most expensive hurricane, the three best fund managers during the last year, the second largest loss due to internal fraud, the biggest one-day change in the credit spread of a particular company, etc. Likewise, the world of records is intimately related to the (general) theory of Poisson processes. In Notes and Comments we shall give several references for this. Below we indicate how an easy example related to a question on records leads to an inhomogeneous Poisson process as a model.

Suppose that the loss rvs $X_i \geqslant 0$ are iid with density function $f(x) > 0$, $x \geqslant 0$. Define the counting process N:

$$N(t) = \sum_{i=1}^{\infty} I_{\{X_i \leqslant t \text{ and } X_i > X_{i-j},\ j=1,\ldots,i-1\}}.$$

$N(t)$ counts the number of records in the sequence $(X_i)_{i\geqslant 1}$ of size less than t, and $(N(t))$ is referred to as the *record process*. It follows that, for $h, t > 0$,

$$\begin{aligned} P(N(t+h) - N(t) \geqslant 1) &= \sum_{i=1}^{\infty} P(X_i \in (t, t+h] \text{ and } X_{i-1} \leqslant t, \ldots, X_1 \leqslant t) \\ &= \sum_{i=1}^{\infty} (F(t+h) - F(t))(F(t))^{i-1} \\ &= \frac{F(t+h) - F(t)}{1 - F(t)} \\ &= \frac{f(t)}{1 - F(t)} h + o(h), \quad \text{as } h \downarrow 0. \end{aligned}$$

Moreover, for $h, t > 0$,

$$\begin{aligned} P(N(t+h) - N(t) \geqslant 2) &\leqslant \sum_{i<j} P(X_1 \leqslant t, \ldots, X_{i-1} \leqslant t,\ X_i \in (t, t+h], \\ &\qquad X_{i+1} \leqslant t+h, \ldots, X_{j-1} \leqslant t+h,\ X_j \in (t, t+h]) \\ &= \left(\int_t^{t+h} f(s)\,\mathrm{d}s \right)^2 \sum_{i<j} (F(t))^{i-1} (F(t))^{j-i-1} \\ &= o(h^2), \quad \text{as } h \downarrow 0. \end{aligned}$$

From these calculations one deduces that the record process N is inhomogeneous Poisson with rate function $\lambda(t) = f(t)/(1 - F(t))$, the so-called *hazard rate* of F, a notion that we encountered in Section 10.4.1.

Suppose now that, as in most practical cases, $\Lambda(t)$ is strictly increasing, so $\Lambda(\Lambda^{-1}(t)) = \Lambda^{-1}(\Lambda(t)) = t$. We can then always transform an inhomogeneous Poisson process N with intensity measure Λ into a homogeneous Poisson process with intensity 1 by a *change of time*.

Proposition 13.31 (time change, operational time). *Suppose that N is an inhomogeneous Poisson process with Λ strictly increasing, and define, for $t \geqslant 0$, $\tilde{N}(t) = N(\Lambda^{-1}(t))$. We then have that $\tilde{N}$ is homogeneous Poisson with intensity* 1.

Proof. For $t > 0$ fixed and $k \geqslant 0$,

$$P(\tilde{N}(t) = k) = P(N(\Lambda^{-1}(t)) = k) = \mathrm{e}^{-\Lambda(\Lambda^{-1}(t))} \frac{(\Lambda(\Lambda^{-1}(t)))^k}{k!} = \mathrm{e}^{-t} \frac{t^k}{k!},$$

so $\tilde{N}(t) \sim \mathrm{Poi}(t)$. By definition, the increments of $\tilde{N}$ are independent; moreover, for $0 < u < v$ we have that

$$\begin{aligned} P(\tilde{N}(v) - \tilde{N}(u) = k) &= P(N(\Lambda^{-1}(v)) - N(\Lambda^{-1}(u)) = k) \\ &= \mathrm{e}^{-(\Lambda(\Lambda^{-1}(v)) - \Lambda(\Lambda^{-1}(u)))} \frac{(\Lambda(\Lambda^{-1}(v)) - \Lambda(\Lambda^{-1}(u)))^k}{k!} \\ &= \mathrm{e}^{-(v-u)} \frac{(v-u)^k}{k!}, \end{aligned}$$

from which stationarity follows. □

This is one of the many examples in insurance and finance where a more complicated process N can be reduced to a standard (easier) model $\tilde{N}$ through the careful choice of a new time clock (a so-called *time change construction*) (see also Section 10.5.1 on credit risk). Proposition 13.31 can be formulated more generally for Λ not strictly increasing, and the converse also holds. Proposition 13.31 justifies the common simplifying assumption that a loss frequency model is homogeneous (unit rate) Poisson, albeit in many cases only in operational time. The original time-scale of N is slowed down or speeded up in such a way that, on average, $\tilde{N}$ has one claim per time unit, whereas N has, on average, $\Lambda(1)$ claims.

Remark 13.32. A standard way in which an inhomogeneous Poisson process can be obtained from a homogeneous Poisson process is by random sampling. Suppose an intensity function λ satisfies $\lambda(s) \leqslant c < \infty$ for $s \geqslant 0$. Start from a homogeneous Poisson process with rate $c > 0$ and denote its arrival times by $T_0 = 0, T_1, T_2, \ldots$. Construct a new process $\tilde{N}$ from $(T_i)_{i \geqslant 0}$ through deletion of each T_i independently of the other T_j with probability $1 - (\lambda(T_i)/c)$. The so-called *thinned* counting process $\tilde{N}$ consists of the remaining (undeleted) points. It can be shown that this process is inhomogeneous Poisson with intensity function $\lambda(\cdot)$.

Mixed Poisson processes. The mixed Poisson rvs of Section 13.2.4 can be embedded into a so-called *mixed Poisson process*. A single realization of such a process cannot be distinguished through statistical means from a realization of a homogeneous Poisson process; indeed, to simulate a sample path, one first draws a realization of the random intensity $\lambda = \Lambda(\omega)$ and then draws the sample path of the homogeneous Poisson process with rate λ. (Here, Λ denotes an rv and not the intensity measure in the inhomogeneous Poisson case above.) Only by repeating this simulation more frequently does one see the different probabilistic nature of the mixed Poisson process: compare Figure 13.9 with Figure 13.8. In the former we have simulated ten sample paths from a mixed Poisson process with mixing variable $\Lambda \sim \mathrm{Ga}(100, 1)$ so that $E(\Lambda) = 100$. Note the much greater variability in the paths.

Example 13.33. When counting processes are used in credit risk modelling the times T_k typically correspond to credit events, for instance default or downgradings. More precisely, a credit event can be constructed as the first jump of a counting process N. The df of the time to the credit event can be easily derived by observing that $P(T_1 > t) = P(N(t) = 0)$. This probability can be calculated in a straightforward way for a homogeneous Poisson process with intensity λ; we obtain $P(N(t) = 0) = \mathrm{e}^{-\lambda t}$. When N is a mixed Poisson process with mixing df F_Λ we obtain

$$P(T_1 > t) = P(N(t) = 0) = \int_0^\infty \mathrm{e}^{-t\lambda} \, \mathrm{d}F_\Lambda(\lambda) = \hat{F}_\Lambda(t),$$

the Laplace–Stieltjes transform of F_Λ in t. In the special case when $\Lambda \sim \mathrm{Ga}(\alpha, \beta)$, the negative binomial case treated in Proposition 13.21, one finds that

$$\begin{aligned} P(T_1 > t) &= \int_0^\infty \mathrm{e}^{-t\lambda} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \mathrm{e}^{-\beta\lambda} \, \mathrm{d}\lambda \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} (t+\beta)^{-\alpha} \int_0^\infty \mathrm{e}^{-s} s^{\alpha-1} \, \mathrm{d}s \\ &= \beta^\alpha (t+\beta)^{-\alpha}, \quad t \geqslant 0, \end{aligned}$$

so that T_1 has a Pareto distribution $T_1 \sim \mathrm{Pa}(\alpha, \beta)$ (see Section A.2.8).

Processes with stochastic intensity. A further important class of models is obtained when λ in the homogeneous Poisson case is replaced by a general stochastic process (λ_t), yielding a two-tier stochastic model or so-called *doubly stochastic* process.

For example, one could take λ_t to be a diffusion or, alternatively, a finite-state Markov chain. The latter case gives rise to a *regime-switching* model: in each state of

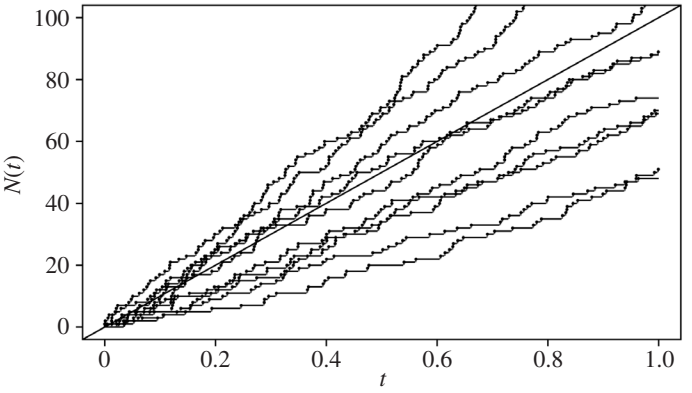

Figure 13.9. Ten realizations of a mixed Poisson process with $\Lambda \sim \text{Ga}(100, 1)$.

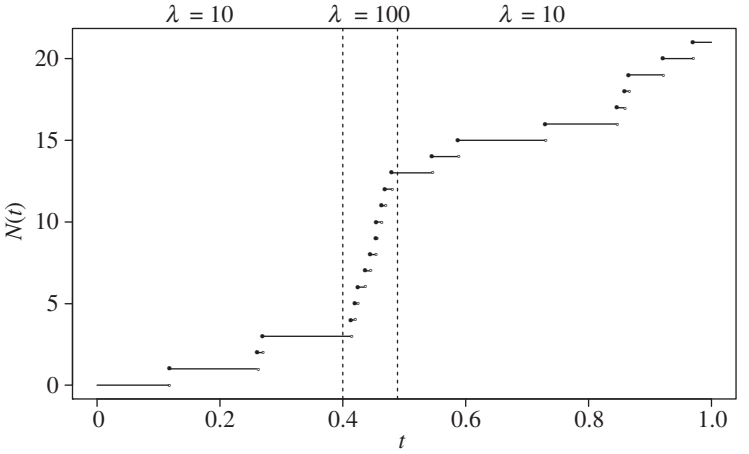

Figure 13.10. Realization of a counting process with a regime switch from $\lambda = 10$ to $\lambda = 100$.

the Markov chain the intensity has a different constant level and the process remains in that state for an exponential length of time, before jumping to another state. In Figure 13.10 we have simulated the sample path of such a process randomly switching between $\lambda = 10$ and $\lambda = 100$. In Section 10.5.1 we looked at doubly stochastic random times, which correspond to the first jump of a doubly stochastic Poisson process.

Notes and Comments

The story behind the name *insurance analytics* is told in Embrechts (2002). A good place to start a search for actuarial literature is the website of the International Actuarial Association: www.actuaries.org. Several interesting books can be found on the website of the Society of Actuaries, www.soa.org (whose offices happen to be on North *Martingale* Road, Schaumburg, Illinois). A standard Society of Actuaries textbook on actuarial mathematics is Bowers et al. (1986); financial economics for

actuaries is to be found in Panjer and Boyle (1998). For our purposes, excellent texts are Mikosch (2004) and Partrat and Besson (2004). A most readable and informative (online) set of lecture notes, with many historical links, is Wüthrich (2013). Rolski et al. (1999) gives a broad, more technical overview of the relevant stochastic process models. In Chapters 7 and 8 we have given several references to actuarial tools relevant to the study of risk measures, including distortion risk measures (Section 8.2.1), comonotonicity (Section 7.2.1) and Fréchet bounds (Sections 7.2.1 and 8.4.4). Finally, an overview of the state of the art of actuarial modelling is to be found in Teugels and Sundt (2004).

Actuarial textbooks dealing in particular with the modelling of loss distributions in insurance are Hogg and Klugman (1984), Klugman, Panjer and Willmot (2012) and Klugman, Panjer and Willmot (2013). Besides the general references above, an early textbook discussion of the use of numerical methods for the calculation of the df of total loss amount rvs is Feilmeier and Bertram (1987); Bühlmann (1984) contains a first comparison between the FFT method and Panjer recursion. More extensive comparisons, taking rounding and discretization errors into account, are found in Grübel and Hermesmeier (1999, 2000). A discussion of the use of the FFT in insurance is given in Embrechts, Gruebel and Pitts (1993). Algorithms for the FFT are freely available on the web, as a search will quickly reveal. The original paper by Panjer (1981) also contains a density version of Theorem 13.16. For an application of Panjer recursion to credit risk measurement within the CreditRisk$^+$ framework, see Credit Suisse Financial Products (1997). Based on Giese (2003), Haaf, Reiss and Schoenmakers (2004) propose an alternative recursive method. For more recent work on Panjer recursion, especially in the multivariate case, see, for example, Hesselager (1996) and Sundt (1999, 2000). For a textbook treatment of recursive methods in insurance, see Sundt and Vernic (2009).

Asymptotic approximation methods going beyond the normal approximation (13.13) are known in statistics under the names Berry–Esséen, Edgeworth and *saddle-point*. The former two are discussed, for example, in Embrechts, Klüppelberg and Mikosch (1997) and are of more theoretical importance. The saddle-point technique is very useful: see Jensen (1995) for an excellent summary, and Embrechts et al. (1985) for an application to compound distributions. Gordy (2002) discusses the importance of saddle-point methods for credit risk modelling, again within the context of CreditRisk$^+$. Wider applications within risk management can be found in Studer (2001) and Glasserman (2004); Glasserman and Li (2003)

Poisson mixture models with insurance applications in mind are summarized in Grandell (1997) (see also Bening and Korolev 2002). In order to delve more deeply into the world of counting processes, one has to study the theory of point processes. Very comprehensive and readable accounts are Daley and Vere-Jones (2003) and Karr (1991). A study of this theory is both mathematically demanding and practically rewarding. Such models are being used increasingly in credit risk. The notion of time change is fundamental to many applications in insurance and finance; for an example of how it can be used to model operational risk, see Embrechts, Kaufmann and Samorodnitsky (2004). For its introduction into finance, see Ané and Geman

(2000) and Dacorogna et al. (2001). An excellent survey is to be found in Peeters (2004).

What have we not included in our brief account of the elements of insurance analytics? We have not treated ruin theory and the general stochastic process theory of insurance risk, credibility theory, dynamic financial analysis, also referred to as dynamic solvency testing, or reinsurance, to name but a few omissions.

The stochastic process theory of insurance risk has a long tradition. The first fundamental summary came through the pioneering work of Cramér (1994a,b). Bühlmann (1970) made the field popular to several generations of actuaries. This early work has now been generalized in every way possible. A standard textbook on *ruin theory* is Asmussen and Albrecher (2010). The modelling of large claims and its consequences for ruin estimates can be found in Embrechts, Klüppelberg and Mikosch (1997).

Credibility theory concerns premium calculation for non-homogeneous portfolios and has a very rich history rooted in non-life insurance mathematics. Its basic concepts were first developed by American actuaries in the 1920s; pioneering papers in this early period were Mowbray (1914) and Whitney (1918). Further important work is found in the papers of Bailey (1945), Robbins (1955, 1964) and Bühlmann (1967, 1969, 1971). An excellent review article tracing the historical development of the basic ideas is Norberg (1979); see also Jewell (1990) for a more recent review. Various textbook versions exist: Bühlmann and Gisler (2005) give an authoritative account of its actuarial usage and hint at applications to financial risk management.

Dynamic financial analysis, also referred to as *dynamic solvency testing*, is a systematic approach, based on large-scale computer simulations, for the integrated financial modelling of non-life insurance and reinsurance companies aimed at assessing the risks and benefits associated with strategic decisions (see Blum 2005; Blum and Dacorogna 2004). An easy introduction can be found in Kaufmann, Gadmer and Klett (2001). The interested reader can consult the website of the Casualty Actuarial Society (www.casact.org/research/drm).

Part IV

Special Topics

14
Multivariate Time Series

In this chapter we consider multivariate time-series models for multiple series of financial risk-factor change data, such as differenced logarithmic price series. The presentation closely parallels the presentation of the corresponding ideas for univariate time series in Chapter 4.

An introduction to concepts in the analysis of multivariate time series and a discussion of multivariate ARMA models is found in Section 14.1, while Section 14.2 presents some of the more important examples of multivariate GARCH models.

14.1 Fundamentals of Multivariate Time Series

Among the fundamentals of multivariate time series discussed in this section are the concepts of cross-correlation, stationarity of multivariate time series, multivariate white noise processes and multivariate ARMA models.

14.1.1 Basic Definitions

A multivariate time-series model for multiple risk factors is a stochastic process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$, i.e. a family of random vectors, indexed by the integers and defined on some probability space $(\Omega, \mathcal{F}, P)$.

Moments of a multivariate time series. Assuming they exist, we define the *mean function* $\boldsymbol{\mu}(t)$ and the *covariance matrix function* $\Gamma(t,s)$ of $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ by

$$\begin{aligned} \boldsymbol{\mu}(t) &= E(\boldsymbol{X}_t), && t \in \mathbb{Z}, \\ \Gamma(t,s) &= E((\boldsymbol{X}_t - \boldsymbol{\mu}(t))(\boldsymbol{X}_s - \boldsymbol{\mu}(s))'), && t,s \in \mathbb{Z}. \end{aligned}$$

Analogously to the univariate case, we have $\Gamma(t,t) = \operatorname{cov}(\boldsymbol{X}_t)$. By observing that the elements $\gamma_{ij}(t,s)$ of $\Gamma(t,s)$ satisfy

$$\gamma_{ij}(t,s) = \operatorname{cov}(X_{t,i}, X_{s,j}) = \operatorname{cov}(X_{s,j}, X_{t,i}) = \gamma_{ji}(s,t), \tag{14.1}$$

it is clear that $\Gamma(t,s) = \Gamma(s,t)'$ for all t, s. However, the matrix Γ need not be symmetric, so in general $\Gamma(t,s) \neq \Gamma(s,t)$, which is in contrast to the univariate case. Lagged values of one of the component series can be more strongly correlated with future values of another component series than vice versa. This property, when observed in empirical data, is known as a *lead–lag* effect and is discussed in more detail in Example 14.7.

Stationarity. The multivariate models we consider will be stationary in one or both of the following senses.

Definition 14.1 (strict stationarity). The multivariate time series $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is *strictly* stationary if

$$(\boldsymbol{X}'_{t_1}, \ldots, \boldsymbol{X}'_{t_n}) \stackrel{\mathrm{d}}{=} (\boldsymbol{X}'_{t_1+k}, \ldots, \boldsymbol{X}'_{t_n+k})$$

for all $t_1, \ldots, t_n, k \in \mathbb{Z}$ and for all $n \in \mathbb{N}$.

Definition 14.2 (covariance stationarity). The multivariate time series $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is *covariance* stationary (or *weakly* or *second-order* stationary) if the first two moments exist and satisfy

$$\begin{aligned} \boldsymbol{\mu}(t) &= \boldsymbol{\mu}, & t &\in \mathbb{Z}, \\ \Gamma(t,s) &= \Gamma(t+k, s+k), & t, s, k &\in \mathbb{Z}. \end{aligned}$$

A strictly stationary multivariate time series with finite covariance matrix is covariance stationary, but we note that, as in the univariate case, it is possible to define infinite-variance processes (including certain multivariate ARCH and GARCH processes) that are strictly stationary but not covariance stationary.

Serial correlation and cross-correlation in stationary multivariate time series. The definition of covariance stationarity implies that for all s, t we have $\Gamma(t-s, 0) = \Gamma(t, s)$, so that the covariance between X_t and X_s only depends on their temporal separation $t - s$, which is known as the *lag*. In contrast to the univariate case, the sign of the lag is important.

For a covariance-stationary multivariate process we write the covariance matrix function as a function of one variable: $\Gamma(h) := \Gamma(h, 0), \forall h \in \mathbb{Z}$. Noting that $\Gamma(0) = \mathrm{cov}(\boldsymbol{X}_t)$, $\forall t$, we can now define the correlation matrix function of a covariance-stationary process. For this definition we recall the operator $\Delta(\cdot)$, defined in (6.4), which when applied to $\Sigma = (\sigma_{ij}) \in \mathbb{R}^{d\times d}$ returns a diagonal matrix with the values $\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{dd}}$ on the diagonal.

Definition 14.3 (correlation matrix function). Writing $\Delta := \Delta(\Gamma(0))$, where $\Delta(\cdot)$ is the operator defined in (6.4), the correlation matrix function $P(h)$ of a covariance-stationary multivariate time series $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is

$$P(h) := \Delta^{-1}\Gamma(h)\Delta^{-1}, \quad \forall h \in \mathbb{Z}. \tag{14.2}$$

The diagonal entries $\rho_{ii}(h)$ of this matrix-valued function give the autocorrelation function of the ith component series $(X_{t,i})_{t\in\mathbb{Z}}$. The off-diagonal entries give so-called cross-correlations between different component series at different times. It follows from (14.1) that $P(h) = P(-h)'$, but $P(h)$ need not be symmetric, and in general $P(h) \neq P(-h)$.

White noise processes. As in the univariate case, *multivariate white noise* processes are building blocks for practically useful classes of time-series model.

Definition 14.4 (multivariate white noise). $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is multivariate white noise if it is covariance stationary with correlation matrix function given by

$$P(h) = \begin{cases} P, & h = 0, \\ 0, & h \neq 0, \end{cases}$$

for some positive-definite correlation matrix P.

A multivariate white noise process with mean 0 and covariance matrix $\Sigma = \text{cov}(\boldsymbol{X}_t)$ will be denoted by WN($\mathbf{0}$, Σ). Such a process has no cross-correlation between component series, except for contemporaneous cross-correlation at lag zero. A simple example is a series of iid random vectors with finite covariance matrix, and this is known as a *multivariate strict white noise*.

Definition 14.5 (multivariate strict white noise). $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is multivariate strict white noise if it is a series of iid random vectors with finite covariance matrix.

A strict white noise process with mean 0 and covariance matrix Σ will be denoted by SWN($\mathbf{0}$, Σ).

The martingale-difference noise concept may also be extended to higher dimensions. As before we assume that the time series $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is adapted to some *filtration* $(\mathcal{F}_t)$, typically the natural filtration $(\sigma(\{\boldsymbol{X}_s : s \leqslant t\}))$, which represents the information available at time t.

Definition 14.6 (multivariate martingale difference). $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ has the multivariate martingale-difference property with respect to the filtration $(\mathcal{F}_t)$ if $E|\boldsymbol{X}_t| < \infty$ and

$$E(\boldsymbol{X}_t \mid \mathcal{F}_{t-1}) = \mathbf{0}, \quad \forall t \in \mathbb{Z}.$$

The unconditional mean of such a process is obviously also zero and, if $\text{cov}(\boldsymbol{X}_t) < \infty$ for all t, the covariance matrix function satisfies $\Gamma(t, s) = 0$ for $t \neq s$. If the covariance matrix is also constant for all t, then a process with the multivariate martingale-difference property is a multivariate white noise process.

14.1.2 Analysis in the Time Domain

We now assume that we have a random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from a covariance-stationary multivariate time-series model $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$. In the time domain we construct empirical estimators of the covariance matrix function and the correlation matrix function from this random sample.

The *sample covariance matrix function* is calculated according to

$$\hat{\Gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(\boldsymbol{X}_{t+h} - \bar{\boldsymbol{X}})(\boldsymbol{X}_t - \bar{\boldsymbol{X}})', \quad 0 \leqslant h < n,$$

where $\bar{\boldsymbol{X}} = \sum_{t=1}^{n} \boldsymbol{X}_t/n$ is the sample mean, which estimates $\boldsymbol{\mu}$, the mean of the time series. Writing $\hat{\Delta} := \Delta(\hat{\Gamma}(0))$, where $\Delta(\cdot)$ is the operator defined in (6.4), the *sample correlation matrix function* is

$$\hat{P}(h) = \hat{\Delta}^{-1}\hat{\Gamma}(h)\hat{\Delta}^{-1}, \quad 0 \leqslant h < n.$$

The information contained in the elements $\hat{\rho}_{ij}(h)$ of the sample correlation matrix function is generally displayed in the *cross-correlogram*, which is a $d \times d$ matrix of plots (see Figure 14.1 for an example). The ith diagonal plot in this graphical display is simply the correlogram of the ith component series, given by $\{(h, \hat{\rho}_{ii}(h)) : h = 0, 1, 2, \dots\}$. For the off-diagonal plots containing the estimates of *cross-correlation* there are various possible presentations and we will consider the following convention: for $i < j$ we plot $\{(h, \hat{\rho}_{ij}(h)) : h = 0, 1, 2, \dots\}$; for $i > j$ we plot $\{(-h, \hat{\rho}_{ij}(h)) : h = 0, 1, 2, \dots\}$. An interpretation of the meaning of the off-diagonal pictures is given in Example 14.7.

It can be shown that for causal processes driven by multivariate strict white noise innovations (see Section 14.1.3) the estimates that comprise the components of the sample correlation matrix function $\hat{P}(h)$ are consistent estimates of the underlying theoretical quantities. For example, if the data themselves are from an SWN process, then the cross-correlation estimators $\hat{\rho}_{ij}(h)$ for $h \neq 0$ converge to zero as the sample size is increased. However, results concerning the asymptotic distribution of cross-correlation estimates are, in general, more complicated than the univariate result for autocorrelation estimates given in Theorem 4.13. Some relevant theory is found in Chapter 11 of Brockwell and Davis (1991) and Chapter 7 of Brockwell and Davis (2002). It is standard to plot the off-diagonal pictures with Gaussian confidence bands at $(-1.96/\sqrt{n}, 1.96/\sqrt{n})$, but these bands should be used as rough guidance for the eye rather than being relied upon too heavily to draw conclusions.

Example 14.7 (cross-correlogram of trivariate index returns). In Figure 14.1 the cross-correlogram of daily log-returns is shown for the Dow Jones, Nikkei and Swiss Market indices for 26 July 1996–25 July 2001. Although every vector observation in this trivariate time series relates to the same trading day, the returns are of course not properly synchronized due to time zones. This picture therefore shows interpretable lead–lag effects that help us to understand the off-diagonal pictures in the cross-correlogram.

Part (b) of the figure shows estimated correlations between the Dow Jones index return on day $t + h$ and the Nikkei index return on day t, for $h \geqslant 0$; these estimates are clearly small and lie mainly within the confidence band, with the obvious exception of the correlation estimate for returns on the same trading day $\hat{P}_{12}(0) \approx 0.14$. Part (d) shows estimated correlations between the Dow Jones index return on day $t + h$ and the Nikkei index return on day t, for $h \leqslant 0$; the estimate corresponding to $h = -1$ is approximately 0.28 and can be interpreted as showing how the American market leads the Japanese market. Comparing parts (c) and (g) we see, unsurprisingly, that the American market also leads the Swiss market, so that returns on day $t - 1$ in the former are quite strongly correlated with returns on day t in the latter.

14.1.3 Multivariate ARMA Processes

We provide a brief excursion into multivariate ARMA models to indicate how the ideas of Section 4.1.2 generalize to higher dimensions. For daily data, capturing multivariate ARMA effects is much less important than capturing multivariate volatility effects (and dynamic correlation effects) through multivariate GARCH

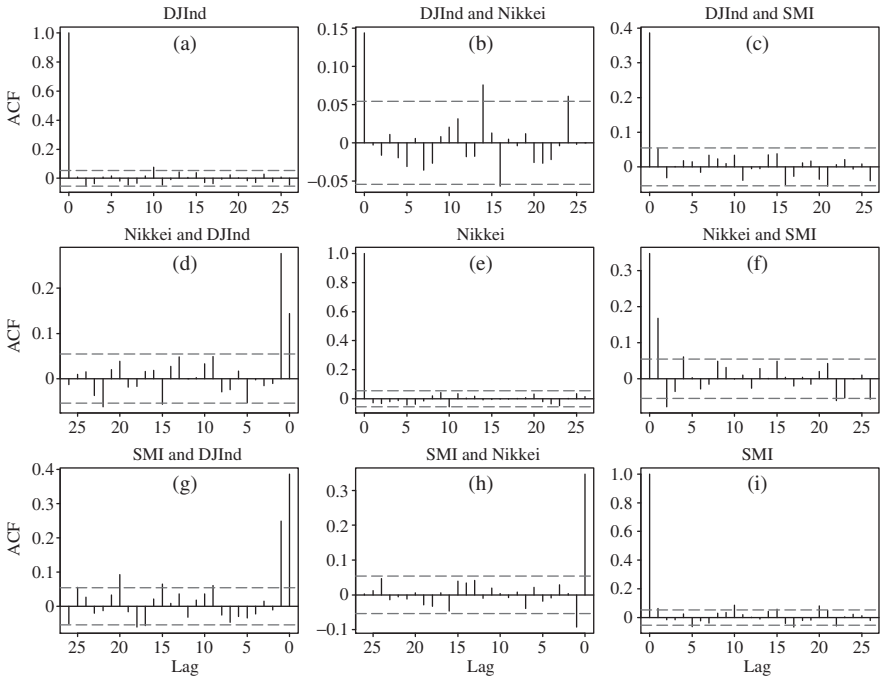

Figure 14.1. Cross-correlogram of daily log-returns of the Dow Jones, Nikkei and Swiss Market indices for 26 July 1996–25 July 2001 (see Example 14.7 for commentary).

modelling, but, for longer-period returns, the more traditional ARMA processes become increasingly useful. In the econometrics literature they are more commonly known as vector ARMA (or VARMA) processes.

Definition 14.8 (VARMA process). Let $(\boldsymbol{\varepsilon}_t)_{t\in\mathbb{Z}}$ be WN$(\mathbf{0}, \Sigma_\varepsilon)$. The process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a zero-mean VARMA(p, q) process if it is a covariance-stationary process satisfying difference equations of the form

$$\boldsymbol{X}_t - \Phi_1 \boldsymbol{X}_{t-1} - \cdots - \Phi_p \boldsymbol{X}_{t-p} = \boldsymbol{\varepsilon}_t + \Theta_1 \boldsymbol{\varepsilon}_{t-1} + \cdots + \Theta_q \boldsymbol{\varepsilon}_{t-q}, \quad \forall t \in \mathbb{Z},$$

for parameter matrices Φ_i and Θ_j in $\mathbb{R}^{d\times d}$. $(\boldsymbol{X}_t)$ is a VARMA process with mean $\boldsymbol{\mu}$ if the centred series $(\boldsymbol{X}_t - \boldsymbol{\mu})_{t\in\mathbb{Z}}$ is a zero-mean VARMA(p, q) process.

Consider a zero-mean VARMA(p, q) process. For practical applications we again consider only causal processes (see Section 4.1.2), which are processes where the solution of the defining equations has a representation of the form

$$\boldsymbol{X}_t = \sum_{i=0}^{\infty} \Psi_i \boldsymbol{\varepsilon}_{t-i}, \tag{14.3}$$

where $(\Psi_i)_{i\in\mathbb{N}_0}$ is a sequence of matrices in $\mathbb{R}^{d\times d}$ with absolutely summable components, meaning that, for any j and k,

$$\sum_{i=0}^{\infty} |\psi_{i,jk}| < \infty. \tag{14.4}$$

As in the univariate case (see Proposition 4.9) it can be verified by direct calculation that such linear processes are covariance stationary. For $h \geqslant 0$ the covariance matrix function is given by

$$\Gamma(t+h,t) = \operatorname{cov}(\boldsymbol{X}_t + h, \boldsymbol{X}_t) = E\bigg(\sum_{i=0}^{\infty} \Psi_i \boldsymbol{\varepsilon}_{t+h-i} \sum_{j=0}^{\infty} \boldsymbol{\varepsilon}'_{t-j} \Psi'_j\bigg).$$

Arguing much as in the univariate case it is easily shown that this depends only on h and not on t and that it is given by

$$\Gamma(h) = \sum_{i=0}^{\infty} \Psi_{i+h} \Sigma_\varepsilon \Psi'_i, \quad h = 0, 1, 2, \ldots. \tag{14.5}$$

The correlation matrix function is easily derived from (14.5) and (14.2).

The requirement that a VARMA process be causal imposes conditions on the values that the parameter matrices Φ_i (in particular) and Θ_j may take. The theory is quite similar to univariate ARMA theory. We will give a single useful example from the VARMA class; this is the first-order vector autoregressive (or VAR(1)) model.

Example 14.9 (VAR(1) process). The first-order VAR process satisfies the set of vector difference equations

$$\boldsymbol{X}_t = \Phi \boldsymbol{X}_{t-1} + \boldsymbol{\varepsilon}_t, \quad \forall t. \tag{14.6}$$

It is possible to find a causal process satisfying (14.3) and (14.4) that is a solution of (14.6) if and only if all eigenvalues of the matrix Φ are less than 1 in absolute value. The causal process

$$\boldsymbol{X}_t = \sum_{i=0}^{\infty} \Phi^i \boldsymbol{\varepsilon}_{t-i} \tag{14.7}$$

is then the unique solution. This solution can be thought of as an infinite-order vector moving-average process, a so-called VMA(∞) process. The covariance matrix function of this process follows from (14.3) and (14.5) and is given by

$$\Gamma(h) = \sum_{i=0}^{\infty} \Phi^{i+h} \Sigma_\varepsilon \Phi^{i\prime} = \Phi^h \Gamma(0), \quad h = 0, 1, 2, \ldots.$$

In practice, full VARMA models are less common than models from the VAR and VMA subfamilies, one reason being that identifiability problems arise when estimating parameters. For example, we can have situations where the first-order VARMA(1, 1) model $\boldsymbol{X}_t - \Phi \boldsymbol{X}_{t-1} = \boldsymbol{\varepsilon}_t + \Theta \boldsymbol{\varepsilon}_{t-1}$ can be rewritten as $\boldsymbol{X}_t - \Phi^* \boldsymbol{X}_{t-1} = \boldsymbol{\varepsilon}_t + \Theta^* \boldsymbol{\varepsilon}_{t-1}$ for completely different parameter matrices Φ^* and Θ^* (see Tsay (2002, p. 323) for an example). Of the two subfamilies, VAR models are easier to estimate. Fitting options for VAR models range from multivariate least-squares estimation without strong assumptions concerning the distribution of the driving white noise, to full ML estimation; models combining VAR and multivariate GARCH features can be estimated using a conditional ML approach in a very similar manner to that described for univariate models in Section 4.2.4.

Notes and Comments

Many standard texts on time series also handle the multivariate theory: see, for example, Brockwell and Davis (1991, 2002) or Hamilton (1994). A key reference aimed at an econometrics audience is Lütkepohl (1993). For examples in the area of finance see Tsay (2002) and Zivot and Wang (2003).

14.2 Multivariate GARCH Processes

In this section we first establish general notation for multivariate GARCH (or MGARCH) models before going on to consider models that are defined via their conditional correlation matrix in Section 14.2.2 and models that are defined via their conditional covariance matrix in Section 14.2.4. We also provide brief notes on model estimation, strategies for dimension reduction, and the use of multivariate GARCH models in quantitative risk management.

14.2.1 General Structure of Models

For the following definition we recall that the Cholesky factor $\Sigma^{1/2}$ of a positive-definite matrix Σ is the lower-triangular matrix satisfying $AA' = \Sigma$ (see the discussion at end of Section 6.1).

Definition 14.10. Let $(\boldsymbol{Z}_t)_{t\in\mathbb{Z}}$ be SWN$(\mathbf{0}, I_d)$. The process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is said to be a multivariate GARCH process if it is strictly stationary and satisfies equations of the form

$$\boldsymbol{X}_t = \Sigma_t^{1/2}\boldsymbol{Z}_t, \quad t \in \mathbb{Z}, \tag{14.8}$$

where $\Sigma_t^{1/2} \in \mathbb{R}^{d\times d}$ is the Cholesky factor of a positive-definite matrix Σ_t that is measurable with respect to $\mathcal{F}_{t-1} = \sigma(\{\boldsymbol{X}_s : s \leqslant t-1\})$, the history of the process up to time $t-1$.

Conditional moments. It is easily calculated that a covariance-stationary process of this type has the multivariate martingale-difference property

$$E(\boldsymbol{X}_t \mid \mathcal{F}_{t-1}) = E(\Sigma_t^{1/2}\boldsymbol{Z}_t \mid \mathcal{F}_{t-1}) = \Sigma_t^{1/2}E(\boldsymbol{Z}_t) = \mathbf{0},$$

and it must therefore be a multivariate white noise process, as argued in Section 14.1. Moreover, Σ_t is the *conditional covariance matrix* since

$$\begin{aligned}\operatorname{cov}(\boldsymbol{X}_t \mid \mathcal{F}_{t-1}) &= E(\boldsymbol{X}_t\boldsymbol{X}_t' \mid \mathcal{F}_{t-1})\\ &= \Sigma_t^{1/2}E(\boldsymbol{Z}_t\boldsymbol{Z}_t')(\Sigma_t^{1/2})'\\ &= \Sigma_t^{1/2}(\Sigma_t^{1/2})'\\ &= \Sigma_t. \end{aligned}\tag{14.9}$$

The conditional covariance matrix Σ_t in a multivariate GARCH model corresponds to the squared volatility σ_t^2 in a univariate GARCH model. The use of the Cholesky factor of Σ_t to describe the relationship to the driving noise in (14.8) is not important, and in fact any type of "square root" of Σ_t could be used (such as the root derived from a symmetric decomposition). (The only practical consequence is that

different choices would lead to different series of residuals when the model is fitted in practice.) We denote the elements of Σ_t by $\sigma_{t,ij}$ and also use the notation $\sigma_{t,i} = \sqrt{\sigma_{t,ii}}$ to denote the conditional standard deviation (or volatility) of the ith component series $(X_{t,i})_{t\in\mathbb{Z}}$.

We recall that we can write $\Sigma_t = \Delta_t P_t \Delta_t$, where

$$\Delta_t = \Delta(\Sigma_t) = \operatorname{diag}(\sigma_{t,1}, \ldots, \sigma_{t,d}), \quad P_t = \wp(\Sigma_t), \tag{14.10}$$

using the operator notation defined in (6.4) and (6.5). The diagonal matrix Δ_t is known as the *volatility matrix* and P_t is known as the *conditional correlation matrix*. The art of building multivariate GARCH models is to specify the dependence of Σ_t (or of Δ_t and P_t) on the past in such a way that Σ_t always remains symmetric and positive definite. A covariance matrix must of course be symmetric and positive semidefinite, and in practice we restrict our attention to the positive-definite case. This facilitates fitting, since the conditional distribution of $\boldsymbol{X}_t \mid \mathcal{F}_{t-1}$ never has a singular covariance matrix.

Unconditional moments. The *unconditional covariance matrix* Σ of a process of this type is given by

$$\Sigma = \operatorname{cov}(\boldsymbol{X}_t) = E(\operatorname{cov}(\boldsymbol{X}_t \mid \mathcal{F}_{t-1})) + \operatorname{cov}(E(\boldsymbol{X}_t \mid \mathcal{F}_{t-1})) = E(\Sigma_t),$$

from which it can be calculated that the *unconditional correlation matrix* P has elements

$$\rho_{ij} = \frac{E(\sigma_{t,ij})}{\sqrt{E(\sigma_{t,ii})E(\sigma_{t,jj})}} = \frac{E(\rho_{t,ij}\sigma_{t,i}\sigma_{t,j})}{\sqrt{E(\sigma_{t,i}^2)E(\sigma_{t,j}^2)}}, \tag{14.11}$$

which is in general difficult to evaluate and is usually not simply the expectation of the conditional correlation matrix.

Innovations. In practical work the innovations are generally taken to be from either a multivariate Gaussian distribution ($\boldsymbol{Z}_t \sim N_d(\boldsymbol{0}, I_d)$) or, more realistically for daily returns, an appropriately scaled spherical multivariate t distribution ($\boldsymbol{Z}_t \sim t_d(\nu, \boldsymbol{0}, (\nu-2)I_d/\nu)$). Any distribution with mean 0 and covariance matrix I_d is permissible, and appropriate members of the normal mixture family of Section 6.2 or the spherical family of Section 6.3.1 may be considered.

Presentation of models. In the following sections we present some of the more important multivariate GARCH specifications. In doing this we concentrate on the following aspects of the models.

- The form of the dynamic equations, with economic arguments and criticisms where appropriate.
- The conditions required to guarantee that the conditional covariance matrix Σ_t remains positive definite. Other mathematical properties of these models, such as conditions for covariance stationarity, are difficult to derive with full mathematical rigour; references in Notes and Comments contain further information.

- The parsimoniousness of the parametrization. A major problem with most multivariate GARCH specifications is that the number of parameters tends to explode with the dimension of the model, making them unsuitable for analyses of many risk factors.
- Simple intuitive fitting methods where available. All models can be fitted by a general global-maximization approach described in Section 14.2.4 but certain models lend themselves to estimation in stages, particularly the models of Section 14.2.2.

14.2.2 Models for Conditional Correlation

In this section we present models that focus on specifying the conditional correlation matrix P_t while allowing volatilities to be described by univariate GARCH models; we begin with a popular and relatively parsimonious model where P_t is assumed to be constant for all t.

Constant conditional correlation (CCC).

Definition 14.11. The process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a CCC–GARCH process if it is a process with the general structure given in Definition 14.10 such that the conditional covariance matrix is of the form $\Sigma_t = \Delta_t P_c \Delta_t$, where

(i) P_c is a constant, positive-definite correlation matrix; and

(ii) Δ_t is a diagonal volatility matrix with elements $\sigma_{t,k}$ satisfying

$$\sigma_{t,k}^2 = \alpha_{k0} + \sum_{i=1}^{p_k} \alpha_{ki} X_{t-i,k}^2 + \sum_{j=1}^{q_k} \beta_{kj} \sigma_{t-j,k}^2, \quad k = 1, \ldots, d, \tag{14.12}$$

where $\alpha_{k0} > 0$, $\alpha_{ki} \geqslant 0$, $i = 1, \ldots, p_k$, $\beta_{kj} \geqslant 0$, $j = 1, \ldots, q_k$.

The CCC–GARCH specification represents a simple way of combining univariate GARCH processes. This can be seen by observing that in a CCC–GARCH model observations and innovations are connected by equations $\boldsymbol{X}_t = \Delta_t P_c^{1/2} \boldsymbol{Z}_t$, which may be rewritten as $\boldsymbol{X}_t = \Delta_t \tilde{\boldsymbol{Z}}_t$ for an SWN($\boldsymbol{0}$, P_c) process $(\tilde{\boldsymbol{Z}}_t)_{t\in\mathbb{Z}}$. Clearly, the component processes are univariate GARCH.

Proposition 14.12. *The CCC–GARCH model is well defined in the sense that Σ_t is almost surely positive definite for all t. Moreover, it is covariance stationary if and only if $\sum_{i=1}^{p_k} \alpha_{ki} + \sum_{j=1}^{q_k} \beta_{kj} < 1$ for $k = 1, \ldots, d$.*

Proof. For a vector $\boldsymbol{v} \neq \boldsymbol{0}$ in $\mathbb{R}^d$ we have

$$\boldsymbol{v}' \Sigma_t \boldsymbol{v} = (\Delta_t \boldsymbol{v})' P_c (\Delta_t \boldsymbol{v}) > 0,$$

since P_c is positive definite and the strict positivity of the individual volatility processes ensures that $\Delta_t \boldsymbol{v} \neq \boldsymbol{0}$ for all t.

If $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is covariance stationary, then each component series $(X_{t,k})_{t\in\mathbb{Z}}$ is a covariance-stationary GARCH process for which a necessary and sufficient condition is $\sum_{i=1}^{p_k} \alpha_{ki} + \sum_{j=1}^{q_k} \beta_{kj} < 1$ by Proposition 4.21. Conversely, if the component

series are covariance stationary, then for all i and j the Cauchy–Schwarz inequality implies

$$\sigma_{ij} = E(\sigma_{t,ij}) = \rho_{ij} E(\sigma_{t,i}\sigma_{t,j}) \leqslant \rho_{ij}\sqrt{E(\sigma_{t,i}^2)}\sqrt{E(\sigma_{t,j}^2)} < \infty.$$

Since $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a multivariate martingale difference with finite, non-time-dependent second moments σ_{ij}, it is a covariance-stationary white noise. □

The CCC model is often a useful starting point from which to proceed to more complex models. In some empirical settings it gives an adequate performance, but it is generally accepted that the constancy of conditional correlation in this model is an unrealistic feature and that the impact of news on financial markets requires models that allow a dynamic evolution of conditional correlation as well as a dynamic evolution of volatilities. A further criticism of the model (which in fact applies to the majority of MGARCH specifications) is the fact that the individual volatility dynamics (14.12) do not allow for the possibility that large returns in one component series at a particular point in time can contribute to the increased volatility of another component time series at future points in time.

To describe a simple method of fitting the CCC model, we introduce the notion of a *devolatized* process. For any multivariate time-series process $\boldsymbol{X}_t$, the devolatized process is the process $\boldsymbol{Y}_t = \Delta_t^{-1}\boldsymbol{X}_t$, where Δ_t is, as usual, the diagonal matrix of volatilities. In the case of a CCC model it is easily seen that the devolatized process $(\boldsymbol{Y}_t)_{t\in\mathbb{Z}}$ is an SWN$(\mathbf{0}, P_c)$ process.

This structure suggests a simple two-stage fitting method in which we first estimate the individual volatility processes for the component series by fitting univariate GARCH processes to data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$; note that, although we have specified in Definition 14.11 that the individual volatilities should follow standard GARCH models, we could of course extend the model to allow any of the univariate models in Section 4.2.3 to be used, such as GARCH with leverage or threshold GARCH.

In a second stage we construct an estimate of the devolatized process by taking $\hat{\boldsymbol{Y}}_t = \hat{\Delta}_t^{-1}\boldsymbol{X}_t$ for $t = 1, \ldots, n$, where $\hat{\Delta}_t^{-1}$ is the estimate of Δ_t; in other words, we collect the standardized residuals from the univariate GARCH models. Note that we have also described this construction in slightly more general terms in the context of a multivariate approach to dynamic historical simulation in Section 9.2.4. If the CCC–GARCH model is adequate, then the $\hat{\boldsymbol{Y}}_t$ data should behave like a realization from an SWN$(\mathbf{0}, P_c)$ process, and this can be investigated with the correlogram and cross-correlogram applied to raw and absolute values. Assuming that the model is adequate, the conditional correlation matrix P_c can then be estimated from the standardized residuals using methods from Chapter 6.

A special case of CCC–GARCH that we call a pure diagonal model occurs when $P_c = I_d$. A covariance-stationary model of this kind is a multivariate white noise where the contemporaneous components $X_{t,i}$ and $X_{t,j}$ are also uncorrelated for $i \neq j$. This subsumes the case of independent GARCH models for each component series. Indeed, if we assume that the driving SWN$(\mathbf{0}, I_d)$ process is multivariate Gaussian, then the component series are independent. However, if, for example, we

assume that the innovations satisfy $\boldsymbol{Z}_t \sim t_d(\nu, \mathbf{0}, ((\nu-2)/\nu)I_d)$, then the component processes are dependent.

Dynamic conditional correlation (DCC). This model generalizes the CCC model to allow conditional correlations to evolve dynamically according to a relatively parsimonious scheme, but it is constructed in a way that still allows estimation in stages using univariate GARCH models. Its formal analysis as a stochastic process is difficult due to the use of the correlation matrix extraction operator $\wp$ in its definition.

Definition 14.13. The process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a DCC–GARCH process if it is a process with the general structure given in Definition 14.10, where the volatilities comprising Δ_t follow univariate GARCH specifications as in (14.12) and the conditional correlation matrices P_t satisfy, for $t \in \mathbb{Z}$, the equations

$$P_t = \wp\left(\left(1 - \sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j\right)P_{\mathrm{c}} + \sum_{i=1}^{p}\alpha_i \boldsymbol{Y}_{t-i}\boldsymbol{Y}_{t-i}' + \sum_{j=1}^{q}\beta_j P_{t-j}\right), \quad (14.13)$$

where P_{c} is a positive-definite correlation matrix, $\wp$ is the operator in (6.5), $\boldsymbol{Y}_t = \Delta_t^{-1}\boldsymbol{X}_t$ denotes the devolatized process, and the coefficients satisfy $\alpha_i \geqslant 0$, $\beta_j \geqslant 0$ and $\sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j < 1$.

Observe first that if all the α_i and β_j coefficients in (14.13) are zero, then the model reduces to the CCC model. If one makes an analogy with a covariance-stationary univariate GARCH model with unconditional variance σ^2, for which the volatility equation can be written

$$\sigma_t^2 = \left(1 - \sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j\right)\sigma^2 + \sum_{i=1}^{p}\alpha_i X_{t-i} + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2,$$

then the correlation matrix P_{c} in (14.13) can be thought of as representing the long-run correlation structure. Although this matrix could be estimated by fitting the DCC model to data by ML estimation in one step, it is quite common to estimate it using an empirical correlation matrix calculated from the devolatized data, as in the estimation of the CCC model.

Observe also that the dynamic equation (14.13) preserves the positive definiteness of P_t. If we define

$$Q_t := \left(1 - \sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j\right)P_{\mathrm{c}} + \sum_{i=1}^{p}\alpha_i \boldsymbol{Y}_{t-i}\boldsymbol{Y}_{t-i}' + \sum_{j=1}^{q}\beta_j P_{t-j}$$

and assume that $P_{t-q}, \ldots, P_{t-1}$ are positive definite, then it follows that, for a vector $\boldsymbol{v} \neq \mathbf{0}$ in $\mathbb{R}^d$, we have

$$\boldsymbol{v}'Q_t\boldsymbol{v} = \left(1 - \sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j\right)\boldsymbol{v}'P_{\mathrm{c}}\boldsymbol{v} + \sum_{i=1}^{p}\alpha_i \boldsymbol{v}'\boldsymbol{Y}_{t-i}\boldsymbol{Y}_{t-i}'\boldsymbol{v} + \sum_{j=1}^{q}\beta_j \boldsymbol{v}'P_{t-j}\boldsymbol{v} > 0,$$

since the first term is strictly positive and the second and third terms are non-negative. If Q_t is positive definite, then so is P_t.

The usual estimation method for the DCC model is as follows.

(1) Fit univariate GARCH-type models to the component series to estimate the volatility matrix Δ_t. Form an estimated realization of the devolatized process by taking $\hat{\boldsymbol{Y}}_t = \hat{\Delta}_t \boldsymbol{X}_t$.

(2) Estimate P_c by taking the sample correlation matrix of the devolatized data (or, better still, some robust estimator of correlation).

(3) Estimate the remaining parameters α_i and β_j in equation (14.13) by fitting a model with structure $\boldsymbol{Y}_t = P_t^{1/2} \boldsymbol{Z}_t$ to the devolatized data. We leave this step vague for the time being and note that this will be a simple application of the methodology for fitting general multivariate GARCH models in Section 14.2.4; in a first-order model ($p = q = 1$), there will be only two remaining parameters to estimate.

14.2.3 Models for Conditional Covariance

The models of this section specify explicitly a dynamic structure for the conditional covariance matrix Σ_t. These models are not designed for multiple-stage estimation based on univariate GARCH estimation procedures.

Vector GARCH models (VEC and DVEC). The most general vector GARCH model—the VEC model—has too many parameters for practical purposes and our task will be to simplify the model by imposing various restrictions on parameter matrices.

Definition 14.14 (VEC model). The process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a VEC process if it has the general structure given in Definition 14.10, and the dynamics of the conditional covariance matrix Σ_t are given by the equations

$$\operatorname{vech}(\Sigma_t) = \boldsymbol{a}_0 + \sum_{i=1}^{p} \bar{A}_i \operatorname{vech}(\boldsymbol{X}_{t-i}\boldsymbol{X}'_{t-i}) + \sum_{j=1}^{q} \bar{B}_j \operatorname{vech}(\Sigma_{t-j}) \tag{14.14}$$

for a vector $\boldsymbol{a}_0 \in \mathbb{R}^{d(d+1)/2}$ and matrices $\bar{A}_i$ and $\bar{B}_j$ in $R^{(d(d+1)/2)\times(d(d+1)/2)}$.

In this definition "vech" denotes the *vector half* operator, which stacks the columns of the lower triangle of a symmetric matrix in a single column vector of length $d(d+1)/2$. Thus (14.14) should be understood as specifying the dynamics for the lower-triangular portion of the conditional covariance matrix, and the remaining elements of the matrix are determined by symmetry.

In this very general form the model has $(1 + (p+q)d(d+1)/2)d(d+1)/2$ parameters; this number grows rapidly with dimension so that even a trivariate model has 78 parameters. The most common simplification has been to restrict attention to cases when $\bar{A}_i$ and $\bar{B}_j$ are diagonal matrices, which gives us the diagonal VEC or DVEC model. This special case can be written very elegantly in terms of a different kind of matrix product, namely the *Hadamard product*, denoted by "$\circ$", which signifies element-by-element multiplication of two matrices of the same size.

We obtain the representation

$$\Sigma_t = A_0 + \sum_{i=1}^{p} A_i \circ (\boldsymbol{X}_{t-i}\boldsymbol{X}_{t-i}') + \sum_{j=1}^{q} B_j \circ \Sigma_{t-j}, \tag{14.15}$$

where A_0 and the A_i and B_j must all be symmetric matrices in $\mathbb{R}^{d\times d}$ such that A_0 has positive diagonal elements and all other matrices have non-negative diagonal elements (standard univariate GARCH assumptions). This representation emphasizes structural similarities with the univariate GARCH model of Definition 4.20.

To better understand the dynamic implications of (14.15), consider a bivariate model of order (1, 1) and write $a_{0,ij}$, $a_{1,ij}$ and b_{ij} for the elements of A_0, A_1 and B_1, respectively. The model amounts to the following three simple equations:

$$\left.\begin{aligned} \sigma_{t,1}^2 &= a_{0,11} + a_{1,11}X_{t-1,1}^2 + b_{11}\sigma_{t-1,1}^2, \\ \sigma_{t,12} &= a_{0,12} + a_{1,12}X_{t-1,1}X_{t-1,2} + b_{12}\sigma_{t-1,12}, \\ \sigma_{t,2}^2 &= a_{0,22} + a_{1,22}X_{t-1,2}^2 + b_{22}\sigma_{t-1,2}^2. \end{aligned}\right\} \tag{14.16}$$

The volatilities of the two component series ($\sigma_{t,1}$ and $\sigma_{t,2}$) follow univariate GARCH updating patterns, and the conditional covariance $\sigma_{t,12}$ has a similar structure driven by the products of the lagged values $X_{t-1,1}X_{t-1,2}$. As for the CCC and DCC models, the volatility of a single component series is only driven by large lagged values of that series and cannot be directly affected by large lagged values in another series; the more general but overparametrized VEC model would allow this feature.

The requirement that Σ_t in (14.15) should be a proper positive-definite covariance matrix does impose conditions on the A_0, A_i and B_j matrices that we have not yet discussed. In practice, in some software implementations of this model, formal conditions are not imposed, other than that the matrices should be symmetric with non-negative diagonal elements; the positive definiteness of the resulting estimates of the conditional covariance matrices can be checked after model fitting.

However, a sufficient condition for Σ_t to be almost surely positive definite is that A_0 should be positive definite and the matrices $A_1, \ldots, A_p, B_1, \ldots, B_q$ should all be positive semidefinite (see Notes and Comments), and this condition is easy to impose. We can constrain all parameter matrices to have a form based on a Cholesky decomposition; that is, we can parametrize the model in terms of lower-triangular Cholesky factor matrices $A_0^{1/2}$, $A_i^{1/2}$ and $B_j^{1/2}$ such that

$$A_0 = A_0^{1/2}(A_0^{1/2})', \qquad A_i = A_i^{1/2}(A_i^{1/2})', \qquad B_j = B_j^{1/2}(B_j^{1/2})'. \tag{14.17}$$

Because the sufficient condition only prescribes that $A_1, \ldots, A_p$ and $B_1, \ldots, B_q$ should be positive semidefinite, we can in fact also consider much simpler parametrizations, such as

$$A_0 = A_0^{1/2}(A_0^{1/2})', \qquad A_i = \boldsymbol{a}_i\boldsymbol{a}_i', \qquad B_j = \boldsymbol{b}_j\boldsymbol{b}_j', \tag{14.18}$$

where $\boldsymbol{a}_i$ and $\boldsymbol{b}_j$ are vectors in $\mathbb{R}^d$. An even cruder model, satisfying the requirement of positive definiteness of Σ_t, would be

$$A_0 = A_0^{1/2}(A_0^{1/2})', \qquad A_i = a_i I_d, \qquad B_j = b_j I_d, \tag{14.19}$$

where a_i and b_j are simply positive constants. In fact, the specifications of the multivariate ARCH and GARCH effects in (14.17)–(14.19) can be mixed and matched in obvious ways.

The BEKK model of Baba, Engle, Kroner and Kraft. The next family of models have the great advantage that their construction ensures the positive definiteness of Σ_t without the need for further conditions.

Definition 14.15. The process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a BEKK process if it has the general structure given in Definition 14.10 and if the conditional covariance matrix Σ_t satisfies, for all $t \in \mathbb{Z}$,

$$\Sigma_t = A_0 + \sum_{i=1}^{p} A_i' \boldsymbol{X}_{t-i} \boldsymbol{X}_{t-i}' A_i + \sum_{j=1}^{q} B_j' \Sigma_{t-j} B_j, \tag{14.20}$$

where all coefficient matrices are in $\mathbb{R}^{d\times d}$ and A_0 is symmetric and positive definite.

Proposition 14.16. *In the BEKK model (14.20), the conditional covariance matrix Σ_t is almost surely positive definite for all t.*

Proof. Consider a first-order model for simplicity. For a vector $\boldsymbol{v} \neq \boldsymbol{0}$ in $\mathbb{R}^d$ we have

$$\boldsymbol{v}'\Sigma_t\boldsymbol{v} = \boldsymbol{v}'A_0\boldsymbol{v} + (\boldsymbol{v}'A_1'\boldsymbol{X}_{t-1})^2 + (B_1\boldsymbol{v})'\Sigma_{t-1}(B_1\boldsymbol{v}) > 0,$$

since the first term is strictly positive and the second and third terms are non-negative. □

To gain an understanding of the BEKK model it is again useful to consider the bivariate special case of order (1, 1) and to consider the dynamics that are implied while comparing these with equations (14.16):

$$\begin{aligned}\sigma_{t,1}^2 &= a_{0,11} + a_{1,11}^2 X_{t-1,1}^2 + 2a_{1,11}a_{1,12}X_{t-1,1}X_{t-1,2} + a_{1,12}^2 X_{t-1,2}^2 \\ &\quad + b_{11}^2\sigma_{t-1,1}^2 + 2b_{11}b_{12}\sigma_{t-1,12} + b_{12}^2\sigma_{t-1,2}^2,\end{aligned} \tag{14.21}$$

$$\begin{aligned}\sigma_{t,12} &= a_{0,12} + (a_{1,11}a_{1,22} + a_{1,12}a_{1,21})X_{t-1,1}X_{t-1,2} \\ &\quad + a_{1,11}a_{1,21}X_{t-1,1}^2 + a_{1,22}a_{1,12}X_{t-1,2}^2 \\ &\quad + (b_{11}b_{22} + b_{12}b_{21})\sigma_{t-1,12} + b_{11}b_{21}\sigma_{t-1,1}^2 + b_{22}b_{12}\sigma_{t-1,2}^2,\end{aligned} \tag{14.22}$$

$$\begin{aligned}\sigma_{t,2}^2 &= a_{0,22} + a_{1,22}^2 X_{t-1,2}^2 + 2a_{1,22}a_{1,21}X_{t-1,1}X_{t-1,2} + a_{1,21}^2 X_{t-1,1}^2 \\ &\quad + b_{22}^2\sigma_{t-1,2}^2 + 2b_{22}b_{21}\sigma_{t-1,21} + b_{21}^2\sigma_{t-1,1}^2.\end{aligned} \tag{14.23}$$

From (14.21) it follows that we now have a model where a large lagged value of the second component $X_{t-1,2}$ can influence the volatility of the first series $\sigma_{t,1}$. The BEKK model has more parameters than the DVEC model and appears to have much richer dynamics. Note, however, that the DVEC model cannot be obtained as a special case of the BEKK model as we have defined it. To eliminate all crossover effects in the conditional variance equations of the BEKK model in (14.21) and (14.23), we would have to set the diagonal terms $a_{1,12}$, $a_{1,21}$, b_{12} and b_{21} to be zero and the parameters governing the individual volatilities would also govern the conditional covariance $\sigma_{t,12}$ in (14.22).

Table 14.1. Summary of numbers of parameters in various multivariate GARCH models: in CCC it is assumed that the numbers of ARCH and GARCH terms for all volatility equations are, respectively, p and q; in DCC it is assumed that the conditional correlation equation has $p + q$ parameters. The second column gives the general formula; the final columns give the numbers for models of dimensions 2, 5 and 10 when $p = q = 1$. Additional parameters in the innovation distribution are not considered.

Model	Parameter count	2	5	10
VEC	$d(d+1)(1+(p+q)d(d+1)/2)/2$	21	465	6105
BEKK	$d(d+1)/2 + d^2(p+q)$	11	65	255
DVEC as in (14.17)	$d(d+1)(1+p+q)/2$	9	45	165
DCC	$d(d+1)/2 + (d+1)(p+q)$	9	27	77
CCC	$d(d+1)/2 + d(p+q)$	7	25	75
DVEC as in (14.18)	$d(d+1)/2 + d(p+q)$	7	25	75
DVEC as in (14.19)	$d(d+1)/2 + (p+q)$	5	17	57

Remark 14.17. A broader definition of the BEKK class, which does subsume all DVEC models, was originally given by Engle and Kroner (1995). In this definition we have

$$\Sigma_t = A_0 A_0' + \sum_{k=1}^{K} \sum_{i=1}^{p} A_{k,i}' \boldsymbol{X}_{t-i} \boldsymbol{X}_{t-i}' A_{k,i} + \sum_{k=1}^{K} \sum_{j=1}^{q} B_{k,j}' \Sigma_{t-j} B_{k,j},$$

where $\frac{1}{2}d(d+1) > K \geqslant 1$ and the choice of K determines the richness of the model. This model class is of largely theoretical interest and tends to be too complex for practical applications; even the case $K = 1$ is difficult to fit in higher dimensions.

In Table 14.1 we have summarized the numbers of parameters in these models. Broad conclusions concerning the practical implications are as follows: the general VEC model is of purely theoretical interest; the BEKK and general DVEC models are for very low-dimensional use; and the remaining models are the most practically useful.

14.2.4 Fitting Multivariate GARCH Models

Model fitting. We have already given notes on fitting some models in stages and it should be stressed that in the high-dimensional applications of risk management this may in fact be the only feasible strategy. Where interest centres on a multivariate risk-factor return series of more modest dimension (fewer than ten dimensions, say), we can attempt to fit multivariate GARCH models by maximizing an appropriate likelihood with respect to all parameters in a single step. The procedure follows from the method for univariate time series described in Section 4.2.4.

The method of building a likelihood for a generic multivariate GARCH model $\boldsymbol{X}_t = \Sigma_t^{1/2} \boldsymbol{Z}_t$ is completely analogous to the univariate case; consider again a first-order model ($p = q = 1$) for simplicity and assume that our data are labelled $\boldsymbol{X}_0, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. A conditional likelihood is based on the conditional joint density of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ given $\boldsymbol{X}_0$ and an initial value Σ_0 for the conditional covariance

matrix. This conditional joint density is

$$f_{\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n|\boldsymbol{X}_0,\Sigma_0}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n \mid \boldsymbol{x}_0, \Sigma_0) = \prod_{t=1}^{n} f_{\boldsymbol{X}_t|\boldsymbol{X}_{t-1},\ldots,\boldsymbol{X}_0,\Sigma_0}(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1},\ldots,\boldsymbol{x}_0, \Sigma_0).$$

If we denote the multivariate innovation density of $\boldsymbol{Z}_t$ by $g(z)$, then we have

$$f_{\boldsymbol{X}_t|\boldsymbol{X}_{t-1},\ldots,\boldsymbol{X}_0,\Sigma_0}(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1},\ldots,\boldsymbol{x}_0, \Sigma_0) = |\Sigma_t|^{-1/2} g(\Sigma_t^{-1/2}\boldsymbol{x}_t),$$

where Σ_t is a matrix-valued function of $\boldsymbol{x}_{t-1},\ldots,\boldsymbol{x}_0$ and Σ_0. Most common choices of $g(z)$ are in the spherical family, so by (6.39) we have $g(z) = h(z'z)$ for some function h of a scalar variable (known as a density generator), yielding a conditional likelihood of the form

$$L(\boldsymbol{\theta}; \boldsymbol{X}_1,\ldots,\boldsymbol{X}_n) = \prod_{t=1}^{n} |\Sigma_t|^{-1/2} h(\boldsymbol{X}_t'\Sigma_t^{-1}\boldsymbol{X}_t),$$

where all parameters appearing in the volatility equation and the innovation distribution are collected in $\boldsymbol{\theta}$. It would of course be possible to add a constant mean term or a conditional mean term with, say, vector autoregressive structure to the model and to adapt the likelihood accordingly.

Evaluation of the likelihood requires us to input a value for Σ_0. Maximization can again be performed in practice using a modified Newton–Raphson procedure, such as that of Berndt et al. (1974). References concerning properties of estimators are given in Notes and Comments, although the literature for multivariate GARCH is small.

Model checking and comparison. Residuals are calculated according to $\hat{Z}_t = \hat{\Sigma}_t^{-1/2}\boldsymbol{X}_t$ and should behave like a realization of an SWN$(\boldsymbol{0}, I_d)$ process. The usual univariate procedures (correlograms, correlograms of absolute values, and portmanteau tests such as Ljung–Box) can be applied to the component series of the residuals. Also, there should not be any evidence of cross-correlations at any lags for either the raw or the absolute residuals in the cross-correlogram.

Model selection is usually performed by a standard comparison of AIC numbers, although there is limited literature on theoretical aspects of the use of the AIC in a multivariate GARCH context.

14.2.5 Dimension Reduction in MGARCH

In view of the large number of parameters in MGARCH models and the practical difficulties involved in their estimation, as well as the fact that many time series of risk-factor changes are likely to have high levels of cross-correlation at lag zero, it makes sense to consider how dimension-reduction strategies from the area of factor modelling (see Section 6.4.1) can be applied in the context of multivariate GARCH.

As discussed in Section 6.4.2 there are a number of different statistical approaches to building a factor model. We give brief notes on strategies for constructing (1) a *macroeconomic* factor model and (2) a *statistical* factor model based on *principal component analysis*.

Macroeconomic factor model. In this approach we attempt to identify a small number of common factors $\boldsymbol{F}_t$ that can explain the variation in many risk factors $\boldsymbol{X}_t$; we might, for example, use stock index returns to explain the variation in individual equity returns. These common factors could be modelled using relatively detailed multivariate models incorporating ARMA and GARCH features.

The dependence of the individual returns on the factor returns can then be modelled by calibrating a factor model of the type

$$\boldsymbol{X}_t = \boldsymbol{a} + B\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, n.$$

In Section 6.4.3 we showed how this may be done in a static way using regression techniques. We now assume that, conditional on the factors $\boldsymbol{F}_t$, the errors $\boldsymbol{\varepsilon}_t$ form a multivariate white noise process with GARCH volatility structure.

In a perfect factor model (corresponding to Definition 6.31) these errors would have a diagonal covariance matrix and would be attributable to idiosyncratic effects alone. In GARCH terms we could assume they follow a pure diagonal model, i.e. a CCC model where the constant conditional correlation matrix is the identity matrix. A pure diagonal model can be fitted in two ways, which correspond to the two ways of estimating a static regression model in Section 6.4.3.

(1) Fit univariate models to the component series $X_{1,k}, \ldots, X_{n,k}$, $k = 1, \ldots, d$. For each k assume that

$$X_{t,k} = \mu_{t,k} + \varepsilon_{t,k}, \quad \mu_{t,k} = a_k + \boldsymbol{b}_k' \boldsymbol{F}_t, \quad t = 1, \ldots, n,$$

where the errors $\varepsilon_{t,k}$ follow some univariate GARCH specification. Most software for GARCH estimation allows covariates to be incorporated into the model for the conditional mean term $\mu_{t,k}$.

(2) Fit in one step the multivariate model

$$\boldsymbol{X}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\mu}_t = \boldsymbol{a} + B\boldsymbol{F}_t, \quad t = 1, \ldots, n,$$

where the errors $\boldsymbol{\varepsilon}_t$ follow a pure diagonal CCC model and the SWN($\mathbf{0}$, I_d) process driving the GARCH model is some non-Gaussian spherical distribution, such as an appropriate scaled t distribution. (If the SWN is Gaussian, approaches (1) and (2) give the same results.)

In practice, it is never possible to find the "right" common factors such that the idiosyncratic errors have a diagonal covariance structure. The pure diagonal assumption can be examined by looking at the errors from the GARCH modelling, estimating their correlation matrix and assessing its closeness to the identity matrix. In the case where correlation structure remains, the formal concept of the factor model can be loosened by allowing errors with a CCC–GARCH structure, which can be calibrated by two-stage estimation as described in Section 14.2.2.

Principal components GARCH (PC-GARCH). In Section 6.4.5 we explained the application of principal component analysis to time-series data under the assumption that the data came from a stationary multivariate time-series process. The idea of principal components GARCH is that we first compute time series of sample principal components and then model the most important series—the ones with highest sample variance—using GARCH models. Usually we attempt to fit a series of independent univariate GARCH models to the sample principal components.

In terms of an underlying model we are implicitly assuming that the data are realizations from a so-called PC-GARCH or orthogonal GARCH model defined as follows.

Definition 14.18. The process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ follows a PC-GARCH (or orthogonal GARCH) model if there exists some orthogonal matrix $\Gamma \in \mathbb{R}^{d\times d}$ satisfying $\Gamma\Gamma' = \Gamma'\Gamma = I_d$ such that $(\Gamma'\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ follows a pure diagonal GARCH model. The process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ follows a PC-GARCH model with mean $\boldsymbol{\mu}$ if $(\boldsymbol{X}_t - \boldsymbol{\mu})_{t\in\mathbb{Z}}$ follows a PC-GARCH model.

If $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ follows a PC-GARCH process for some matrix Γ, then we can introduce the process $(\boldsymbol{Y}_t)_{t\in\mathbb{Z}}$ defined by $\boldsymbol{Y}_t = \Gamma'\boldsymbol{X}_t$, which satisfies $\boldsymbol{Y}_t = \Delta_t\boldsymbol{Z}_t$, where $(\boldsymbol{Z}_t)_{t\in\mathbb{Z}}$ is SWN$(\mathbf{0}, I_d)$ and Δ_t is a (diagonal) volatility matrix with elements that are updated according to univariate GARCH schemes and past values of the components of $\boldsymbol{Y}_t$. Since $\boldsymbol{X}_t = \Gamma\Delta_t\boldsymbol{Z}_t$, the conditional and unconditional covariance matrices have the structure

$$\Sigma_t = \Gamma\Delta_t^2\Gamma', \qquad \Sigma = \Gamma E(\Delta_t^2)\Gamma', \tag{14.24}$$

and they are obviously symmetric and positive definite.

Comparing (14.24) with (6.59) we see that the PC-GARCH model implies a spectral decomposition of the conditional and unconditional covariance matrices. The *eigenvalues* of the conditional covariance matrix, which are the elements of the diagonal matrix Δ_t^2, are given a GARCH updating structure. The *eigenvectors* form the columns of Γ and are used to construct the time series $(\boldsymbol{Y}_t)_{t\in\mathbb{Z}}$, the principal components transform of $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$. It should be noted that despite the simple structure of (14.24), the conditional correlation matrix of $\boldsymbol{X}_t$ is not constant in this model.

As noted above, this model is estimated in two stages. Let us suppose we have risk-factor change data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ assumed to come from a PC-GARCH model with mean $\boldsymbol{\mu}$. In the first step we calculate the spectral decomposition of the sample covariance matrix S of the data as in Section 6.4.5; this gives us an estimator G of Γ. We then rotate the data to obtain sample principal components $\{\boldsymbol{Y}_t = G'(\boldsymbol{X}_t - \bar{\boldsymbol{X}}) : t = 1, \ldots, n\}$, where $\bar{\boldsymbol{X}} = n^{-1}\sum_{t=1}^n \boldsymbol{X}_t$. These should be consistent with a pure diagonal model if the PC-GARCH model is appropriate for the original data; there should be no cross-correlation between the series at any lag. In a second stage we fit univariate GARCH models to the time series of principal components; the residuals from these GARCH models should behave like SWN$(\mathbf{0}, I_d)$.

To turn the model output into a factor model we use the idea embodied in equation (6.65)—that the first k loading vectors in the matrix G specify the most important

principal components—and we write these columns in the submatrix $G_1 \in \mathbb{R}^{d\times k}$. We define the factors to be $\boldsymbol{F}_t := (Y_{t,1}, \ldots, Y_{t,k})'$, $t = 1, \ldots, n$, the first k principal component time series. We have now calibrated an approximate factor model of the form $\boldsymbol{X}_t = \bar{\boldsymbol{X}} + G_1\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t$, where the components of $\boldsymbol{F}_t$ are assumed to follow a pure diagonal GARCH model of dimension k. We can use this model to forecast the behaviour of the risk-factor changes $\boldsymbol{X}_t$ by first forecasting the behaviour of the factors $\boldsymbol{F}_t$; the error term is usually ignored in practice.

14.2.6 MGARCH and Conditional Risk Measurement

Suppose we calibrate an MGARCH model (possibly with VARMA conditional mean structure) having the general structure $\boldsymbol{X}_t = \boldsymbol{\mu}_t + \Sigma_t^{1/2}\boldsymbol{Z}_t$ to historical risk-factor return data $\boldsymbol{X}_{t-n+1}, \ldots, \boldsymbol{X}_t$. We are interested in the loss distribution of $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$ conditional on $\mathcal{F}_t = \sigma(\{\boldsymbol{X}_s : s \leqslant t\})$, as described in Sections 9.1 and 9.2.1. (We may also be interested in longer-period losses as in Section 9.2.7.)

A general method that could be applied is the Monte Carlo method of Section 9.2.5: we could simulate many times the next value $\boldsymbol{X}_{t+1}$ (and subsequent values for losses over longer periods) of the stochastic process $(\boldsymbol{X}_t)_{t\in\mathbb{Z}}$ using estimates of $\boldsymbol{\mu}_{t+1}$ and Σ_{t+1}.

Alternatively, a variance–covariance calculation could be made as in Section 9.2.2. Considering a linearized loss operator with the general form $l^{\Delta}_{[t]}(\boldsymbol{x}) = -(c_t + \boldsymbol{b}_t'\boldsymbol{x})$, the moments of the conditional loss distribution would be

$$E(L^{\Delta}_{t+1} \mid \mathcal{F}_t) = -c_t - \boldsymbol{b}_t'\boldsymbol{\mu}_{t+1}, \qquad \operatorname{cov}(L^{\Delta}_{t+1} \mid \mathcal{F}_t) = \boldsymbol{b}_t'\Sigma_{t+1}\boldsymbol{b}_t.$$

Under an assumption of multivariate Gaussian innovations, the conditional distribution of L^{Δ}_{t+1} given $\mathcal{F}_t$ would be univariate Gaussian as in equation (2.14). Under an assumption of (scaled) multivariate t innovations, it would be univariate t. To calculate VaR and ES we can follow the calculations in Examples 2.11, 2.14 and 2.15 but we would first need to compute estimates of Σ_{t+1} and $\boldsymbol{\mu}_{t+1}$ using our multivariate time-series model, as indicated in Example 14.19.

Example 14.19. Consider the simple stock portfolio in Example 2.1, where $c_t = 0$ and $\boldsymbol{b}_t = V_t\boldsymbol{w}_t$, and suppose our time-series model is a first-order DVEC model with a constant mean term. The model takes the form

$$\boldsymbol{X}_t - \boldsymbol{\mu} = \Sigma_t^{1/2}\boldsymbol{Z}_t, \qquad \Sigma_t = A_0 + A_1 \circ ((\boldsymbol{X}_{t-1} - \boldsymbol{\mu})(\boldsymbol{X}_{t-1}' - \boldsymbol{\mu}')) + B \circ \Sigma_{t-1}. \tag{14.25}$$

Suppose we assume that the innovations are multivariate Student t. The standard risk measures applied to the linearized loss distribution would take the form

$$\mathrm{VaR}^t_\alpha = -V_t\boldsymbol{w}_t'\boldsymbol{\mu} + V_t\sqrt{\frac{\boldsymbol{w}_t'\Sigma_{t+1}\boldsymbol{w}_t(\nu-2)}{\nu}}\,t_\nu^{-1}(\alpha),$$

$$\mathrm{ES}^t_\alpha = -V_t\boldsymbol{w}_t'\boldsymbol{\mu} + V_t\sqrt{\frac{\boldsymbol{w}_t'\Sigma_{t+1}\boldsymbol{w}_t(\nu-2)}{\nu}}\,\frac{g_\nu(t_\nu^{-1}(\alpha))}{1-\alpha}\left(\frac{\nu + (t_\nu^{-1}(\alpha))^2}{\nu-1}\right),$$

where the notation is as in Example 2.15. Estimates of the risk measures are obtained by replacing $\boldsymbol{\mu}$, ν and Σ_{t+1} by estimates. The latter can be calculated iteratively from (14.25) using estimates of A_0, A_1 and B and a starting value for Σ_0.

As an alternative to formal estimation of MGARCH models we can also use the technique of multivariate exponentially weighted moving average (EWMA) forecasting to compute an estimate of Σ_{t+1}. The recursive procedure has been described in Section 9.2.2.

Notes and Comments

The CCC–GARCH model was suggested by Bollerslev (1990), who used it to model European exchange-rate data before and after the introduction of the European Monetary System (EMS) and came to the expected conclusion that conditional correlations after the introduction of the EMS were higher. The idea of the DCC model is explored by Engle (2002), Engle and Sheppard (2001) and Tse and Tsui (2002). Fitting in stages is promoted in the formulation of Engle and Sheppard (2001), and asymptotic statistical theory for this procedure is given. Hafner and Franses (2009) suggest that the dynamics of CCC are too simple for collections of many asset returns and give a generalization. See also the textbook by Engle (2009).

The DVEC model was proposed by Bollerslev, Engle and Wooldridge (1988). The more general (but overparametrized) VEC model is discussed in Engle and Kroner (1995) alongside the BEKK model, named after these two authors as well as Baba and Kraft, who co-authored an earlier unpublished manuscript. The condition for the positive definiteness of Σ_t in (14.15), which suggests the parametrizations (14.17)–(14.19), is described in Attanasio (1991).

There is limited work on statistical properties of quasi-maximum likelihood estimates in multivariate models: Jeantheau (1998) shows consistency for a general formulation and Comte and Lieberman (2003) show asymptotic normality for the BEKK formulation.

The PC-GARCH model was first described by Ding (1994) in a PhD thesis; under the name of orthogonal GARCH it has been extensively investigated by Alexander (2001). The latter shows how PC-GARCH can be used as a dimension-reduction tool for expressing the conditional covariances of a number of asset return series in terms of a much smaller number of principal component return series.

Survey articles by Bollerslev, Engle and Nelson (1994) and Bauwens, Laurent and Rombouts (2006) are useful sources of additional information and references for all of these multivariate models.

15
Advanced Topics in Multivariate Modelling

This chapter contains selected advanced topics that extend the presentation of multivariate models and copulas in Chapters 6 and 7. In Section 15.1 we treat topics that relate to multivariate normal mixture distributions and elliptical distributions. In Section 15.2 we extend the theory of Archimedean copulas and consider models that are more general or more flexible than those presented in Section 7.4.

15.1 Normal Mixture and Elliptical Distributions

This section addresses two statistical issues that were raised in Chapter 6. In Section 15.1.1 we give the algorithm for fitting multivariate generalized hyperbolic distributions to data, which was used to create the examples in Section 6.2.4. In Section 15.1.2 we discuss empirical tests for elliptical symmetry.

15.1.1 Estimation of Generalized Hyperbolic Distributions

While univariate GH models have been fitted to return data in many empirical studies, there has been relatively little applied work with the multivariate distributions. However, normal mixture distributions of the kind we have described may be fitted with algorithms of the EM (expectation–maximization) type. In this section we present an algorithm for that purpose and sketch the ideas behind it. Similar methods have been developed independently by other authors and references may be found in Notes and Comments.

Assume we have iid data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ and wish to fit the multivariate GH distribution, or one of its special cases. Summarizing the parameters by $\boldsymbol{\theta} = (\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})'$, the problem is to maximize

$$\ln L(\boldsymbol{\theta}; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \sum_{i=1}^{n} \ln f_{\boldsymbol{X}}(\boldsymbol{X}_i; \boldsymbol{\theta}), \tag{15.1}$$

where $f_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\theta})$ denotes the GH density in (6.29).

This problem is not particularly easy at first sight due to the number of parameters and the necessity of maximizing over covariance matrices Σ. However, if we were able to "observe" the latent mixing variables $W_1, \ldots, W_n$ coming from the mixture representation in (6.24), it would be much easier. Since the joint density of any pair $\boldsymbol{X}_i$ and W_i is given by

$$f_{\boldsymbol{X},W}(\boldsymbol{x}, w; \boldsymbol{\theta}) = f_{\boldsymbol{X}|W}(\boldsymbol{x} \mid w; \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}) h_W(w; \lambda, \chi, \psi), \tag{15.2}$$

we could construct the likelihood

$$\ln \tilde{L}(\boldsymbol{\theta}; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, W_1, \ldots, W_n) = \sum_{i=1}^{n} \ln f_{\boldsymbol{X}|W}(\boldsymbol{X}_i \mid W_i; \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}) + \sum_{i=1}^{n} \ln h_W(W_i; \lambda, \chi, \psi), \quad (15.3)$$

where the two terms could be maximized separately with respect to the parameters they involve. The apparently more problematic parameters of Σ and $\boldsymbol{\gamma}$ are in the first term of the likelihood, and estimates are relatively easy to derive due to the Gaussian form of this term.

To overcome the latency of the W_i data the EM algorithm is used. This is an iterative procedure consisting of an E-step, or expectation step (where essentially W_i is replaced with an estimate given the observed data and current parameter estimates), and an M-step, or maximization step (where the parameter estimates are updated). Suppose at the beginning of step k we have the vector of parameter estimates $\boldsymbol{\theta}^{[k]}$. We proceed as follows.

E-step. We calculate the conditional expectation of the so-called augmented likelihood (15.3) given the data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ using the parameter values $\boldsymbol{\theta}^{[k]}$. This results in the objective function

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{[k]}) = E(\ln \tilde{L}(\boldsymbol{\theta}; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, W_1, \ldots, W_n) \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{\theta}^{[k]}).$$

M-step. We maximize the objective function with respect to $\boldsymbol{\theta}$ to obtain the next set of estimates $\boldsymbol{\theta}^{[k+1]}$.

Alternating between these steps, the EM algorithm produces improved parameter estimates at each step (in the sense that the value of the original likelihood (15.1) is continually increased) and they converge to the maximum likelihood (ML) estimates.

In practice, performing the E-step amounts to replacing any functions $g(W_i)$ of the latent mixing variables that arise in (15.3) by the quantities $E(g(W_i) \mid \boldsymbol{X}_i; \boldsymbol{\theta}^{[k]})$. To calculate these quantities we can observe that the conditional density of W_i given $\boldsymbol{X}_i$ satisfies $f_{W|\boldsymbol{X}}(w \mid \boldsymbol{x}; \boldsymbol{\theta}) \propto f_{W,\boldsymbol{X}}(w, \boldsymbol{x}; \boldsymbol{\theta})$, up to some constant of proportionality. It may therefore be deduced from (15.2) that

$$W_i \mid \boldsymbol{X}_i \sim N^-(\lambda - \tfrac{1}{2}d, (\boldsymbol{X}_i - \boldsymbol{\mu})' \tilde{\Sigma}^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}) + \chi, \psi + \boldsymbol{\gamma}' \Sigma^{-1} \boldsymbol{\gamma}). \quad (15.4)$$

If we write out the likelihood (15.3) using (6.25) for the first term and the generalized inverse Gaussian (GIG) density (A.14) for the second term, we find that the functions $g(W_i)$ arising in (15.3) are $g_1(w) = w$, $g_2(w) = 1/w$ and $g_3(w) = \ln(w)$. The conditional expectation of these functions in model (15.4) may be evaluated using information about the GIG distribution in Section A.2.5; note that $E(\ln(W_i) \mid \boldsymbol{X}_i; \boldsymbol{\theta}^{[k]})$ involves derivatives of a Bessel function with respect to order and must be approximated numerically. We will introduce the notation

$$\delta_i^{[\cdot]} = E(W_i^{-1} \mid \boldsymbol{X}_i; \boldsymbol{\theta}^{[\cdot]}), \quad \eta_i^{[\cdot]} = E(W_i \mid \boldsymbol{X}_i; \boldsymbol{\theta}^{[\cdot]}), \quad \xi_i^{[\cdot]} = E(\ln(W_i) \mid \boldsymbol{X}_i; \boldsymbol{\theta}^{[\cdot]}), \quad (15.5)$$

which allows us to describe the basic EM scheme as well as a variant below.

In the M-step there are two terms to maximize, coming from the two terms in (15.3); we write these as $Q_1(\boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}; \boldsymbol{\theta}^{[k]})$ and $Q_2(\lambda, \chi, \psi; \boldsymbol{\theta}^{[k]})$. To address the identifiability issue mentioned in Section 6.2.3 we constrain the determinant of Σ to be some fixed value (in practice, we take the determinant of the sample covariance matrix S) in the maximization of Q_1. The maximizing values of $\boldsymbol{\mu}$, Σ and $\boldsymbol{\gamma}$ may then be derived analytically by calculating partial derivatives and setting these equal to zero; the resulting formulas are embedded in Algorithm 15.1 below (see steps (3) and (4)). The maximization of $Q_2(\lambda, \chi, \psi; \boldsymbol{\theta}^{[k]})$ with respect to the parameters of the mixing distribution is performed numerically; the function $Q_2(\lambda, \chi, \psi; \boldsymbol{\theta}^{[\cdot]})$ is

$$(\lambda - 1)\sum_{i=1}^{n} \xi_i^{[\cdot]} - \tfrac{1}{2}\chi \sum_{i=1}^{n} \delta_i^{[\cdot]} - \tfrac{1}{2}\psi \sum_{i=1}^{n} \eta_i^{[\cdot]} \\ - \tfrac{1}{2}n\lambda \ln(\chi) + \tfrac{1}{2}n\lambda \ln(\psi) - n \ln(2K_\lambda(\sqrt{\chi\psi})). \quad (15.6)$$

This would complete one iteration of a standard EM algorithm. However, there are a couple of variants on the basic scheme; both involve modification of the final step described above, namely the maximization of Q_2.

Assuming the parameters $\boldsymbol{\mu}$, Σ and $\boldsymbol{\gamma}$ have been updated first in iteration k, we define

$$\boldsymbol{\theta}^{[k,2]} = (\lambda^{[k]}, \chi^{[k]}, \psi^{[k]}, \boldsymbol{\mu}^{[k+1]}, \Sigma^{[k+1]}, \boldsymbol{\gamma}^{[k+1]})',$$

recalculate the weights $\delta_i^{[k,2]}$, $\eta_i^{[k,2]}$ and $\xi_i^{[k,2]}$ in (15.5), and then maximize the function $Q_2(\lambda, \eta, \xi; \boldsymbol{\theta}^{[k,2]})$ in (15.6). This results in a so-called MCECM algorithm (multi-cycle, expectation, conditional maximization), which is the one we present below.

Alternatively, instead of maximizing Q_2 we may maximize the original likelihood (15.1) with respect to λ, χ and ψ with the other parameters held fixed at the values $\boldsymbol{\mu}^{[k]}$, $\Sigma^{[k]}$ and $\boldsymbol{\gamma}^{[k]}$; this results in an ECME algorithm.

Algorithm 15.1 (EM estimation of GH distribution).

(1) Set iteration count $k = 1$ and select starting values for $\boldsymbol{\theta}^{[1]}$. In particular, reasonable starting values for $\boldsymbol{\mu}$, $\boldsymbol{\gamma}$ and Σ, respectively, are the sample mean, the zero vector and the sample covariance matrix S.

(2) Calculate weights $\delta_i^{[k]}$ and $\eta_i^{[k]}$ using (15.5), (15.4) and (A.15). Average the weights to get

$$\bar{\delta}^{[k]} = n^{-1}\sum_{i=1}^{n} \delta_i^{[k]} \quad \text{and} \quad \bar{\eta}^{[k]} = n^{-1}\sum_{i=1}^{n} \eta_i^{[k]}.$$

(3) For a symmetric model set $\boldsymbol{\gamma}^{[k+1]} = \mathbf{0}$. Otherwise set

$$\boldsymbol{\gamma}^{[k+1]} = \frac{n^{-1}\sum_{i=1}^{n} \delta_i^{[k]}(\bar{\boldsymbol{X}} - \boldsymbol{X}_i)}{\bar{\delta}^{[k]}\bar{\eta}^{[k]} - 1}.$$

(4) Update estimates of the location vector and dispersion matrix by

$$\boldsymbol{\mu}^{[k+1]} = \frac{n^{-1}\sum_{i=1}^{n}\delta_i^{[k]}\boldsymbol{X}_i - \boldsymbol{\gamma}^{[k+1]}}{\bar{\delta}^{[k]}},$$

$$\Psi = \frac{1}{n}\sum_{i=1}^{n}\delta_i^{[k]}(\boldsymbol{X}_i - \boldsymbol{\mu}^{[k+1]})(\boldsymbol{X}_i - \boldsymbol{\mu}^{[k+1]})' - \bar{\eta}^{[k]}\boldsymbol{\gamma}^{[k+1]}\boldsymbol{\gamma}^{[k+1]\prime},$$

$$\Sigma^{[k+1]} = \frac{|S|^{1/d}\Psi}{|\Psi|^{1/d}}.$$

(5) Set

$$\boldsymbol{\theta}^{[k,2]} = (\lambda^{[k]}, \chi^{[k]}, \psi^{[k]}, \boldsymbol{\mu}^{[k+1]}, \Sigma^{[k+1]}, \boldsymbol{\gamma}^{[k+1]})'.$$

Calculate weights $\delta_i^{[k,2]}$, $\eta_i^{[k,2]}$ and $\xi_i^{[k,2]}$ using (15.5), (15.4) and information in Section A.2.5.

(6) Maximize $Q_2(\lambda, \chi, \psi; \boldsymbol{\theta}^{[k,2]})$ in (15.6) with respect to λ, χ and ψ to complete the calculation of $\boldsymbol{\theta}^{[k,2]}$. Increment iteration count $k \to k+1$ and go to step (2).

This algorithm may be easily adapted to fit special cases of the GH distribution. This involves holding certain parameters fixed throughout and maximizing with respect to the remaining parameters: for the hyperbolic distribution we set $\lambda = 1$; for the NIG distribution $\lambda = -\frac{1}{2}$; for the t distribution $\psi = 0$; for the VG distribution $\chi = 0$. In the case of the t and VG distributions, in step (6) we have to work with the function Q_2 that results from assuming an inverse gamma or gamma density for h_W.

15.1.2 Testing for Elliptical Symmetry

The general problem of this section is to test whether a sample of identically distributed data vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ has an elliptical distribution $E_d(\boldsymbol{\mu}, \Sigma, \psi)$ for some $\boldsymbol{\mu}$, Σ and generator ψ. In all of the methods we require estimates of $\boldsymbol{\mu}$ and Σ, and these can be obtained using approaches discussed in Section 6.3.4, such as fitting t distributions, calculating M-estimates or perhaps using (6.49) in the bivariate case. We denote the estimates simply by $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$.

In finance we cannot generally assume that the observations are of iid random vectors, but we assume that they at least have an identical distribution. Note that, even if the data were independent, the fact that we generally estimate $\boldsymbol{\mu}$ and Σ from the whole data set would introduce dependence in the procedures that we describe below.

Stable correlation estimates: an exploratory method. An easy exploratory graphical method can be based on Proposition 6.28. We could attempt to estimate

$$\rho(\boldsymbol{X} \mid h(\boldsymbol{X}) \geqslant c), \qquad h(\boldsymbol{x}) = (\boldsymbol{x} - \hat{\boldsymbol{\mu}})'\hat{\Sigma}^{-1}(\boldsymbol{x} - \hat{\boldsymbol{\mu}})$$

for various values of $c \geqslant 0$. We expect that for elliptically distributed data the estimates will remain roughly stable over a range of different c values. Of course the estimates of this correlation should again be calculated using some method that is more efficient than the standard correlation estimator for heavy-tailed data. The

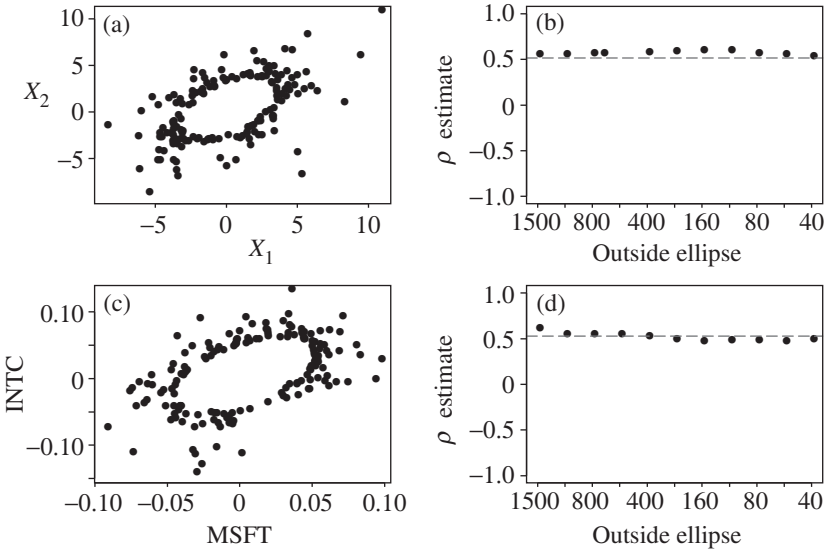

Figure 15.1. Correlations are estimated using the Kendall's tau method for points lying outside ellipses of progressively larger size (as shown in (a) and (c)). (a), (b) Two thousand t-distributed data with four degrees of freedom and $\rho = 0.5$. (c), (d) Two thousand daily log-returns on Microsoft and Intel. Points show estimates for an ellipse that is allowed to grow until there are only forty points outside; dashed lines show estimates of correlation for all data.

method is most natural as a bivariate method, and in this case the correlation of $\boldsymbol{X} \mid h(\boldsymbol{X}) \geqslant c$ can be estimated by applying the Kendall's tau transform method to those data points $\boldsymbol{X}_i$ that lie outside the ellipse defined by $h(\boldsymbol{x}) = c$. In Figure 15.1 we give an example with both simulated and real data, neither of which show any marked departure from the assumption of stable correlations. The method is of course exploratory and does not allow us to come to any formal conclusion.

Q–Q plots. The remaining methods that we describe rely on the link (6.41) between non-singular elliptical and spherical distributions. If $\boldsymbol{\mu}$ and Σ were known, then we would test for elliptical symmetry by testing the data $\{\Sigma^{-1/2}(\boldsymbol{X}_i - \boldsymbol{\mu}) : i = 1, \ldots, n\}$ for spherical symmetry. Replacing these parameters by estimates as above, we consider whether the data

$$\{\boldsymbol{Y}_i = \hat{\Sigma}^{-1/2}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}) : i = 1, \ldots, n\} \tag{15.7}$$

are consistent with a spherical distribution, while ignoring the effect of estimation error.

Some graphical methods based on Q–Q plots have been suggested by Li, Fang and Zhu (1997) and these are particularly useful for large d. These essentially rely on the following result.

Lemma 15.2. *Suppose that $T(\boldsymbol{Y})$ is a statistic such that, almost surely,*

$$T(a\boldsymbol{Y}) = T(\boldsymbol{Y}) \quad \text{for every } a > 0. \tag{15.8}$$

Then $T(\boldsymbol{Y})$ has the same distribution for every spherical vector $\boldsymbol{Y} \sim S_d^+(\psi)$.

Proof. From Theorem 6.21 we have $T(\boldsymbol{Y}) \stackrel{\mathrm{d}}{=} T(R\boldsymbol{S})$, and $T(R\boldsymbol{S}) \stackrel{\text{a.s.}}{=} T(\boldsymbol{S})$ follows from (15.8). Since the distribution of $T(\boldsymbol{Y})$ only depends on $\boldsymbol{S}$ and not R, it must be the same for all $\boldsymbol{Y} \sim S_d^+(\psi)$. □

We exploit this result by looking for statistics $T(\boldsymbol{Y})$ with the property (15.8), whose distribution we know when $\boldsymbol{Y} \sim N_d(\boldsymbol{0}, I_d)$. Two examples are

$$\left.\begin{aligned} T_1(\boldsymbol{Y}) &= \frac{d^{1/2}\bar{Y}}{\sqrt{(1/(d-1))\sum_{i=1}^d (Y_i - \bar{Y})^2}}, \quad \bar{Y} = \frac{1}{d}\sum_{i=1}^d Y_i, \\ T_2(\boldsymbol{Y}) &= \frac{\sum_{i=1}^k Y_i^2}{\sum_{i=1}^d Y_i^2}. \end{aligned}\right\} \tag{15.9}$$

For $\boldsymbol{Y} \sim N_d(\boldsymbol{0}, I_d)$, and hence for $\boldsymbol{Y} \sim S_d^+(\psi)$, we have $T_1(\boldsymbol{Y}) \sim t_{d-1}$ and $T_2(\boldsymbol{Y}) \sim \text{Beta}(\frac{1}{2}k, \frac{1}{2}(d-k))$.

Our experience suggests that the beta plot is the more revealing of the resulting Q–Q plots. Li, Fang and Zhu (1997) suggest choosing k such that it is roughly equal to $d - k$. In Figure 15.2 we show examples of the Q–Q plots obtained for 2000 simulated data from a ten-dimensional t distribution with four degrees of freedom and for the daily, weekly and monthly return data on ten Dow Jones 30 stocks that were analysed in Example 6.3 and Section 6.2.4. The curvature in the plots for daily and weekly returns seems to be evidence against the elliptical hypothesis.

Numerical tests. We restrict ourselves to simple ideas for bivariate tests; references to more general test ideas are found in Notes and Comments. If we neglect the error involved in estimating location and dispersion, testing for elliptical symmetry amounts to testing the $\boldsymbol{Y}_i$ data in (15.7) for spherical symmetry. For $i = 1, \ldots, n$, if we set $R_i = \|\boldsymbol{Y}_i\|$ and $\boldsymbol{S}_i = \boldsymbol{Y}_i / \|\boldsymbol{Y}_i\|$, then under the null hypothesis the $\boldsymbol{S}_i$ data should be uniformly distributed on the unit sphere $\mathcal{S}^{d-1}$, and the paired data $(R_i, \boldsymbol{S}_i)$ should form realizations of independent pairs.

In the bivariate case, testing for uniformity on the unit circle $\mathcal{S}^1$ amounts to a univariate test of uniformity on $[0, 2\pi]$ for the angles Θ_i described by the points $\boldsymbol{S}_i = (\cos\Theta_i, \sin\Theta_i)'$ on the perimeter of the circle; equivalently, we may test the data $\{U_i := \Theta_i/(2\pi) \colon i = 1, \ldots, n\}$ for uniformity on $[0, 1]$. Neglecting issues of serial dependence in the data, this may be done, for instance, by a standard chi-squared goodness-of-fit test (see Rice 1995, p. 241) or a Kolmogorov–Smirnov test (see Conover 1999). Again neglecting issues of serial dependence, the independence of the components of the pairs $\{(R_i, U_i) \colon i = 1, \ldots, n\}$ could be examined by performing a test of association with Spearman's rank correlation coefficient (see, for example, Conover 1999, pp. 312–328).

We have performed these tests for the two data sets used in Figure 15.1, these being 2000 simulated bivariate t data with four degrees of freedom and 2000 daily log-returns for Intel and Microsoft. In Figure 15.3 the transformed data on the unit circle $\boldsymbol{S}_i$ and the implied angles U_i on the $[0, 1]$ scale are shown; the dispersion matrices have been estimated using the construction (6.49) based on Kendall's tau.

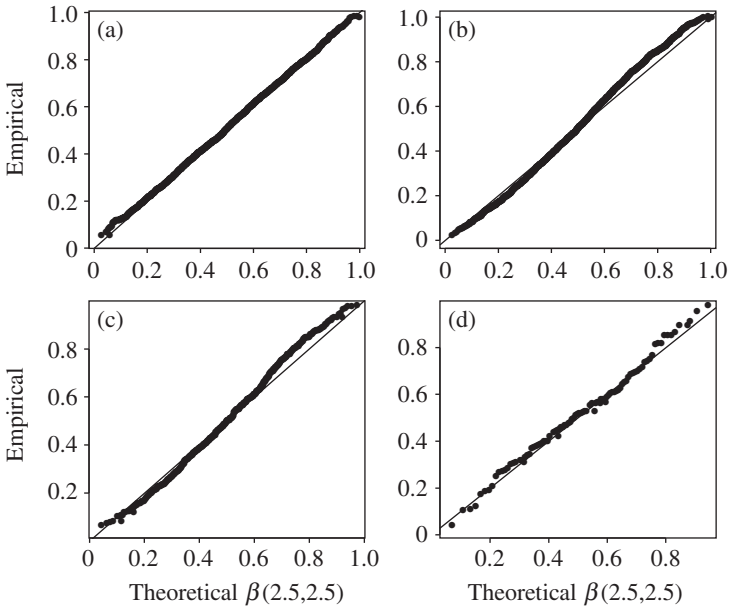

Figure 15.2. Q–Q plots of the beta statistic (15.9) for four data sets with dimension $d = 10$; we have set $k = 5$. (a) Two thousand simulated observations from a t distribution with four degrees of freedom. (b) Daily, (c) weekly and (d) monthly returns on Dow Jones stocks as analysed in Example 6.3 and Section 6.2.4. Daily and weekly returns show evidence against elliptical symmetry.

Neither of these data sets shows significant evidence against the elliptical hypothesis. For the bivariate t data the p-values for the chi-squared and Kolmogorov–Smirnov tests of uniformity and the Spearman's rank test of association are, respectively, 0.99, 0.90 and 0.10. For the stock-return data they are 0.08, 0.12 and 0.19. Note that simulated data from lightly skewed members of the GH family often fail these tests.

Notes and Comments

EM algorithms for the multivariate GH distribution have been independently proposed by Protassov (2004) and Neil Shephard (personal communication). Our approach is based on EM-type algorithms for fitting the multivariate t distribution with unknown degrees of freedom. A good starter reference on this subject is Liu and Rubin (1995), where the use of the MCECM algorithm of Meng and Rubin (1993) and the ECME algorithm proposed in Liu and Rubin (1994) is discussed. Further refinements of these algorithms are discussed in Liu (1997) and Meng and van Dyk (1997).

The Q–Q plots for testing spherical symmetry were suggested by Li, Fang and Zhu (1997). There is a large literature on tests of spherical symmetry, including Smith (1977), Kariya and Eaton (1977), Beran (1979) and Baringhaus (1991). This work is also related to tests of uniformity for directional data: see Mardia (1972), Giné (1975) and Prentice (1978).

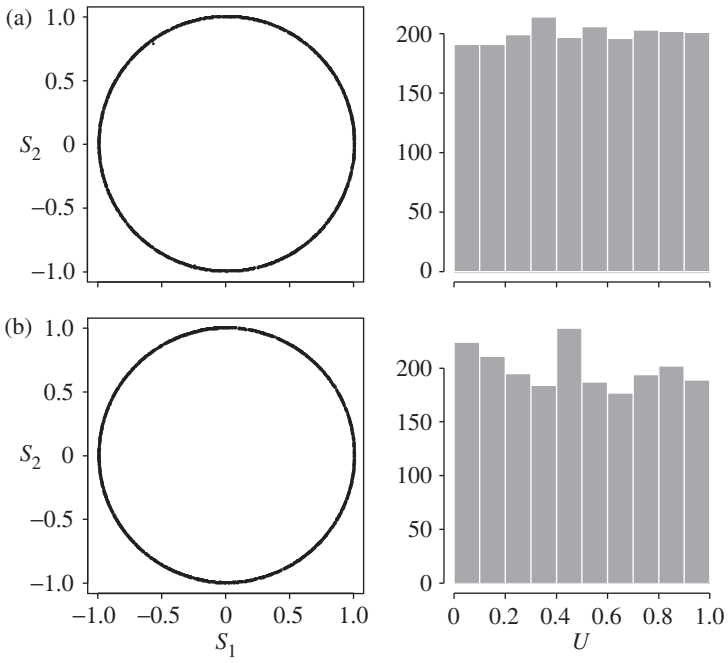

Figure 15.3. Illustration of the transformation of bivariate data to points on the unit circle $\mathcal{S}^1$ using the transformation $\boldsymbol{S}_i = \boldsymbol{Y}_i/\|\boldsymbol{Y}_i\|$, where the $\boldsymbol{Y}_i$ data are defined in (15.7); the angles of these points are then transformed on to the [0, 1] scale, where they can be tested for uniformity. (a) Two thousand simulated t data with four degrees of freedom. (b) Two thousand Intel and Microsoft log-returns. Neither show strong evidence against elliptical symmetry.

15.2 Advanced Archimedean Copula Models

In Section 7.4.2 we gave a characterization result for Archimedean copula generators that can be used to generate copulas in any dimension d (Theorem 7.50). In Section 15.2.1 we provide a more general result that characterizes the larger class of generators that may be used in a given dimension d.

The Archimedean copulas discussed so far have all been models for exchangeable random vectors. We discuss non-exchangeable, asymmetric extensions of the Archimedean family in Section 15.2.2.

15.2.1 Characterization of Archimedean Copulas

Recall that multivariate Archimedean copulas take the form

$$C(u_1, \ldots, u_d) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)),$$

where ψ is an Archimedean generator function, i.e. a decreasing, continuous, convex function $\psi : [0, \infty) \to [0, 1]$ satisfying $\psi(0) = 1$ and $\lim_{t\to\infty} \psi(t) = 0$. The necessary and sufficient condition for ψ to generate a copula in every dimension $d \geqslant 2$ is that it should be a completely monotonic function satisfying

$$(-1)^k \frac{\mathrm{d}^k}{\mathrm{d}t^k}\psi(t) \geqslant 0, \quad k \in \mathbb{N},\ t \in (0, \infty).$$

We also recall that these generators may be characterized as the Laplace–Stieltjes transforms of dfs G on $[0, \infty)$ such that $G(0) = 0$.

While Archimedean copulas with completely monotonic generators can be used in any dimension, if we are interested in Archimedean copulas in a given dimension d, we can relax the requirement of complete monotonicity and substitute the weaker requirement of *d-monotonicity*. A generator ψ is d-monotonic if it is differentiable up to order $(d-2)$ on $(0, \infty)$ with derivatives satisfying

$$(-1)^k \frac{\mathrm{d}^k}{\mathrm{d}t^k}\psi(t) \geqslant 0, \quad k = 0, 1, \ldots, d-2,$$

and if $(-1)^{d-2}\psi^{(d-2)}$ is decreasing and convex.

Theorem 15.3. *If $\psi : [0, \infty) \to [0, 1]$ is an Archimedean copula generator, then the construction (7.46) gives a copula in dimension d if and only if ψ is d-monotonic.*

Proof. See McNeil and Nešlehová (2009). □

If ψ is a d-monotonic generator, we write $\psi \in \Psi_d$. Examples of generators in Ψ_d that are not in Ψ_{d+1} (or Ψ_∞) are $\psi(t) = \max((1-t)^{d-1}, 0)$ and the Clayton generator $\psi(t) = \max((1+\theta t)^{-1/\theta}, 0)$ for $-1/(d-1) \leqslant \theta < -1/d$ (see McNeil and Nešlehová (2009) for details). An elegant way of characterizing the d-monotonic copula generators is in terms of a lesser-known integral transform.

Let G be a df on $[0, \infty)$ satisfying $G(0) = 0$ and let X be an rv with df G. Then, for $t \geqslant 0$ and $d \geqslant 2$, the Williamson d-transform of G (or X) is the function

$$\begin{aligned}\mathfrak{W}_d G(t) &= \int_0^\infty \max\left(\left(1 - \frac{t}{x}\right)^{d-1}, 0\right) \mathrm{d}G(x) \\ &= E\left(\max\left(\left(1 - \frac{t}{X}\right)^{d-1}, 0\right)\right).\end{aligned}$$

Every Williamson d-transform of a df G satisfying $G(0) = 0$ is a d-monotonic copula generator, and every d-monotonic copula generator is the Williamson d-transform of a unique df G. In the same way that a rich variety of copulas for arbitrary dimensions can be created by taking Laplace–Stieltjes transforms of particular distributions, a rich variety of copulas for dimension d can be created by taking Williamson d-transforms. For example, in McNeil and Nešlehová (2010) it is shown that by taking X to have a gamma, inverse-gamma or Pareto distribution we can obtain interesting families of copulas that give rise to a wide range of Spearman's or Kendall's rank correlation values.

There is an elegant connection between d-monotonic copulas generators and *simplex* distributions, which can form the basis of a sampling algorithm for the corresponding copulas. A d-dimensional random vector $\boldsymbol{X}$ is said to have a simplex distribution if $\boldsymbol{X} \overset{\mathrm{d}}{=} R\boldsymbol{S}_d$, where $\boldsymbol{S}_d$ is an rv that is distributed uniformly on the unit simplex $\mathcal{S}_d = \{\boldsymbol{x} \in \mathbb{R}_+^d : x_1 + \cdots + x_d = 1\}$ and R is an independent non-negative scalar rv with df F_R known as the radial distribution of the simplex distribution. A key theorem shows that the class of d-dimensional Archimedean copulas is equivalent

to the class of survival copulas of d-dimensional simplex distributions (excluding those with point mass at the origin).

Theorem 15.4. *If $\boldsymbol{X}$ has a simplex distribution with radial distribution F_R satisfying $F_R(0) = 0$, then the survival copula of $\boldsymbol{X}$ is Archimedean with generator $\psi = \mathfrak{W}_d F_R$.*

Proof. Let $\boldsymbol{x} \in \mathbb{R}^d_+$ and write $\boldsymbol{S}_d = (S_1, \ldots, S_d)'$. We use the fact that the survival function of $\boldsymbol{S}_d$ is given by

$$P(S_1 > x_1, \ldots, S_d > x_d) = \max\left(\left(1 - \sum_{i=1}^{d} x_i\right)^{d-1}, 0\right)$$

(Fang, Kotz and Ng 1990, p. 115) to establish that

$$\begin{aligned} P(X_1 > x_1, \ldots, X_d > x_d) &= E\left(P\left(S_1 > \frac{x_1}{R}, \ldots, S_d > \frac{x_d}{R}\right)\right) \\ &= E\left(\max\left(\left(1 - \frac{x_1 + \cdots + x_d}{R}\right)^{d-1}, 0\right)\right) \\ &= \psi(x_1 + \cdots + x_d). \end{aligned}$$

The marginal survival functions $P(X_i > x) = \psi(x)$ are continuous and strictly decreasing on $\{x\colon \psi(x) > 0\}$ and it may be verified that, for any $\boldsymbol{x} \in \mathbb{R}^d_+$, we have $\psi(x_1 + \cdots + x_d) = \psi(\psi^{-1}(\psi(x_1)) + \cdots + \psi^{-1}(\psi(x_d)))$, so we can write

$$P(X_1 > x_1, \ldots, X_d > x_d) = \psi(\psi^{-1}(\psi(x_1)) + \cdots + \psi^{-1}(\psi(x_d))),$$

which proves the assertion; for more technical details see McNeil and Nešlehová (2009). □

15.2.2 Non-exchangeable Archimedean Copulas

A copula obtained from construction (7.46) is obviously an *exchangeable* copula conforming to (7.20). While exchangeable bivariate Archimedean copulas are widely used in modelling applications, their exchangeable multivariate extensions represent a very specialized form of dependence structure and have more limited applications. An exception to this is in the area of credit risk, although even here more general models with group structures are also needed. It is certainly natural to enquire whether there are extensions to the Archimedean class that are not rigidly exchangeable, and we devote this section to a short discussion of some possible extensions.

Asymmetric Archimedean copulas. Let C_θ be any exchangeable d-dimensional copula. A parametric family of asymmetric copulas $C_{\theta,\alpha_1,\ldots,\alpha_d}$ is then obtained by setting

$$C_{\theta,\alpha_1,\ldots,\alpha_d}(u_1, \ldots, u_d) = C_\theta(u_1^{\alpha_1}, \ldots, u_d^{\alpha_d}) \prod_{i=1}^{d} u_i^{1-\alpha_i}, \quad (u_1, \ldots, u_d) \in \mathbb{R}^d, \tag{15.10}$$

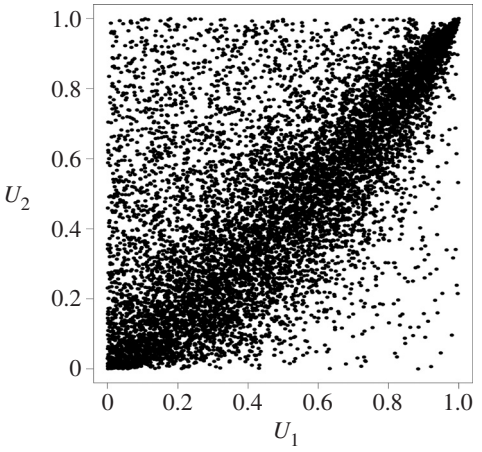

Figure 15.4. Pairwise scatterplots of 10 000 simulated points from an asymmetric Gumbel copula $C^{\mathrm{Gu}}_{4,0.95,0.7}$ based on construction (15.10). This is simulated using Algorithm 15.5.

where $0 \leqslant \alpha_i \leqslant 1$ for all i. Only in the special case $\alpha_1 = \cdots = \alpha_d$ is the copula (15.10) exchangeable. Note also that when the α_i parameters are 0, $C_{\theta,0,\ldots,0}$ is the independence copula, and when the α_i parameters are 1, $C_{\theta,1,\ldots,1}$ is simply C_θ. When C_θ is an Archimedean copula, we refer to copulas constructed by (15.10) as asymmetric Archimedean copulas.

We check that (15.10) defines a copula by constructing a random vector with this df and observing that its margins are standard uniform. The construction is presented as a simulation algorithm.

Algorithm 15.5 (asymmetric Archimedean copula).

(1) Generate a random vector $(V_1, \ldots, V_d)$ with df C_θ.

(2) Generate, independently of $(V_1, \ldots, V_d)$, independent standard uniform variates $\tilde{U}_i$ for $i = 1, \ldots, d$.

(3) Return $U_i = \max(V_i^{1/\alpha_i}, \tilde{U}_i^{1/(1-\alpha_i)})$ for $i = 1, \ldots, d$.

It may be easily verified that $(U_1, \ldots, U_d)$ have the df (15.10). See Figure 15.4 for an example of simulated data from an asymmetric bivariate copula based on Gumbel's copula. Note that an alternative copula may be constructed by taking $(\tilde{U}_1, \ldots, \tilde{U}_d)$ in Algorithm 15.5 to be distributed according to some copula other than the independence copula.

Nested Archimedean copulas. Non-exchangeable, higher-dimensional Archimedean copulas with exchangeable bivariate margins can be constructed by recursive application (or nesting) of Archimedean generators and their inverses, and we will give examples in this section. The biggest problem with these constructions lies in checking that they lead to valid multivariate distributions satisfying (7.1). The necessary theory is complicated and we will simply indicate the nature of the conditions that are necessary without providing justification; a comprehensive reference is Joe (1997). It turns out that with some care we can construct situations of *partial*

exchangeability. We give three- and four-dimensional examples that indicate the pattern of construction.

Example 15.6 (three-dimensional non-exchangeable Archimedean copulas). Suppose that ϕ_1 and ϕ_2 are two Archimedean copula generators and consider

$$C(u_1, u_2, u_3) = \psi_2(\psi_2^{-1} \circ \psi_1(\psi_1^{-1}(u_1) + \psi_1^{-1}(u_2)) + \psi_2^{-1}(u_3)). \tag{15.11}$$

Conditions that ensure that this is a copula are that the generators ψ_1 and ψ_2 are completely monotonic functions, as in (7.47), and that the composition $\psi_2^{-1} \circ \psi_1 \colon [0, \infty) \to [0, \infty)$ is a function whose derivative is a completely monotonic function.

Observe that when $\psi_2 = \psi_1 = \psi$ we are back in the situation of full exchangeability, as in (7.46). Otherwise, if $\psi_1 \neq \psi_2$ and (U_1, U_2, U_3) is a random vector with df given by (15.11), then only U_1 and U_2 are exchangeable, i.e. $(U_1, U_2, U_3) \stackrel{\mathrm{d}}{=} (U_2, U_1, U_3)$, but no other swapping of subscripts is possible. All bivariate margins of (15.11) are themselves Archimedean copulas. The margins C_{13} and C_{23} have generator ψ_2 and C_{12} has generator ψ_1.

Example 15.7 (four-dimensional non-exchangeable Archimedean copulas). A possible four-dimensional construction is

$$\begin{aligned} C(u_1, u_2, u_3, u_4) = \psi_3(\psi_3^{-1} \circ \psi_1(\psi_1^{-1}(u_1) + \psi_1^{-1}(u_2)) \\ + \psi_3^{-1} \circ \psi_2(\psi_2^{-1}(u_3) + \psi_2^{-1}(u_4))), \end{aligned} \tag{15.12}$$

where ψ_1, ψ_2 and ψ_3 are three distinct, completely monotonic Archimedean copula generators and we assume that the composite functions $\psi_3^{-1} \circ \psi_1$ and $\psi_3^{-1} \circ \psi_2$ have derivatives that are completely monotonic to obtain a proper distribution function. This is not the only possible four-dimensional construction (Joe 1997), but it is a useful construction because it gives two exchangeable groups. If (U_1, U_2, U_3, U_4) has the df (15.12), then U_1 and U_2 are exchangeable, as are U_3 and U_4.

The same kinds of construction can be extended to higher dimensions, subject again to complete monotonicity conditions on the compositions of generator inverses and generators (see Notes and Comments).

LT-Archimedean copulas with p-factor structure. Recall from Definition 7.52 the family of LT-Archimedean copulas. The arguments of Proposition 7.51 establish that these have the form

$$\begin{aligned} C(u_1, \ldots, u_d) &= \hat{G}(\hat{G}^{-1}(u_1) + \cdots + \hat{G}^{-1}(u_d)) \\ &= E\left(\exp\left(-V \sum_{i=1}^{d} \hat{G}^{-1}(u_i)\right)\right) \end{aligned} \tag{15.13}$$

for a strictly positive rv V with Laplace–Stieltjes transform $\hat{G}$.

It is possible to generalize the construction (15.13) to obtain a larger family of non-exchangeable copulas. An LT-Archimedean copula with p-factor structure is

constructed from a p-dimensional random vector $\boldsymbol{V} = (V_1, \ldots, V_p)'$ with independent, strictly positive components and a matrix $A \in \mathbb{R}^{d\times p}$ with elements $a_{ij} > 0$ as follows:

$$C(u_1, \ldots, u_d) = E\left(\exp\left(-\sum_{i=1}^{d} \boldsymbol{a}_i' \boldsymbol{V} \hat{G}_i^{-1}(u_i)\right)\right), \tag{15.14}$$

where $\boldsymbol{a}_i$ is the ith row of A and $\hat{G}_i^{-1}$ is the Laplace–Stieltjes transform of the strictly positive rv $\boldsymbol{a}_i'\boldsymbol{V}$.

We can write (15.14) in a different way, which facilitates the computation of $C(u_1, \ldots, u_d)$. Note that

$$\sum_{i=1}^{d} \boldsymbol{a}_i' \boldsymbol{V} \hat{G}_i^{-1}(u_i) = \sum_{j=1}^{p} V_j \sum_{i=1}^{d} a_{ij} \hat{G}_i^{-1}(u_i).$$

It follows from the independence of the V_j that

$$\begin{aligned} C(u_1, \ldots, u_d) &= \prod_{j=1}^{p} E\left(\exp\left(-V_j \sum_{i=1}^{d} a_{ij} \hat{G}_i^{-1}(u_i)\right)\right) \\ &= \prod_{j=1}^{p} \hat{G}_{V_j}\left(\sum_{i=1}^{d} a_{ij} \hat{G}_i^{-1}(u_i)\right). \end{aligned} \tag{15.15}$$

Note that (15.15) is fairly easy to evaluate when $\hat{G}_{V_j}$, the Laplace–Stieltjes transform of the V_j, is available in closed form, because $\hat{G}_i(t) = \prod_{j=1}^{p} \hat{G}_{V_j}(a_{ij}t)$ by the independence of the V_j.

Notes and Comments

The relationship between d-monotonic functions and Archimedean copulas in dimension d, as well as the link to simplex distributions, is developed in McNeil and Nešlehová (2009); see also McNeil and Nešlehová (2010), which provides many examples of such copulas and generalizes the theory to so-called Liouville copulas, the survival copulas of Liouville distributions.

For more details on the asymmetric copulas obtained from construction (15.10) and ideas for more general asymmetric copulas see Genest, Ghoudi and Rivest (1998). These copula classes were introduced in the PhD thesis of Khoudraji (1995). For additional theory concerning nested higher-dimensional Archimedean copulas and sampling algorithms see Joe (1997), McNeil (2008), Hofert (2008, 2011, 2012) and Hofert and Mächler (2011). LT-Archimedean copulas with p-factor structure were proposed by Rogge and Schönbucher (2003) with applications in credit risk in mind. Krupskii and Joe (2013) extend the idea to suggest even more general ways of constructing factor copula models.

16
Advanced Topics in Extreme Value Theory

This chapter extends the treatment of EVT in Chapter 5. In Section 16.1 we provide more information about the tails of some of the distributions and models that are prominent in this book, including the tails of normal variance mixture models (Chapter 6) and strictly stationary GARCH models (Chapter 4).

In Section 16.2 we build on the point process framework for describing the occurrence and magnitude of threshold exceedances (the POT model) that was described in Section 5.3. We show how self-exciting processes for extremes may be developed based on the idea of Hawkes processes.

Sections 16.3 and 16.4 provide a concise summary of the more important ideas in multivariate EVT; they deal, respectively, with multivariate maxima and multivariate threshold exceedances. The novelty of these sections is that the ideas are presented as far as possible using the copula methodology of Chapter 7. The style is similar to Sections 5.1 and 5.2, with the main results being mostly stated without proof and an emphasis being given to examples relevant for applications.

16.1 Tails of Specific Models

In this section we survey the tails of some of the more important distributions and models that we have encountered in this book.

16.1.1 Domain of Attraction of the Fréchet Distribution

As stated in Section 5.1.2, the domain of attraction of the Fréchet distribution consists of distributions with regularly varying tails of the form $\bar{F}(x) = x^{-\alpha}L(x)$ for $\alpha > 0$, where α is known as the tail index. These are heavy-tailed models where higher-order moments cease to exist. Normalized maxima of random samples from such distributions converge to a Fréchet distribution with shape parameter $\xi = 1/\alpha$, and excesses over sufficiently high thresholds converge to a generalized Pareto distribution with shape parameter $\xi = 1/\alpha$.

We now show that the Student t distribution and the inverse gamma distribution are in this class; we analyse the former because of its general importance in financial modelling and the latter because it appears as the mixing distribution that yields the Student t in the class of normal variance mixture models (see Example 6.7). In Section 16.1.3 we will see that the mixing distribution in a normal variance mixture model essentially determines the tail of that model.

Both the t and inverse gamma distributions are presented in terms of their density, and the analysis of their tails proves to be a simple application of a useful result known as Karamata's Theorem, which is given in Section A.1.4.

Example 16.1 (Student t distribution). It is easily verified that the standard univariate t distribution with $\nu \geqslant 1$ has a density of the form $f_\nu(x) = x^{-(\nu+1)}L(x)$ where L is a slowly varying function. Karamata's Theorem (see Theorem A.7) therefore allows us to calculate the form of the tail $\bar{F}_\nu(x) = \int_x^\infty f_\nu(y)\,\mathrm{d}y$ by essentially treating the slowly varying function as a constant and taking it out of the integral. We get

$$\bar{F}_\nu(x) = \int_x^\infty y^{-(\nu+1)}L(y)\,\mathrm{d}y \sim \nu^{-1}x^{-\nu}L(x), \quad x \to \infty,$$

from which we conclude that the df F_ν of a t distribution has tail index ν and $F_\nu \in \mathrm{MDA}(H_{1/\nu})$ by Theorem 5.8.

Example 16.2 (inverse gamma distribution). The density of the inverse gamma distribution is given in (A.17). It is of the form $f_{\alpha,\beta}(x) = x^{-(\alpha+1)}L(x)$, since $\mathrm{e}^{-\beta/x} \to 1$ as $x \to \infty$. Using the same technique as in Example 16.1, we deduce that this distribution has tail index α, so $F_{\alpha,\beta} \in \mathrm{MDA}(H_{1/\alpha})$.

16.1.2 Domain of Attraction of the Gumbel Distribution

The Gumbel class consists of the so-called *von Mises* distribution functions and any other distributions that are *tail equivalent* to von Mises distributions (see Embrechts, Klüppelberg and Mikosch 1997, pp. 138–150). We give the definitions of both of these concepts below. Note that distributions in this class can have both infinite and finite right endpoints; again we write $x_F = \sup\{x \in \mathbb{R}\colon F(x) < 1\} \leqslant \infty$ for the right endpoint of F.

Definition 16.3 (von Mises distribution function). Suppose there exists some $z < x_F$ such that F has the representation

$$\bar{F}(x) = c\exp\left\{-\int_z^x \frac{1}{a(t)}\,\mathrm{d}t\right\}, \quad z < x < x_F,$$

where c is some positive constant, $a(t)$ is a positive and absolutely continuous function with derivative a', and $\lim_{x\uparrow x_F} a'(x) = 0$. Then F is called a von Mises distribution function.

Definition 16.4 (tail equivalence). Two dfs F and G are called *tail equivalent* if they have the same right endpoints (i.e. $x_F = x_G$) and $\lim_{x\uparrow x_F} \bar{F}(x)/\bar{G}(x) = c$ for some constant $0 < c < \infty$.

To decide whether a particular df F is a von Mises df, the following condition is extremely useful. Assume there exists some $z < x_F$ such that F is twice differentiable on (z, x_F) with density $f = F'$ and $F'' < 0$ in (z, x_F). Then F is a von Mises df if and only if

$$\lim_{x\uparrow x_F} \frac{\bar{F}(x)F''(x)}{f^2(x)} = -1. \tag{16.1}$$

We now use this condition to show that the gamma df is a von Mises df.

Example 16.5 (gamma distribution). The density $f = f_{\alpha,\beta}$ of the gamma distribution is given in (A.13), and a straightforward calculation yields $F''(x) = f'(x) = -f(x)(\beta + (1-\alpha)/x) < 0$, provided $x > \max((\alpha - 1)/\beta, 0)$. Clearly, $\lim_{x\to\infty} F''(x)/f(x) = -\beta$. Moreover, using L'Hôpital's rule we get $\lim_{x\to\infty} \bar{F}(x)/f(x) = \lim_{x\to\infty} -f(x)/f'(x) = \beta^{-1}$. Combining these two limits establishes (16.1).

Example 16.6 (GIG distribution). The density of an rv $X \sim N^-(\lambda, \chi, \psi)$ with the GIG distribution is given in (A.14). Let $F_{\lambda,\chi,\psi}(x)$ denote the df and consider the case where $\psi > 0$. If $\psi = 0$, then the GIG is an inverse gamma distribution, which was shown in Example 16.2 to be in the Fréchet class. If $\psi > 0$, then $\lambda \geqslant 0$, and a similar technique to Example 16.5 could be used to establish that the GIG is a von Mises df. In the case where $\lambda > 0$ it is easier to demonstrate tail equivalence with a gamma distribution, which is the special case when $\chi = 0$. We observe that

$$\lim_{x\to\infty} \frac{\bar{F}_{\lambda,\chi,\psi}(x)}{\bar{F}_{\lambda,0,\psi}(x)} = \lim_{x\to\infty} \frac{f_{\lambda,\chi,\psi}(x)}{f_{\lambda,0,\psi}(x)} = c_{\lambda,\chi,\psi}$$

for some constant $0 < c_{\lambda,\chi,\psi} < \infty$. It follows that $F_{\lambda,\chi,\psi} \in \mathrm{MDA}(H_0)$.

16.1.3 Mixture Models

In this book we have considered a number of models for financial risk-factor changes that arise as mixtures (or products) of rvs. In Chapter 6 we introduced multivariate normal variance mixture models including the Student t and (symmetric) GH distributions, which have the general structure given in (6.18). A one-dimensional normal variance mixture (or the marginal distribution of a d-dimensional normal variance mixture) is of the same type (see Section A.1.1) as an rv X satisfying

$$X \overset{\mathrm{d}}{=} \sqrt{W}Z, \tag{16.2}$$

where $Z \sim N(0, 1)$ and W is an independent, positive-valued scalar rv. We would like to know more about the tails of distributions satisfying (16.2).

More generally, to understand the tails of the marginal distributions of elliptical distributions it suffices to consider spherical distributions, which have the stochastic representation

$$\boldsymbol{X} \overset{\mathrm{d}}{=} R\boldsymbol{S} \tag{16.3}$$

for a random vector $\boldsymbol{S}$ that is uniformly distributed on the unit sphere $\mathcal{S}^{d-1} = \{\boldsymbol{s} \in \mathbb{R}^d : \boldsymbol{s}'\boldsymbol{s} = 1\}$, and for an independent radial variate R (see Section 6.3.1, and Theorem 6.21 in particular). Again we would like to know more about the tails of the marginal distributions of the vector $\boldsymbol{X}$ in (16.3).

In Section 4.2 we considered strictly stationary stochastic processes (X_t), such as GARCH processes satisfying equations of the form

$$X_t = \sigma_t Z_t, \tag{16.4}$$

where (Z_t) are strict white noise innovations, typically with a Gaussian distribution or (more realistically) a scaled Student t distribution, and where σ_t is an

$\mathcal{F}_{t-1}$-measurable rv representing volatility. These models can also be seen as mixture models and we would like to know something about the tail of the stationary distribution of (X_t).

A useful result for analysing the tails of mixtures is the following theorem due to Breiman (1965), which we immediately apply to spherical distributions.

Theorem 16.7 (tails of mixture distributions). *Let X be given by $X = YZ$ for independent, non-negative rvs Y and Z such that*

(1) *Y has a regularly varying tail with tail index α;*

(2) *$E(Z^{\alpha+\varepsilon}) < \infty$ for some $\varepsilon > 0$.*

Then X has a regularly varying tail with tail index α and

$$P(X > x) \sim E(Z^{\alpha})P(Y > x), \quad x \to \infty.$$

Proposition 16.8 (tails of spherical distributions). *Let $\boldsymbol{X} \stackrel{\mathrm{d}}{=} R\boldsymbol{S} \sim S_d(\psi)$ have a spherical distribution. If R has a regularly varying tail with tail index α, then so does $|X_i|$ for $i = 1, \ldots, d$. If $E(R^k) < \infty$ for all $k > 0$, then $|X_i|$ does not have a regularly varying tail.*

Proof. Suppose that R has a regularly varying tail with tail index α and consider RS_i. Since $|S_i|$ is a non-negative rv with finite support $[0, 1]$ and finite moments, it follows from Theorem 16.7 that $R|S_i|$, and hence $|X_i|$, are regularly varying with tail index α. If $E(R^k) < \infty$ for all $k > 0$, then $E|X_i|^k < \infty$ for all $k > 0$, so that $|X_i|$ cannot have a regularly varying tail. □

Example 16.9 (tails of normal variance mixtures). Suppose that $\boldsymbol{X} \stackrel{\mathrm{d}}{=} \sqrt{W}\boldsymbol{Z}$ with $\boldsymbol{Z} \sim N_d(\mathbf{0}, I_d)$ and W an independent scalar rv, so that both $\boldsymbol{Z}$ and $\boldsymbol{X}$ have spherical distributions and $\boldsymbol{X}$ has a normal variance mixture distribution. The vector $\boldsymbol{Z}$ has the spherical representation $\boldsymbol{Z} \stackrel{\mathrm{d}}{=} \tilde{R}\boldsymbol{S}$, where $\tilde{R}^2 \sim \chi^2_d$ (see Example 6.23). The vector $\boldsymbol{X}$ has the spherical representation $\boldsymbol{X} \stackrel{\mathrm{d}}{=} R\boldsymbol{S}$, where $R \stackrel{\mathrm{d}}{=} \sqrt{W}\tilde{R}$.

Now, the chi-squared distribution (being a gamma distribution) is in the domain of attraction of the Gumbel distribution, so $E(\tilde{R}^k) = E((\tilde{R}^2)^{k/2}) < \infty$ for all $k > 0$. We first consider the case when W has a regularly varying tail with tail index α so that $\bar{F}_W(w) = L(w)w^{-\alpha}$. It follows that $P(\sqrt{W} > x) = P(W > x^2) = L_2(x)x^{-2\alpha}$, where $L_2(x) := L(x^2)$ is also slowly varying, so that $\sqrt{W}$ has a regularly varying tail with tail index 2α. By Theorem 16.7, $R \stackrel{\mathrm{d}}{=} \sqrt{W}\tilde{R}$ also has a regularly varying tail with tail index 2α, and by Proposition 16.8, so do the components of $|\boldsymbol{X}|$.

To consider a particular case, suppose that $W \sim \mathrm{Ig}(\frac{1}{2}\nu, \frac{1}{2}\nu)$, so that, by Example 16.2, W is regularly varying with tail index $\frac{1}{2}\nu$. Then $\sqrt{W}$ has a regularly varying tail with tail index ν and so does $|X_i|$; this is hardly surprising because $\boldsymbol{X} \sim t_d(\nu, \mathbf{0}, I_d)$, implying that X_i has a univariate Student t distribution with ν degrees of freedom, and we already know from Example 16.1 that this has tail index ν.

On the other hand, if $F_W \in \mathrm{MDA}(H_0)$, then $E(R^k) < \infty$ for all $k > 0$ and $|X_i|$ cannot have a regularly varying tail by Proposition 16.8. This means, for example,

Table 16.1. Approximate theoretical values of the tail index κ solving (16.5) for various GARCH(1, 1) processes with Gaussian and Student t innovation distributions.

Parameters	Gauss	t distribution $\nu = 8$	t distribution $\nu = 4$
$\alpha_1 = 0.2,\ \beta = 0.75$	4.4	3.5	2.7
$\alpha_1 = 0.1,\ \beta = 0.85$	9.1	5.8	3.4
$\alpha_1 = 0.05,\ \beta = 0.95$	21.1	7.9	3.9

that univariate GH distributions do not have power tails (except for the special boundary case corresponding to Student t) because the GIG is in the maximum domain of attraction of the Gumbel distribution, as was shown in Example 16.6.

Analysis of the tails of the stationary distribution of GARCH-type models is more challenging. In view of Theorem 16.7 and the foregoing examples, it is clear that when the innovations (Z_t) are Gaussian, then the law of the process (X_t) in (16.4) will have a regularly varying tail if the volatility σ_t has a regularly varying tail. Mikosch and Stărică (2000) analyse the GARCH(1, 1) model (see Definition 4.20), where the squared volatility satisfies $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta\sigma_{t-1}^2$. They show that under relatively weak conditions on the innovation distribution of (Z_t), the volatility σ_t has a regularly varying tail with tail index κ given by the solution of the equation

$$E((\alpha_1 Z_t^2 + \beta)^{\kappa/2}) = 1. \tag{16.5}$$

In Table 16.1 we have calculated approximate values of κ for various innovation distributions and parameter values using numerical integration and root-finding procedures. By Theorem 16.7 these are the values of the tail index for the stationary distribution of the GARCH(1, 1) model itself.

Two main findings are obvious: for any fixed set of parameter values, the tail index gets smaller and the tail of the GARCH model gets heavier as we move to heavier-tailed innovation distributions; for any fixed innovation distribution, the tail of the GARCH model gets lighter as we decrease the ARCH effect (α_1) and increase the GARCH effect (β).

Tail dependence in elliptical distributions. We close this section by giving a result that reveals an interesting connection between tail dependence in elliptical distributions and regular variation of the radial rv R in the representation $\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + R\boldsymbol{A}\boldsymbol{S}$ of an elliptically symmetric distribution given in Proposition 6.27.

Theorem 16.10. *Let $\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + R\boldsymbol{A}\boldsymbol{S} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$, where $\boldsymbol{\mu}$, R, A and $\boldsymbol{S}$ are as in Proposition 6.27 and we assume that $\sigma_{ii} > 0$ for all $i = 1, \ldots, d$. If R has a regularly varying tail with tail index $\alpha > 0$, then the coefficient of upper and lower tail dependence between X_i and X_j is given by*

$$\lambda(X_i, X_j) = \frac{\int_{\pi/2 - \arcsin \rho_{ij}}^{\pi/2} \cos^\alpha(t)\, \mathrm{d}t}{\int_0^{\pi/2} \cos^\alpha(t)\, \mathrm{d}t}, \tag{16.6}$$

where ρ_{ij} is the (i, j)th element of $P = \wp(\Sigma)$ and $\wp$ is the correlation operator defined in (6.5).

An example where R has a regularly varying tail occurs in the case of the multivariate t distribution $\boldsymbol{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$. It is obvious from the arguments used in Example 16.9 that the tail of the df of R is regularly varying with tail index $\alpha = \nu$. Thus (16.6) with α replaced by ν gives an alternative expression to (7.38) for calculating tail-dependence coefficients for the t copula $C^t_{\nu,P}$.

Arguably, the original expression (7.38) is easier to work with, since the df of a univariate t distribution is available in statistical software packages. Moreover, the equivalence of the two formulas allows us to conclude that we can use (7.38) to evaluate tail-dependence coefficients for any bivariate elliptical distribution with correlation parameter ρ when the distribution of the radial rv R has a regularly varying tail with tail index ν.

Notes and Comments

Section 16.1 is a highly selective account tailored to the study of a number of very specific models, and all of the theoretical subjects touched upon—regular variation, von Mises distributions, tails of products, tails of stochastic recurrence equations—can be studied in much greater detail.

For more about regular variation, slow variation and Karamata's Theorem see Bingham, Goldie and Teugels (1987) and Seneta (1976). A summary of the more important ideas with regard to the study of extremes is found in Resnick (2008). Section 16.1.2, with the exception of the examples, is taken from Embrechts, Klüppelberg and Mikosch (1997), and detailed references to results on von Mises distributions and the maximum domain of attraction of the Gumbel distribution are found therein.

Theorem 16.7 follows from results of Breiman (1965). Related results on distributions of products are found in Embrechts and Goldie (1980). The discussion of tails of GARCH models is based on Mikosch and Stărică (2000); the theory involves the study of stochastic recurrence relations and is essentially due to Kesten (1973). See also Mikosch (2003) for an excellent introduction to these ideas.

The formula for tail-dependence coefficients in elliptical distributions when the radial rv has a regularly varying tail is taken from Hult and Lindskog (2002). Similar results were derived independently by Schmidt (2002); see also Frahm, Junker and Szimayer (2003) for a discussion of the applicability of such results to financial returns.

16.2 Self-exciting Models for Extremes

The models described in this section build on the point process formulation of the POT model for threshold exceedances in Section 5.3. However, instead of assuming Poisson behaviour as we did before, we now attempt to explain the clustering of extreme events in financial return data in terms of self-exciting behaviour.

16.2.1 Self-exciting Processes

We first consider modelling the occurrence of threshold exceedances in time using a simple form of self-exciting point process known as a Hawkes process. The basic idea is that instead of modelling the instantaneous risk of a threshold exceedance as being constant over time, we assume that threshold exceedances in the recent past cause the instantaneous risk of a threshold exceedance to be higher. The main area of application of these models has traditionally been in the modelling of earthquakes and their aftershocks, although there is a growing number of applications in financial modelling (see Notes and Comments).

Given data $X_1, \ldots, X_n$ and a threshold u, we will assume as usual that there are N_u exceedances, comprising the data $\{(i, X_i) : 1 \leqslant i \leqslant n,\ X_i > u\}$. Note that from now on we will express the time of an exceedance on the natural timescale of the time series, so if, for example, the data are daily observations, then our times are expressed in days. It will also be useful to have the alternative notation $\{(T_j, \tilde{X}_j) : j = 1, \ldots, N_u\}$, which enumerates exceedances consecutively.

First we consider a model for exceedance times only. In point process notation we let $Y_i = i I_{\{X_i > u\}}$, so Y_i returns an exceedance time, in the event that one takes place at time i, and returns 0 otherwise. The point process of exceedances is the process $N(\cdot)$ with state space $\mathcal{X} = (0, n]$ given by $N(A) = \sum_{i=1}^{n} I_{\{Y_i \in A\}}$ for $A \subset \mathcal{X}$.

We assume that the point process $N(\cdot)$ is a self-exciting process with *conditional intensity*

$$\lambda^*(t) = \tau + \psi \sum_{j\,:\,0<T_j<t} h(t - T_j, \tilde{X}_j - u), \tag{16.7}$$

where $\tau > 0$, $\psi \geqslant 0$ and h is some positive-valued function. Each previous exceedance $(T_j, \tilde{X}_j)$ contributes to the conditional intensity and the amount that it contributes can depend on both the elapsed time $(t - T_j)$ since that exceedance and the amount of the excess loss $(\tilde{X}_j - u)$ over the threshold. Informally, we understand the conditional intensity as expressing the instantaneous chance of a new exceedance of the threshold at time t, like the rate or intensity of an ordinary Poisson process. However, in the self-exciting model, the conditional intensity is itself a stochastic process that depends on ω, the state of nature, through the history of threshold exceedances up to (but not including) time t.

Possible parametric specifications of the h function are

- $h(s, x) = \mathrm{e}^{\delta x - \gamma s}$, where $\delta, \gamma > 0$; or
- $h(s, x) = \mathrm{e}^{\delta x}(s + \gamma)^{-(\rho+1)}$, where $\delta, \gamma, \rho > 0$.

Collecting all parameters in $\boldsymbol{\theta}$, the likelihood takes the form

$$L(\boldsymbol{\theta}; \text{data}) = \exp\left(-\int_0^n \lambda^*(s)\,\mathrm{d}s\right) \prod_{i=1}^{N_u} \lambda^*(T_i),$$

and may be maximized numerically to obtain parameter estimates.

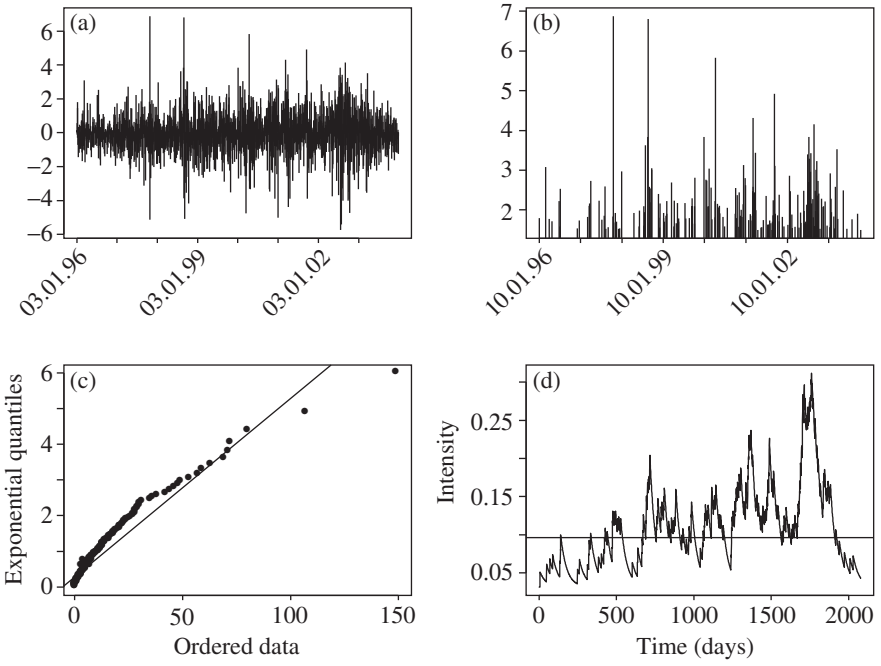

Figure 16.1. (a) S&P daily percentage loss data. (b) Two hundred largest losses. (c) A Q–Q plot of inter-exceedance times against an exponential reference distribution. (d) The estimated intensity of exceeding the threshold in a self-exciting model. See Example 16.11 for details.

Example 16.11 (S&P daily percentage losses 1996–2003). We apply the self-exciting process methodology to all daily percentage losses incurred by the Standard & Poor's index in the eight-year period 1996–2003 (2078 values). In Figure 16.1 the loss data are shown as well as the point process of the 200 largest daily losses exceeding a threshold of 1.5%. Clearly, there is clustering in the pattern of exceedance data, and the Q–Q plot shows that the inter-exceedance times are not exponential.

We fit the simpler self-exciting model with $h(s, x) = \mathrm{e}^{\delta x - \gamma s}$. The parameter estimates (and standard errors) are $\hat{\tau} = 0.032(0.011)$, $\hat{\psi} = 0.016(0.0069)$, $\hat{\gamma} = 0.026(0.011)$, $\hat{\delta} = 0.13(0.27)$, suggesting that all parameters except δ are significant. The log-likelihood for the fitted model is -648.2, whereas the log-likelihood for a homogeneous Poisson model is -668.2; the Poisson special case can therefore clearly be rejected in a likelihood ratio test with a p-value less than 0.001. Figure 16.1 (d) shows the estimated intensity $\lambda^*(t)$ of crossing the threshold throughout the data observation period, which seems to reflect the pattern of exceedances observed.

Note that a simple refinement of this model (and those of the following section) would be to consider a self-exciting structure where both extreme negative and extreme positive returns contributed to the conditional intensity; this would involve setting upper and lower thresholds and considering exceedances of both.

16.2.2 A Self-exciting POT Model

We now consider how the POT model of Section 5.3.2 might be generalized to incorporate a self-exciting component. We first develop a marked self-exciting model where marks have a generalized Pareto distribution but are *unpredictable*, meaning that the excess losses are iid GPD. In the second model we consider the case of *predictable* marks. In this model the excess losses are conditionally generalized Pareto, given the exceedance history up to the time of the mark, with a scaling parameter that depends on that history. In this way we get a model where, in a period of excitement, both the temporal intensity of occurrence and the magnitude of the exceedances increase.

In point process language, our models are processes $N(\cdot)$ on a state space of the form $\mathcal{X} = (0, n] \times (u, \infty)$ such that $N(A) = \sum_{i=1}^{n} I_{\{(i,X_i)\in A\}}$ for sets $A \subset \mathcal{X}$. To build these models we start with the intensity of the reparametrized version of the standard POT model given in (5.31). We recall that this model simply says that exceedances of the threshold u occur as a homogeneous Poisson process with rate τ and that excesses have a generalized Pareto distribution with df $G_{\xi,\beta}$.

Model with unpredictable marks. We first introduce the notation

$$v^*(t) = \sum_{j:\, 0<T_j<t} h(t - T_j, \tilde{X}_j - u)$$

for the *self-excitement function*, where the function h is as in Section 16.2.1. We generalize (5.31) and consider a self-exciting model with conditional intensity

$$\lambda^*(t, x) = \frac{\tau + \psi v^*(t)}{\beta}\left(1 + \xi\frac{x-u}{\beta}\right)^{-1/\xi-1} \tag{16.8}$$

on a state space $\mathcal{X} = (0, n] \times (u, \infty)$, where $\tau > 0$ and $\psi \geqslant 0$. Effectively, we have combined the one-dimensional intensity in (16.7) with a GPD density. When $\psi = 0$ we have an ordinary POT model with no self-exciting structure.

It is easy to calculate that the conditional rate of crossing the threshold $x \geqslant u$ at time t, given information up to that time, is

$$\tau^*(t, x) = \int_x^\infty \lambda^*(t, y)\,\mathrm{d}y = (\tau + \psi v^*(t))\left(1 + \xi\frac{x-u}{\beta}\right)^{-1/\xi}, \tag{16.9}$$

which, for fixed x, is simply a one-dimensional self-exciting process of the form (16.7). The implied distribution of the excess losses when an exceedance takes place is generalized Pareto, because

$$\frac{\tau^*(t, u + x)}{\tau^*(t, u)} = \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi} = \bar{G}_{\xi,\beta}(x), \tag{16.10}$$

independently of t. Statistical fitting of this model is performed by maximizing a likelihood of the form

$$L(\boldsymbol{\theta}; \text{data}) = \exp\left(-n\tau - \psi\int_0^n v^*(s)\,\mathrm{d}s\right)\prod_{j=1}^{N_u}\lambda^*(T_j, \tilde{X}_j). \tag{16.11}$$

A model with predictable marks. A model with predictable marks can be obtained by generalizing (16.8) to get

$$\lambda^*(t,x) = \frac{\tau + \psi v^*(t)}{\beta + \alpha v^*(t)} \left(1 + \xi \frac{x-u}{\beta + \alpha v^*(t)}\right)^{-1/\xi - 1}, \tag{16.12}$$

where $\beta > 0$ and $\alpha \geqslant 0$. For simplicity we have assumed that the GPD scaling is also linear in the self-excitement function $v^*(t)$. The properties of this model follow immediately from the model with unpredictable marks. The conditional crossing rate of the threshold $x \geqslant u$ at time t is as in (16.9) with the parameter β replaced by the time-dependent self-exciting function $\beta + \alpha v^*(t)$. By repeating the calculation in (16.10) we find that the distribution of the excess loss over the threshold, given that an exceedance takes place at time t and given the history of exceedances up to time t, is generalized Pareto with df $G_{\xi,\beta+\alpha v^*(t)}$. The likelihood for fitting the model is again (16.11), where the function $\lambda^*(t,x)$ is now given by (16.12). Note that by comparing a model with $\alpha = 0$ and a model with $\alpha > 0$ we can formally test the hypothesis that the marks are unpredictable using a likelihood ratio test.

Example 16.12 (self-exciting POT model for S&P daily loss data). We continue the analysis of the data of Example 16.11 by fitting self-exciting POT models with both unpredictable and predictable marks to the 200 exceedances of the threshold $u = 1.5\%$. The former is equivalent to fitting a self-exciting model to the exceedance times as in Example 16.11 and then fitting a GPD to the excess losses over the threshold; the estimated intensity of crossing the threshold is therefore identical to the one shown in Figure 16.1. The log-likelihood for this model is -783.4, whereas a model with predictable marks gives a value of -779.3 for one extra parameter α; in a likelihood ratio test the p-value is 0.004, showing a significant improvement.

In Figure 16.2 we show the exceedance data as well as the estimated intensity $\tau^*(t,u)$ of exceeding the threshold in the model with predictable marks. We also show the estimated mean of the GPD for the conditional distribution of the excess loss above the threshold, given that an exceedance takes place at time t. The GPD mean $(\beta + \alpha v^*(t))/(1-\xi)$ and the intensity $\tau^*(t,u)$ are both affine functions of the self-excitement function $v^*(t)$ and obviously follow its path.

Calculating conditional risk measures. Finally, we note that self-exciting POT models can be used to estimate a conditional VaR and also a conditional expected shortfall. If we have analysed n daily data ending on day t and want to calculate, say, a 99% VaR, then we treat the problem as a (conditional) return-level problem; we look for the level at which the conditional exceedance intensity at a time point just after t (denoted by $t+$) is 0.01. In general, to calculate a conditional estimate of VaR_α^t (for α sufficiently large) we would attempt to solve the equation $\tau^*(t+,x) = (1-\alpha)$ for some value of x satisfying $x \geqslant u$. In the model with predictable marks this is

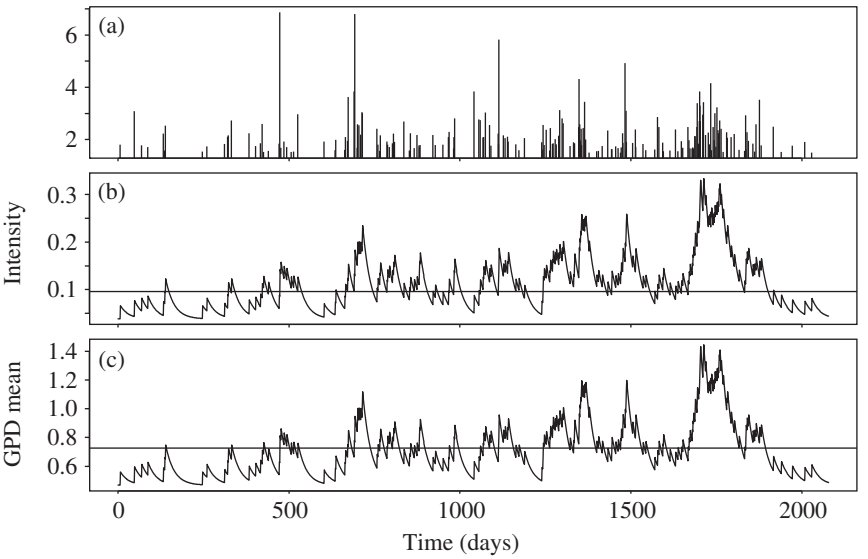

Figure 16.2. (a) Exceedance pattern for 200 largest daily losses in S&P data. (b) Estimated intensity of exceeding the threshold in a self-exciting POT model with predictable marks. (c) Mean of the conditional generalized Pareto distribution of the excess loss above the threshold. See Example 16.12 for details.

possible if $\tau + \psi v^*(t+) > 1 - \alpha$ and gives the formula

$$\mathrm{VaR}_\alpha^t = u + \frac{\beta + \alpha v^*(t+)}{\xi}\left(\left(\frac{1-\alpha}{\tau + \psi v^*(t+)}\right)^{-\xi} - 1\right).$$

The associated conditional expected shortfall could then be calculated by observing that the conditional distribution of excess losses above VaR_α^t given information up to time t is GPD with shape parameter ξ and scaling parameter given by $\beta + \alpha v^*(t+) + \xi(\mathrm{VaR}_\alpha^t - u)$.

Notes and Comments

The original reference to the Hawkes self-exciting process is Hawkes (1971). There is a large literature on the application of such processes to earthquake modelling; a starter reference is Ogata (1988). To our knowledge, the earliest contribution on Hawkes processes in financial econometrics was Bowsher (2002); this paper finally appeared as Bowsher (2007). The application to extremes in financial time series was suggested in Chavez-Demoulin, Davison and McNeil (2005). Embrechts, Liniger and Lin (2011) give an application of multivariate Hawkes processes to multiple financial time series. Self-exciting processes have also been applied in credit risk (Errais, Giesecke and Goldberg 2010) and in the modelling of high-frequency financial data (Chavez-Demoulin and McGill 2012).

The idea detailed in Section 16.2.2 (of a POT model with self-exciting structure) was first proposed in the first edition of this textbook. Grothe, Korniichuk and Manner (2014) extend the idea and consider multivariate models for extremes where component processes may excite themselves or other component processes.

16.3 Multivariate Maxima

In this section we give a brief overview of the theory of multivariate maxima, stating the main results in terms of copulas. A class of copulas known as extreme value copulas emerges as the class of natural limiting dependence structures for multivariate maxima. These provide useful dependence structures for modelling the joint tail behaviour of risk factors that appear to show tail dependence. A useful reference is Galambos (1987), which is one of the few texts to treat the theory of multivariate maxima as a copula theory (although Galambos does not use the word, referring to copulas simply as dependence functions).

16.3.1 Multivariate Extreme Value Copulas

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be iid random vectors in $\mathbb{R}^d$ with joint df F and marginal dfs $F_1, \ldots, F_d$. We label the components of these vectors $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})'$ and interpret them as losses of d different types. We define the maximum of the jth component to be $M_{n,j} = \max(X_{1,j}, \ldots, X_{n,j})$, $j = 1, \ldots, d$. In classical multivariate EVT, the object of interest is the vector of *componentwise block maxima*: $\boldsymbol{M}_n = (M_{n,1}, \ldots, M_{n,d})'$. In particular, we are interested in the possible multivariate limiting distributions for $\boldsymbol{M}_n$ under appropriate normalizations, much as in the univariate case. It should, however, be observed that the vector $\boldsymbol{M}_n$ will in general not correspond to any of the vector observations $\boldsymbol{X}_i$.

We seek limit laws for

$$\frac{\boldsymbol{M}_n - \boldsymbol{d}_n}{\boldsymbol{c}_n} = \left(\frac{M_{n,1} - d_{n,1}}{c_{n,1}}, \ldots, \frac{M_{n,d} - d_{n,d}}{c_{n,d}} \right)'$$

as $n \to \infty$, where $\boldsymbol{c}_n = (c_{n,1}, \ldots, c_{n,d})'$ and $\boldsymbol{d}_n = (d_{n,1}, \ldots, d_{n,d})'$ are vectors of normalizing constants, the former satisfying $\boldsymbol{c}_n > \boldsymbol{0}$. Note that in this and other statements in this section, arithmetic operations on vectors of equal length are understood as componentwise operations. Supposing that $(\boldsymbol{M}_n - \boldsymbol{d}_n)/\boldsymbol{c}_n$ converges in distribution to a random vector with joint df H, we have

$$\lim_{n\to\infty} P\left(\frac{\boldsymbol{M}_n - \boldsymbol{d}_n}{\boldsymbol{c}_n} \leqslant \boldsymbol{x} \right) = \lim_{n\to\infty} F^n(\boldsymbol{c}_n \boldsymbol{x} + \boldsymbol{d}_n) = H(\boldsymbol{x}). \tag{16.13}$$

Definition 16.13 (multivariate extreme value distribution and domain of attraction). If (16.13) holds for some F and some H, we say that F is in the maximum domain of attraction of H, written $F \in \mathrm{MDA}(H)$, and we refer to H as a multivariate extreme value (MEV) distribution.

The convergence issue for multivariate maxima is already partly solved by the univariate theory. If H has non-degenerate margins, then these must be univariate extreme value distributions of Fréchet, Gumbel or Weibull type. Since these are continuous, Sklar's Theorem tells us that H must have a unique copula. The following theorem asserts that this copula C must have a particular kind of scaling behaviour.

Theorem 16.14 (extreme value copula). *If (16.13) holds for some F and some H with GEV margins, then the unique copula C of H satisfies*

$$C(\boldsymbol{u}^t) = C^t(\boldsymbol{u}), \quad \forall t > 0. \tag{16.14}$$

Any copula with the property (16.14) is known as an *extreme value (EV) copula* and can be the copula of an MEV distribution. The independence and comonotonicity copulas are EV copulas and the Gumbel copula provides an example of a parametric EV copula family. The bivariate version in (7.12) obviously has property (16.14), as does the exchangeable higher-dimensional Gumbel copula based on (7.46) as well as the non-exchangeable versions based on (15.10)–(15.12).

There are a number of mathematical results characterizing MEV distributions and EV copulas. One such result is the following.

Theorem 16.15 (Pickands representation). *The copula C is an EV copula if and only if it has the representation*

$$C(\boldsymbol{u}) = \exp\left\{B\left(\frac{\ln u_1}{\sum_{k=1}^d \ln u_k}, \ldots, \frac{\ln u_d}{\sum_{k=1}^d \ln u_k}\right)\sum_{i=1}^d \ln u_i\right\}, \tag{16.15}$$

where $B(\boldsymbol{w}) = \int_{\mathcal{S}_d} \max(x_1 w_1, \ldots, x_d w_d)\,\mathrm{d}S(\boldsymbol{x})$ and S is a finite measure on the d-dimensional simplex, i.e. the set $\mathcal{S}_d = \{\boldsymbol{x}\colon x_i \geqslant 0,\ i = 1, \ldots, d,\ \sum_{i=1}^d x_i = 1\}$.

The function $B(\boldsymbol{w})$ is sometimes referred to as the dependence function of the EV copula. In the general case, such functions are difficult to visualize and work with, but in the bivariate case they have a simple form that we discuss in more detail.

In the bivariate case we redefine $B(\boldsymbol{w})$ as a function of a scalar argument by setting $A(w) := B((w, 1-w)')$ with $w \in [0, 1]$. It follows from Theorem 16.15 that a bivariate copula is an EV copula if and only if it takes the form

$$C(u_1, u_2) = \exp\left\{(\ln u_1 + \ln u_2)A\left(\frac{\ln u_1}{\ln u_1 + \ln u_2}\right)\right\}, \tag{16.16}$$

where $A(w) = \int_0^1 \max((1-x)w, x(1-w))\,\mathrm{d}S(x)$ for a measure S on $[0, 1]$. It can be inferred that such bivariate dependence functions must satisfy

$$\max(w, 1-w) \leqslant A(w) \leqslant 1, \quad 0 \leqslant w \leqslant 1, \tag{16.17}$$

and that they must, moreover, be convex. Conversely, a differentiable convex function $A(w)$ satisfying (16.17) can be used to construct an EV copula using (16.16).

The upper and lower bounds in (16.17) have intuitive interpretations. If $A(w) = 1$ for all w, then the copula (16.16) is clearly the independence copula, and if $A(w) = \max(w, 1-w)$, then it is the comonotonicity copula. It is also useful to note, and easy to show, that we can extract the dependence function from the EV copula in (16.16) by setting

$$A(w) = -\ln C(\mathrm{e}^{-w}, \mathrm{e}^{-(1-w)}), \quad w \in [0, 1]. \tag{16.18}$$

Example 16.16 (Gumbel copula). We consider the asymmetric version of the bivariate Gumbel copula defined by (7.12) and construction (15.10), i.e. the copula

$$C^{\mathrm{Gu}}_{\theta,\alpha,\beta}(u_1, u_2) = u_1^{1-\alpha} u_2^{1-\beta} \exp\{-((-\alpha \ln u_1)^\theta + (-\beta \ln u_2)^\theta)^{1/\theta}\}.$$

As already remarked, this copula has the scaling property (16.14) and is an EV copula. Using (16.18) we calculate that the dependence function is given by

$$A(w) = (1-\alpha)w + (1-\beta)(1-w) + ((\alpha w)^\theta + (\beta(1-w))^\theta)^{1/\theta}.$$

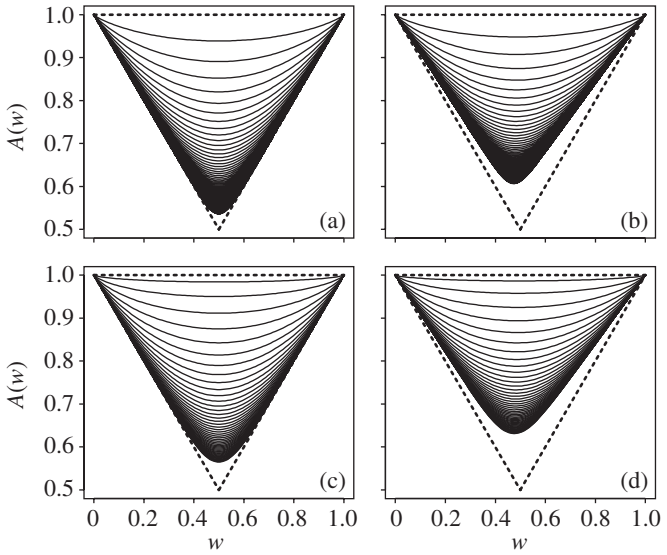

Figure 16.3. Plot of dependence functions for the (a) symmetric Gumbel, (b) asymmetric Gumbel, (c) symmetric Galambos and (d) asymmetric Galambos copulas (asymmetric cases have $\alpha = 0.9$ and $\beta = 0.8$) as described in Examples 16.16 and 16.17. Dashed lines show boundaries of the triangle in which the dependence function must reside; solid lines show dependence functions for a range of parameter values.

We have plotted this function in Figure 16.3 for a range of θ values running from 1.1 to 10 in steps of size 0.1. Part (a) shows the standard symmetric Gumbel copula with $\alpha = \beta = 1$; the dependence function essentially spans the whole range from independence, represented by the upper edge of the dashed triangle, to comonotonicity, represented by the two lower edges of the dashed triangle, which comprise the function $A(w) = \max(w, 1 - w)$. Part (b) shows an asymmetric example with $\alpha = 0.9$ and $\beta = 0.8$; in this case we still have independence when $\theta = 1$, but the limit as $\theta \to \infty$ is no longer the comonotonicity model. The Gumbel copula model is also sometimes known as the logistic model.

Example 16.17 (Galambos copula). This time we begin with the dependence function given by

$$A(w) = 1 - ((\alpha w)^{-\theta} + (\beta(1 - w))^{-\theta})^{-1/\theta}, \tag{16.19}$$

where $0 \leqslant \alpha, \beta \leqslant 1$ and $0 < \theta < \infty$. It can be verified that this is a convex function satisfying $\max(w, 1 - w) \leqslant A(w) \leqslant 1$ for $0 \leqslant w \leqslant 1$, so it can be used to create an EV copula in (16.16). We obtain the copula

$$C^{\mathrm{Gal}}_{\theta,\alpha,\beta}(u_1, u_2) = u_1 u_2 \exp\{((-\alpha \ln u_1)^{-\theta} + (-\beta \ln u_2)^{-\theta})^{-1/\theta}\},$$

which has also been called the negative logistic model. We have plotted this function in Figure 16.3 for a range of θ values running from 0.2 to 5 in steps of size 0.1. Part (c) shows the standard symmetric case with $\alpha = \beta = 1$ spanning the whole range from independence to comonotonicity. Part (d) shows an asymmetric example

with $\alpha = 0.9$ and $\beta = 0.8$; in this case we still approach independence as $\theta \to 0$, but the limit as $\theta \to \infty$ is no longer the comonotonicity model.

A number of other bivariate EV copulas have been described in the literature (see Notes and Comments).

16.3.2 Copulas for Multivariate Minima

The structure of limiting copulas for multivariate minima can be easily inferred from the structure of limiting copulas for multivariate maxima; moving from maxima to minima essentially involves the same considerations that we made at the end of Section 5.1.1 and uses identity (5.2) in particular.

Normalized componentwise minima of iid random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ with df F will converge in distribution to a non-degenerate limit if the df $\tilde{F}$ of the random vectors $-\boldsymbol{X}_1, \ldots, -\boldsymbol{X}_n$ is in the maximum domain of attraction of an MEV distribution (see Definition 16.13), written $\tilde{F} \in \mathrm{MDA}(H)$. Of course, for a radially symmetric distribution, $\tilde{F}$ coincides with F.

Let $\boldsymbol{M}_n^*$ be the vector of componentwise maxima of $-\boldsymbol{X}_1, \ldots, -\boldsymbol{X}_n$ such that $M_{n,j}^* = \max(-X_{1,j}, \ldots, -X_{n,j})$. If $\tilde{F} \in \mathrm{MDA}(H)$ for some non-degenerate H, we have

$$\lim_{n\to\infty} P\left(\frac{\boldsymbol{M}_n^* - \boldsymbol{d}_n}{\boldsymbol{c}_n} \leqslant \boldsymbol{x}\right) = \lim_{n\to\infty} \tilde{F}^n(\boldsymbol{c}_n\boldsymbol{x} + \boldsymbol{d}_n) = H(\boldsymbol{x}) \tag{16.20}$$

for appropriate sequences of normalizing vectors $\boldsymbol{c}_n$ and $\boldsymbol{d}_n$, and an MEV distribution H of the form $H(\boldsymbol{x}) = C(H_{\xi_1}(x_1), \ldots, H_{\xi_d}(x_d))$, where H_{ξ_j} denotes a GEV distribution with shape parameter ξ_j and C is an EV copula satisfying (16.14).

Defining the vector of componentwise minima by $\boldsymbol{m}_n$ and using (5.2), it follows from (16.20) that

$$\lim_{n\to\infty} P\left(\frac{\boldsymbol{m}_n + \boldsymbol{d}_n}{\boldsymbol{c}_n} \geqslant \boldsymbol{x}\right) = H(-\boldsymbol{x}),$$

so normalized minima converge in distribution to a limit with survival function $H(-\boldsymbol{x}) = C(H_{\xi_1}(-x_1), \ldots, H_{\xi_d}(-x_d))$. It follows that the copula of the limiting distribution of the minima is the survival copula of C (see Section 7.1.5 for discussion of survival copulas). In general, the limiting copulas for minima are *survival copulas of EV copulas* and concrete examples of such copulas are the Gumbel and Galambos survival copulas.

In the special case of a radially symmetric underlying distribution, the limiting copula of the minima is precisely the survival copula of the limiting EV copula of the maxima.

16.3.3 Copula Domains of Attraction

As in the case of univariate maxima we would like to know which underlying multivariate dfs F are attracted to which MEV distributions H. We now give a useful result in terms of copulas that is essentially due to Galambos (see Notes and Comments).

Theorem 16.18. *Let $F(\boldsymbol{x}) = C(F_1(x_1), \ldots, F_d(x_d))$ for continuous marginal dfs $F_1, \ldots, F_d$ and some copula C. Let $H(\boldsymbol{x}) = C_0(H_1(x_1), \ldots, H_d(x_d))$ be an MEV distribution with EV copula C_0. Then $F \in \mathrm{MDA}(H)$ if and only if $F_i \in \mathrm{MDA}(H_i)$ for $1 \leqslant i \leqslant d$ and*

$$\lim_{t\to\infty} C^t(u_1^{1/t}, \ldots, u_d^{1/t}) = C_0(u_1, \ldots, u_d), \quad \boldsymbol{u} \in [0,1]^d. \tag{16.21}$$

This result shows that the copula C_0 of the limiting MEV distribution is determined solely by the copula C of the underlying distribution according to (16.21); the marginal distributions of F determine the margins of the MEV limit but are irrelevant to the determination of its dependence structure. This motivates us to introduce the concept of a copula domain of attraction.

Definition 16.19. If (16.21) holds for some C and some EV copula C_0, we say that C is in the copula domain of attraction of C_0, written $C \in \mathrm{CDA}(C_0)$.

There are a number of equivalent ways of writing (16.21). First, by taking logarithms and using the asymptotic identity $\ln(x) \sim x - 1$ as $x \to 1$, we get, for $\boldsymbol{u} \in (0,1]^d$,

$$\left.\begin{aligned} \lim_{t\to\infty} t(1 - C(u_1^{1/t}, \ldots, u_d^{1/t})) &= -\ln C_0(u_1, \ldots, u_d), \\ \lim_{s\to 0^+} \frac{1 - C(u_1^s, \ldots, u_d^s)}{s} &= -\ln C_0(u_1, \ldots, u_d). \end{aligned}\right\} \tag{16.22}$$

By inserting $u_i = \mathrm{e}^{-x_i}$ in the latter identity and using $\mathrm{e}^{-sx} \sim 1 - sx$ as $s \to 0$, we get, for $\boldsymbol{x} \in [0,\infty)^d$,

$$\lim_{s\to 0^+} \frac{1 - C(1 - sx_1, \ldots, 1 - sx_d)}{s} = -\ln C_0(\mathrm{e}^{-x_1}, \ldots, \mathrm{e}^{-x_d}). \tag{16.23}$$

Example 16.20 (limiting copula for bivariate Pareto distribution). In Example 7.14 we saw that the bivariate Pareto distribution has univariate Pareto margins $F_i(x) = 1 - (\kappa_i/(\kappa_i + x))^\alpha$ and a Clayton survival copula. It follows from Example 5.6 that $F_i \in \mathrm{MDA}(H_{1/\alpha})$, $i = 1, 2$. Using (7.16), the Clayton survival copula is calculated to be $C(u_1, u_2) = u_1 + u_2 - 1 + ((1-u_1)^{-1/\alpha} + (1-u_2)^{-1/\alpha} - 1)^{-\alpha}$. Using (16.23), it is easily calculated that $C_0(u_1, u_2) = u_1 u_2 \exp(((-\ln u_1)^{-1/\alpha} + (-\ln u_2)^{-1/\alpha})^{-\alpha})$, which is the standard exchangeable Galambos copula of Example 16.17. The limiting distribution of maxima therefore consists of two Fréchet dfs connected by the Galambos copula.

The coefficients of upper tail dependence play an interesting role in the copula domain of attraction theory. In particular, they can help us to recognize copulas that lie in the copula domain of attraction of the independence copula.

Proposition 16.21. *Let C be a bivariate copula with upper tail-dependence coefficient λ_u, and assume that C satisfies $C \in \mathrm{MDA}(C_0)$ for some EV copula C_0. Then λ_u is also the upper tail-dependence coefficient of C_0 and is related to its dependence function by $\lambda_\mathrm{u} = 2(1 - A(\frac{1}{2}))$.*

Proof. We use (7.35) and (7.16) to see that

$$\lambda_u = \lim_{q\to 1^-} \frac{\hat{C}(1-q,1-q)}{1-q} = 2 - \lim_{q\to 1^-} \frac{1-C(q,q)}{1-q}.$$

By using the asymptotic identity $\ln x \sim x-1$ as $x \to 1$ and the CDA condition (16.22) we can calculate

$$\begin{aligned}
\lim_{q\to 1^-} \frac{1-C_0(q,q)}{1-q} &= \lim_{q\to 1^-} \frac{\ln C_0(q,q)}{\ln q} \\
&= \lim_{q\to 1^-} \lim_{s\to 0^+} \frac{1-C(q^s,q^s)}{-s\ln q} \\
&= \lim_{q\to 1^-} \lim_{s\to 0^+} \frac{1-C(q^s,q^s)}{-\ln(q^s)} \\
&= \lim_{v\to 1^-} \frac{1-C(v,v)}{1-v},
\end{aligned}$$

which shows that C and C_0 share the same coefficient of upper tail dependence. Using the formula $\lambda_u = 2 - \lim_{q\to 1^-} \ln C_0(q,q)/\ln q$ and the representation (16.16) we easily obtain that $\lambda_u = 2(1-A(\frac{1}{2}))$. □

In the case when $\lambda_u = 0$ we must have $A(\frac{1}{2}) = 1$, and the convexity of dependence functions dictates that $A(w)$ is identically 1, so C_0 must be the independence copula. In the higher-dimensional case this is also true: if C is a d-dimensional copula with all upper tail-dependence coefficients equal to 0, then the bivariate margins of the limiting copula C_0 must all be independence copulas, and, in fact, it can be shown that C_0 must therefore be the d-dimensional independence copula (see Notes and Comments).

As an example consider the limiting distribution of multivariate maxima of Gaussian random vectors. Since the pairwise coefficients of tail dependence of Gaussian vectors are 0 (see Example 7.38), the limiting distribution is a product of marginal Gumbel distributions. The convergence is extremely slow, but ultimately normalized componentwise maxima are independent in the limit.

Now consider the multivariate t distribution, which has been an important model throughout this book. If $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are iid random vectors with a $t_d(\nu, \boldsymbol{\mu}, \Sigma)$ distribution, we know from Example 16.1 that univariate maxima of the individual components are attracted to univariate Fréchet distributions with parameter $1/\nu$. Moreover, we know from Example 7.39 that tail dependence coefficients for the t copula are strictly positive; the limiting EV copula cannot be the independence copula.

In fact, the limiting EV copula for t-distributed random vectors can be calculated using (16.23), although the calculations are tedious. In the bivariate case it is found that the limiting copula, which we call the t–EV copula, has dependence function

$$A(w) = wt_{\nu+1}\left(\frac{(w/(1-w))^{1/\nu}-\rho}{\sqrt{(1-\rho^2)/(\nu+1)}}\right) + (1-w)t_{\nu+1}\left(\frac{((1-w)/w)^{1/\nu}-\rho}{\sqrt{(1-\rho^2)/(\nu+1)}}\right), \tag{16.24}$$

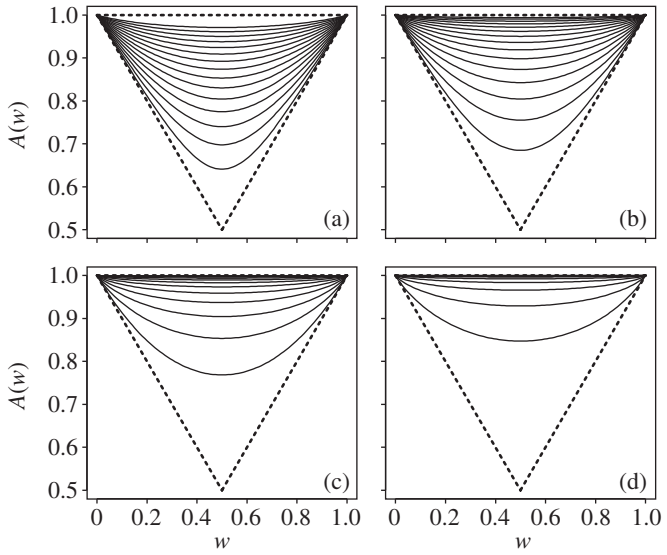

Figure 16.4. Plots of the dependence function for the t–EV copula for (a) $\nu = 2$, (b) $\nu = 4$, (c) $\nu = 10$ and (d) $\nu = 20$, and with various values of ρ.

where ρ is the off-diagonal component of $P = \wp(\Sigma)$. This dependence function is shown in Figure 16.4 for four different values of ν and for ρ values ranging from -0.5 to 0.9 with increments of 0.1. As $\rho \to 1$ the t–EV copula converges to the comonotonicity copula; as $\rho \to -1$ or as $\nu \to \infty$ it converges to the independence copula.

16.3.4 Modelling Multivariate Block Maxima

A multivariate block maxima method analogous to the univariate method of Section 5.1.4 could be developed, although similar criticisms apply, namely that the block maxima method is not the most efficient way of making use of extreme data. Also, the kind of inference that this method allows may not be exactly what is desired in the multivariate case, as will be seen.

Suppose we divide our underlying data into blocks as before and we denote the realizations of the block maxima vectors by $\boldsymbol{M}_{n,1}, \ldots, \boldsymbol{M}_{n,m}$, where m is the total number of blocks. The distributional model suggested by the univariate and multivariate maxima theory consists of GEV margins connected by an extreme value copula.

In the multivariate theory there is, in a sense, a "correct" EV copula to use, which is the copula C_0 to which the copula C of the underlying distribution of the raw data is attracted. However, the underlying copula C is unknown and so the approach is generally to work with any tractable EV copula that appears appropriate for the task in hand. In a bivariate application, if we restrict to exchangeable copulas, then we have at our disposal the Gumbel, Galambos and t–EV copulas, and a number of other possibilities for which references in Notes and Comments should be consulted. As will be apparent from Figures 16.3 and 16.4, the essential functional form of all

these families is really very similar; it is mostly sufficient to work with either the Gumbel copula or the Galambos copula as these have simple forms that permit a relatively easy calculation of the copula density (which is needed for likelihood inference). Even if the "true" underlying copula were t, it would seldom make sense to use the more complicated t–EV copula, since the dependence function in (16.24) can be accurately approximated by the dependence function of a Gumbel copula for many values of ν and ρ.

The Gumbel copula also allows us to explore the possibility of asymmetry by using the general non-exchangeable family described in Example 16.16. For applications in dimensions higher than two, the higher-dimensional extensions of Gumbel discussed in Sections 7.4.2 and 15.2.2 may be useful, although we should stress again that multivariate extreme value models are best suited to low-dimensional applications.

Putting these considerations together, data on multivariate maxima could be modelled using the df $H_{\boldsymbol{\xi},\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\theta}}(\boldsymbol{x}) = C_{\boldsymbol{\theta}}(H_{\xi_1,\mu_1,\sigma_1}(x_1), \ldots, H_{\xi_d,\mu_d,\sigma_d}(x_d))$ for some tractable parametric EV copula $C_{\boldsymbol{\theta}}$. The usual method involves maximum likelihood inference and the maximization can either be performed in one step for all parameters of the margins and copula or broken into two steps, whereby marginal models are estimated first and then a parametric copula is fitted using the ideas in Sections 7.5.2 and 7.5.3. The following bivariate example gives an idea of the kind of inference that can be made with such a model.

Example 16.22. Let $M_{65,1}$ represent the quarterly maximum of daily percentage falls of the US dollar against the euro and let $M_{65,2}$ represent the quarterly maximum of daily percentage falls of the US dollar against the yen. We define a stress event for each of these daily return series: for the dollar against the euro we might be concerned about a 4% fall in any one day; for the dollar against the yen we might be concerned about a 5% fall in any one day. We want to estimate the unconditional probability that one or both of these stress events occurs over any quarter. The probability p of interest is given by $p = 1 - P(M_{65,1} \leqslant 4\%, M_{65,2} \leqslant 5\%)$ and is approximated by $1 - H_{\boldsymbol{\xi},\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\theta}}(0.04, 0.05)$, where the parameters are estimated from the block maxima data. Of course, a more worrying scenario might be that both of these stress events should occur on the *same* day. To calculate the probability of simultaneous extreme events we require a different methodology, which is developed in Section 16.4.

Notes and Comments

Early works on distributions for bivariate extremes include Geffroy (1958), Tiago de Oliveira (1958) and Sibuya (1960). A selection of further important papers in the development of the subject include Galambos (1975), de Haan and Resnick (1977), Balkema and Resnick (1977), Deheuvels (1980) and Pickands (1981). The texts by Galambos (1987) and Resnick (2008) have both been influential; our presentation more closely resembles the former.

Theorem 16.14 is proved in Galambos (1987): see Theorem 5.2.1 and Lemma 5.4.1 therein (see also Joe 1997, p. 173). Theorem 16.15 is essentially a result of Pickands (1981). A complete version of the proof is given in Theorem 5.4.5 of

Galambos (1987), although it is given in the form of a characterization of MEV distributions with Gumbel margins. This is easily reformulated as a characterization of the EV copulas. In the bivariate case, necessary and sufficient conditions for $A(w)$ in (16.16) to define a bivariate EV copula are given in Joe (1997, Theorem 6.4).

The copula of Example 16.17 appears in Galambos (1975). A good summary of other bivariate and multivariate extreme value copulas is found in Kotz and Nadarajah (2000); they are presented as MEV distributions with unit Fréchet margins but the EV copulas are easily inferred from this presentation. See also Joe (1997, Chapters 5 and 6), in which EV copulas and their higher-dimensional extensions are discussed. Many parametric models for extremes have been suggested by Tawn (1988, 1990).

Theorem 16.18 is found in Galambos (1987), where the necessary and sufficient copula convergence criterion is given as $\lim_{n\to\infty} C^n(\boldsymbol{u}^{1/n}) = C_0(\boldsymbol{u})$ for positive integers n; by noting that for any $t > 0$ we have the inequalities

$$C^{[t]+1}(\boldsymbol{u}^{1/[t]}) \leqslant C^t(\boldsymbol{u}^{1/t}) \leqslant C^{[t]}(\boldsymbol{u}^{1/([t]+1)}),$$

it can be inferred that this is equivalent to $\lim_{t\to\infty} C^t(\boldsymbol{u}^{1/t}) = C_0(\boldsymbol{u})$. Further equivalent CDA conditions are found in Takahashi (1994). The idea of a domain of attraction of an EV copula also appears in Abdous, Ghoudi and Khoudraji (1999). Not every copula is in a copula domain of attraction; a counterexample may be found in Schlather and Tawn (2002).

We have shown that pairwise asymptotic independence for the components of random vectors implies pairwise independence of the corresponding components in the limiting MEV distribution of the maxima. Pairwise independence for an MEV distribution in fact implies mutual independence, as recognized and described by a number of authors: see Galambos (1987, Corollary 5.3.1), Resnick (2008, Theorem 5.27), and the earlier work of Geffroy (1958) and Sibuya (1960).

16.4 Multivariate Threshold Exceedances

In this section we describe practically useful models for multivariate extremes (again in low-dimensional applications) that build on the basic idea of modelling excesses over high thresholds with the generalized Pareto distribution (GPD) as in Section 5.2. The idea is to use GPD-based tail models of the kind discussed in Section 5.2.3 together with appropriate copulas to obtain models for multivariate threshold exceedances.

16.4.1 Threshold Models Using EV Copulas

Assume that the vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ have unknown joint distribution $F(\boldsymbol{x}) = C(F_1(x_1), \ldots, F_d(x_d))$ for some unknown copula C and margins $F_1, \ldots, F_d$, and that F is in the domain of attraction of an MEV distribution. Much as in the univariate case we would like to approximate the upper tail of $F(\boldsymbol{x})$ above some vector of high thresholds $\boldsymbol{u} = (u_1, \ldots, u_d)'$. The univariate theory of Sections 5.2.2 and 5.2.3 tells us that, for $x_j \geqslant u_j$ and u_j high enough, the tail of the marginal distribution

F_j may be approximated by a GPD-based functional form

$$\tilde{F}_j(x_j) = 1 - \lambda_j \left(1 + \xi_j \frac{x_j - u_j}{\beta_j}\right)^{-1/\xi_j}, \tag{16.25}$$

where $\lambda_j = \bar{F}_j(u_j)$. This suggests that for $\boldsymbol{x} \geqslant \boldsymbol{u}$ we use the approximation $F(\boldsymbol{x}) \approx C(\tilde{F}_1(x_1), \ldots, \tilde{F}_d(x_d))$. But C is also unknown and must itself be approximated in the tail. The following heuristic argument suggests that we should be able to replace C by its limiting copula C_0.

The CDA condition (16.21) suggests that for any value $\boldsymbol{v} \in (0, 1)^d$ and t sufficiently large we may make the approximation $C(\boldsymbol{v}^{1/t}) \approx C_0^{1/t}(\boldsymbol{v})$. If we now write $\boldsymbol{w} = \boldsymbol{v}^{1/t}$, we have

$$C(\boldsymbol{w}) \approx C_0^{1/t}(\boldsymbol{w}^t) = C_0(\boldsymbol{w}) \tag{16.26}$$

by the scaling property of EV copulas. The approximation (16.26) will be best for large values of $\boldsymbol{w}$, since $\boldsymbol{v}^{1/t} \to \mathbf{1}$ as $t \to \infty$.

We assume then that we can substitute the copula C with its EV limit C_0 in the tail, and this gives us the overall model

$$\tilde{F}(\boldsymbol{x}) = C_0(\tilde{F}_1(x_1), \ldots, \tilde{F}_d(x_d)), \quad \boldsymbol{x} \geqslant \boldsymbol{u}. \tag{16.27}$$

We complete the model specification by choosing a flexible and tractable parametric EV copula for C_0. As before, the Gumbel copula family is particularly convenient.

16.4.2 Fitting a Multivariate Tail Model

Assume we have observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from a df F with a tail that permits the approximation (16.27). Of these observations, only a minority are likely to be in the joint tail ($\boldsymbol{x} \geqslant \boldsymbol{u}$); other observations may exceed some of the individual thresholds but lie below others. The usual way of making inferences about all the parameters of such a model (the marginal parameters $\xi_j, \beta_j, \lambda_j$ for $j = 1, \ldots, d$ and the copula parameter (or parameter vector) θ is to maximize a likelihood for *censored data*.

Let us suppose that m_i components of the data vector $\boldsymbol{X}_i$ exceed their respective thresholds in the vector $\boldsymbol{u}$. The only relevant information that the remaining components convey is that they lie below their thresholds; such a component $X_{i,j}$ is said to be censored at the value u_j. The contribution to the likelihood of $\boldsymbol{X}_i$ is given by

$$L_i = L_i(\boldsymbol{\xi}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\theta}; \boldsymbol{X}_i) = \left. \frac{\partial^{m_i} \tilde{F}(x_1, \ldots, x_d)}{\partial x_{j_1} \cdots \partial x_{j_{m_i}}} \right|_{\max(\boldsymbol{X}_i, \boldsymbol{u})},$$

where the indices $j_1, \ldots, j_{m_i}$ are those of the components of $\boldsymbol{X}_i$ exceeding their thresholds.

For example, in a bivariate model with Gumbel copula (7.12), the likelihood contribution would be

$$L_i = \begin{cases} C_\theta^{\mathrm{Gu}}(1 - \lambda_1, 1 - \lambda_2), & X_{i,1} \leqslant u_1,\ X_{i,2} \leqslant u_2, \\ C_{\theta,1}^{\mathrm{Gu}}(\tilde{F}_1(X_{i,1}), 1 - \lambda_2)\tilde{f}_1(X_{i,1}), & X_{i,1} > u_1,\ X_{i,2} \leqslant u_2, \\ C_{\theta,2}^{\mathrm{Gu}}(1 - \lambda_1, \tilde{F}_2(X_{i,2}))\tilde{f}_2(X_{i,2}), & X_{i,1} \leqslant u_1,\ X_{i,2} > u_2, \\ c_\theta^{\mathrm{Gu}}(\tilde{F}_1(X_{i,1}), \tilde{F}_2(X_{i,2}))\tilde{f}_1(X_{i,1})\tilde{f}_2(X_{i,2}), & X_{i,1} > u_1,\ X_{i,2} > u_2, \end{cases} \tag{16.28}$$

Table 16.2. Parameter estimates and standard errors (in brackets) for a bivariate tail model fitted to exchange-rate return data; see Example 16.23 for details.

	\$/€	\$/¥
u	0.75	1.00
N_u	189	126
λ	0.094 (0.0065)	0.063 (0.0054)
ξ	−0.049 (0.066)	0.095 (0.11)
β	0.33 (0.032)	0.38 (0.053)
θ	1.10 (0.030)	

where $\tilde{f}_j$ denotes the density of the univariate tail model $\tilde{F}_j$ in (16.25), $c_\theta^{\mathrm{Gu}}(u_1, u_2)$ denotes the Gumbel copula density, and $C_{\theta,j}^{\mathrm{Gu}}(u_1, u_2) := (\partial/\partial u_j)C_\theta^{\mathrm{Gu}}(u_1, u_2)$ denotes a conditional distribution of the copula, as in (7.17). The overall likelihood is a product of such contributions and is maximized with respect to all parameters of the marginal models and copula.

In a simpler approach, parameters of the marginal GPD models could be estimated as in Section 5.2.3 and only the parameters of the copula obtained from the above likelihood. In fact, this is also a sensible way of getting starting values before going on to the global maximization over all parameters.

The model described by the likelihood (16.28) has been studied in some detail by Ledford and Tawn (1996) and a number of related models have been studied in the statistical literature on multivariate EVT (see Notes and Comments for more details).

Example 16.23 (bivariate tail model for exchange-rate return data). We analyse daily percentage falls in the value of the US dollar against the euro and the Japanese yen, taking data for the eight-year period 1996–2003. We have 2008 daily returns and choose to set thresholds at 0.75% and 1.00%, giving 189 and 126 exceedances, respectively. In a full maximization of the likelihood over all parameters we obtained the estimates and standard errors shown in Table 16.2. The value of the maximized log-likelihood is −1064.7, compared with −1076.4 in a model where independence in the tail is assumed (i.e. a Gumbel copula with $\theta = 1$), showing strong evidence against an independence assumption.

We can now use the fitted model (16.27) to make various calculations about stress events. For example, an estimate of the probability that on any given day the dollar falls by more than 2% against both currencies is given by

$$p_{12} := 1 - \tilde{F}_1(2.00) - \tilde{F}_2(2.00) + C_\theta^{\mathrm{Gu}}(\tilde{F}_1(2.00), \tilde{F}_2(2.00)) = 0.000\,315,$$

with $\tilde{F}_j$ as in (16.25), making this approximately a 13-year event (assuming 250 trading days per year). The marginal probabilities of falls in value of this magnitude are $p_1 := 1 - \tilde{F}_1(2.00) = 0.0014$ and $p_2 := 1 - \tilde{F}_2(2.00) = 0.0061$. We can use this information to calculate so-called spillover probabilities for the conditional occurrence of stress events; for example, the probability that the dollar falls 2% against the yen given that it falls 2% against the euro is estimated to be $p_{12}/p_1 = 0.23$.

16.4.3 Threshold Copulas and Their Limits

An alternative approach to multivariate extremes looks explicitly at the kind of copulas we get when we condition observations to lie above or below extreme thresholds. Just as the GPD is a natural limiting model for univariate threshold exceedances, so we can find classes of copula that are natural limiting models for the dependence structure of multivariate exceedances.

The theory has been studied in most detail in the case of exchangeable bivariate copulas, and we concentrate on this case. Moreover, it proves slightly easier to switch our focus at this stage and first consider the lower-left tail of a probability distribution, before showing how the theory is adapted to the upper-right tail.

Lower threshold copulas and their limits. Consider a random vector (X_1, X_2) with continuous margins F_1 and F_2 and an exchangeable copula C. We consider the distribution of (X_1, X_2) conditional on both being below their v-quantiles, an event we denote by $A_v = \{X_1 \leqslant F_1^{\leftarrow}(v),\ X_2 \leqslant F_2^{\leftarrow}(v)\}$, $0 < v \leqslant 1$. Assuming $C(v, v) \neq 0$, the probability that X_1 lies below its x_1-quantile and X_2 lies below its x_2-quantile conditional on this event is

$$P(X_1 \leqslant F_1^{\leftarrow}(x_1), X_2 \leqslant F_2^{\leftarrow}(x_2) \mid A_v) = \frac{C(x_1, x_2)}{C(v, v)}, \quad x_1, x_2 \in [0, v].$$

Considered as a function of x_1 and x_2 this defines a bivariate df on $[0, v]^2$, and by Sklar's Theorem we can write

$$\frac{C(x_1, x_2)}{C(v, v)} = C_v^0(F_{(v)}(x_1), F_{(v)}(x_2)), \quad x_1, x_2 \in [0, v], \tag{16.29}$$

for a unique copula C_v^0 and continuous marginal distribution functions

$$F_{(v)}(x) = P(X_1 \leqslant F_1^{\leftarrow}(x) \mid A_v) = \frac{C(x, v)}{C(v, v)}, \quad 0 \leqslant x \leqslant v. \tag{16.30}$$

This unique copula may be written as

$$C_v^0(u_1, u_2) = \frac{C(F_{(v)}^{\leftarrow}(u_1), F_{(v)}^{\leftarrow}(u_2))}{C(v, v)}, \tag{16.31}$$

and it will be referred to as the *lower threshold copula* of C at level v. Juri and Wüthrich (2002), who developed the approach we describe in this section, refer to it as a lower tail dependence copula. It is of interest to attempt to evaluate limits for this copula as $v \to 0$; such a limit will be known as a *limiting lower threshold copula*.

Much like the GPD in Example 5.19, limiting lower threshold copulas must possess a stability property under the operation of calculating lower threshold copulas in (16.31). A copula C is a limiting lower threshold copula if, for any threshold $0 < v \leqslant 1$, it satisfies

$$C_v^0(u_1, u_2) = C(u_1, u_2). \tag{16.32}$$

Example 16.24 (Clayton copula as limiting lower threshold copula). For the standard bivariate Clayton copula in (7.13) we can easily calculate that $F_{(v)}$ in (16.30) is

$$F_{(v)}(x) = \frac{(x^{-\theta} + v^{-\theta} - 1)^{-1/\theta}}{(2v^{-\theta} - 1)^{-1/\theta}}, \quad 0 \leqslant x \leqslant v,$$

and its inverse is

$$F_{(v)}^{\leftarrow}(u) = u(2v^{-\theta} - 1 + u^{\theta}(1 - v^{-\theta}))^{-1/\theta}, \quad 0 \leqslant u \leqslant 1.$$

The lower threshold copula for the Clayton copula can therefore be calculated from (16.31) and it may be verified that this is again the Clayton copula. In other words, the Clayton copula is a limiting lower threshold copula because (16.32) holds.

Upper threshold copulas. To define upper threshold copulas we consider again a random vector (X_1, X_2) with copula C and margins F_1 and F_2. We now condition on the event $\bar{A}_v = \{X_1 > F_1^{\leftarrow}(v),\ X_2 > F_2^{\leftarrow}(v)\}$ for $0 \leqslant v < 1$. We have the identity

$$P(X_1 > F_1^{\leftarrow}(x_1), X_2 > F_2^{\leftarrow}(x_2) \mid \bar{A}_v) = \frac{\bar{C}(x_1, x_2)}{\bar{C}(v, v)}, \quad x_1, x_2 \in [v, 1].$$

Since $\bar{C}(x_1, x_2)/\bar{C}(v, v)$ defines a bivariate survival function on $[v, 1]^2$, by (7.14) we can write

$$\frac{\bar{C}(x_1, x_2)}{\bar{C}(v, v)} = \hat{C}_v^1(\bar{G}_{(v)}(x_1), \bar{G}_{(v)}(x_2)), \quad x_1, x_2 \in [v, 1], \tag{16.33}$$

for some survival copula $\hat{C}_v^1$ of a copula C_v^1 and marginal survival functions

$$\bar{G}_{(v)}(x) = P(X_1 > F_1^{\leftarrow}(x) \mid \bar{A}_v) = \frac{\bar{C}(x, v)}{\bar{C}(v, v)}, \quad v \leqslant x \leqslant 1. \tag{16.34}$$

The copula C_v^1 is known as the *upper threshold copula* at level v and it is now of interest to find limits as $v \to 1$, which are known as limiting upper threshold copulas. In fact, as the following lemma shows, it suffices to study either lower or upper threshold copulas because results for one follow easily from results for the other.

Lemma 16.25. *The survival copula of the upper threshold copula of C at level v is the lower threshold copula of $\hat{C}$ at level $1 - v$.*

Proof. We use the identity $\bar{C}(u_1, u_2) = \hat{C}(1 - u_1, 1 - u_2)$ and (16.34) to rewrite (16.33) as

$$\frac{\hat{C}(1 - x_1, 1 - x_2)}{\hat{C}(1 - v, 1 - v)} = \hat{C}_v^1\left(\frac{\hat{C}(1 - x_1, 1 - v)}{\hat{C}(1 - v, 1 - v)}, \frac{\hat{C}(1 - v, 1 - x_2)}{\hat{C}(1 - v, 1 - v)}\right).$$

Writing $y_1 = 1 - x_1$, $y_2 = 1 - x_2$ and $w = 1 - v$ we have

$$\frac{\hat{C}(y_1, y_2)}{\hat{C}(w, w)} = \hat{C}_{1-w}^1\left(\frac{\hat{C}(y_1, w)}{\hat{C}(w, w)}, \frac{\hat{C}(w, y_2)}{\hat{C}(w, w)}\right), \quad y_1, y_2 \in [0, w],$$

and comparison with (16.29) and (16.30) shows that $\hat{C}_{1-w}^1$ must be the lower threshold copula of $\hat{C}$ at the level $w = 1 - v$. □

It follows that the survival copulas of limiting lower threshold copulas are limiting upper threshold copulas. The Clayton survival copula is a limiting upper threshold copula.

Relationship between limiting threshold copulas and EV copulas. We give one result that shows how limiting upper threshold copulas may be calculated for underlying exchangeable copulas C that are in the domain of attraction of EV copulas with tail dependence, thus linking the study of threshold copulas to the theory of Section 16.3.3.

Theorem 16.26. *If C is an exchangeable copula with upper tail-dependence coefficient $\lambda_u > 0$ satisfying $C \in \mathrm{CDA}(C_0)$, then C has a limiting upper threshold copula that is the survival copula of the df*

$$G(x_1, x_2) = \frac{(x_1 + x_2)(1 - A(x_1/(x_1 + x_2)))}{\lambda_u}, \tag{16.35}$$

where A is the dependence function of C_0. Also, $\hat{C}$ has a limiting lower threshold copula that is the copula of G.

Example 16.27 (upper threshold copula of Galambos copula). We use this result to calculate the limiting upper threshold copula for the Galambos copula. We recall that this is an EV copula with dependence function given in (16.19) and consider the standard exchangeable case with $\alpha = \beta = 1$. Using the methods of Section 7.2.4 it may easily be calculated that the coefficient of upper tail dependence of this copula is $\lambda_u = 2^{-1/\theta}$. The bivariate distribution $G(x_1, x_2)$ in (16.35) is therefore given by

$$G(x_1, x_2) = (\tfrac{1}{2}(x_1^{-\theta} + x_2^{-\theta}))^{-1/\theta}, \quad (x_1, x_2) \in (0, 1]^2,$$

the copula of which is the Clayton copula. The limiting upper threshold copula in this case is therefore the Clayton survival copula. Moreover, the limiting lower threshold copula of the Galambos survival copula is the Clayton copula.

The Clayton copula turns out to be an important attractor for a large class of underlying exchangeable copulas. Juri and Wüthrich (2003) have shown that all Archimedean copulas whose generators are regularly varying at 0 with negative parameter (meaning that $\phi(t)$ satisfies $\lim_{t\to 0} \phi(xt)/\phi(t) = x^{-\alpha}$ for all x and some $\alpha > 0$) share the Clayton copula C_α^{Cl} as their limiting lower threshold copula.

It is of interest to calculate limiting lower and upper threshold copulas for the t copula, and this can be done using Theorem 16.26 and the expression for the dependence function in (16.24). However, the resulting limit is not convenient for practical purposes because of the complexity of this dependence function. We have already remarked in Section 16.3.4 that the dependence function of the t–EV copula can be well approximated by the dependence functions of other exchangeable EV copulas, such as Gumbel and Galambos, for most practical purposes. Theorem 16.26 therefore suggests that instead of working with the true limiting upper threshold copula of the t copula we could instead work with the limiting upper threshold copula of, say, the Galambos copula, i.e. the Clayton survival copula. Similarly, we could work with the Clayton copula as an approximation for the true limiting lower threshold copula of the t copula.

Limiting threshold copulas in practice Limiting threshold copulas in dimensions higher than two have not yet been extensively studied, nor have limits for non-exchangeable bivariate copulas or limits when we define two thresholds v_1 and v_2 and let these tend to zero (or one) at different rates. The practical use of these ideas is therefore largely confined to bivariate applications when thresholds are set at approximately similar quantiles and a symmetric dependence structure is assumed.

Let us consider a situation where we have a bivariate distribution that appears to exhibit tail dependence in both the upper-right and lower-left corners. While true lower and upper limiting threshold copulas may exist for this unknown distribution, we could in practice simply adopt a tractable and flexible parametric limiting threshold copula family. It is particularly easy to use the Clayton copula and its survival copula as lower and upper limits, respectively.

Suppose, for example, that we set high thresholds at $\boldsymbol{u} = (u_1, u_2)'$, so that $P(X_1 > u_1) \approx P(X_2 > u_2)$ and both probabilities are small. For the conditional distribution of (X_1, X_2) over the threshold $\boldsymbol{u}$ we could assume a model of the form

$$P(\boldsymbol{X} \leqslant \boldsymbol{x} \mid \boldsymbol{X} > \boldsymbol{u}) \approx \hat{C}_\theta^{\mathrm{Cl}}(G_{\xi_1,\beta_1}(x_1 - u_1), G_{\xi_2,\beta_2}(x_2 - u_2)), \quad \boldsymbol{x} > \boldsymbol{u},$$

where $\hat{C}_\theta^{\mathrm{Cl}}$ is the Clayton survival copula and G_{ξ_j,β_j} denotes a GPD, as defined in 5.16. Inference about the model parameters $(\theta, \xi_1, \beta_1, \xi_2, \beta_2)$ would be based on the exceedance data above the thresholds and would use the methods discussed in Section 7.5.

Similarly, for a vector of low thresholds $\boldsymbol{u}$ satisfying $P(X_1 \leqslant u_1) \approx P(X_2 \leqslant u_2)$ with both these probabilities small, we could approximate the conditional distribution of (X_1, X_2) below the threshold $\boldsymbol{u}$ by a model of the form

$$P(\boldsymbol{X} \leqslant \boldsymbol{x} \mid \boldsymbol{X} < \boldsymbol{u}) \approx C_\theta^{\mathrm{Cl}}(\bar{G}_{\xi_1,\beta_1}(u_1 - x_1), \bar{G}_{\xi_2,\beta_2}(u_2 - x_2)), \quad \boldsymbol{x} < \boldsymbol{u},$$

where C_θ^{Cl} is the Clayton copula and $\bar{G}_{\xi_j,\beta_j}$ denotes a GPD survival function. Inference about the model parameters would be based on the data below the thresholds and would use the methods of Section 7.5.

Note and Comments

The GPD-based tail model (16.27) and inference for censored data using a likelihood of the form (16.28) have been studied by Ledford and Tawn (1996), although the derivation of the model uses somewhat different asymptotic reasoning based on a characterization of multivariate domains of attraction of MEV distributions with unit Fréchet margins found in Resnick (2008). The authors of the former paper concentrate on the model with Gumbel (logistic) dependence structure and discuss, in particular, testing for asymptotic independence of extremes. Likelihood inference is non-problematic (the problem being essentially regular) when $\theta > 0$ and $\xi_j > -\frac{1}{2}$, but testing for independence of extremes $\theta = 1$ is not quite so straightforward since this is a boundary point of the parameter space. This case is possibly more interesting in environmental applications than in financial ones, where we tend to expect dependence of extreme values.

A related bivariate GPD model is presented in Smith, Tawn and Coles (1997). In our notation they essentially consider a model of the form

$$\bar{F}(x_1, \ldots, x_d) = 1 + \ln C_0(\mathrm{e}^{\tilde{F}(x_1)-1}, \ldots, \mathrm{e}^{\tilde{F}(x_d)-1}), \quad \boldsymbol{x} \geqslant \boldsymbol{k},$$

where C_0 is an extreme value copula. This model is also discussed in Smith (1994) and Ledford and Tawn (1996); it is pointed out that $\bar{F}$ does not reduce to a product of marginal distributions in the case when C_0 is the independence copula, unlike the model in (16.27).

Another style of statistical model for multivariate extremes is based on the point process theory of multivariate extremes developed in de Haan (1985), de Haan and Resnick (1977) and Resnick (2008). Statistical models using this theory are found in Coles and Tawn (1991) and Joe, Smith and Weissman (1992); see also the texts of Joe (1997) and Coles (2001). New approaches to modelling multivariate extremes can be found in Heffernan and Tawn (2004) and Balkema and Embrechts (2007); the latter paper considers applications to stress testing high-dimensional portfolios in finance.

Limiting threshold copulas are studied in Juri and Wüthrich (2002, 2003). In the latter paper it is demonstrated that the Clayton copula is an attractor for the threshold copulas of a wide class of Archimedean copulas; moreover, a version of our Theorem 16.26 is proved. Limiting threshold copulas for the t copula are investigated in Demarta and McNeil (2005). The usefulness of Clayton's copula and the Clayton survival copula for describing the dependence in the tails of bivariate financial return data was confirmed in a large-scale empirical study of high-frequency exchange-rate returns by Breymann, Dias and Embrechts (2003).

17

Dynamic Portfolio Credit Risk Models and Counterparty Risk

This chapter is concerned with dynamic reduced-form models of portfolio credit risk. It is also the natural place for an analysis of counterparty credit risk for over-the-counter (OTC) credit derivatives, since this risk can only be satisfactorily modelled in the framework of dynamic models.

In Section 17.1 we give an informal introduction to the subject of dynamic models in which we explain why certain risk-management tasks for portfolios of credit derivatives cannot be properly handled in the copula framework of Chapter 12; we also give an overview of the different types of dynamic model used in portfolio credit risk.

A detailed analysis of counterparty credit risk management is presented in Section 17.2, while the remainder of the chapter focusses on two different approaches to dynamic modelling. In Section 17.3 we consider dynamic models with conditionally independent default times, and in Section 17.4 we treat credit risk models with incomplete information.

We use a number of concepts and techniques from continuous-time finance and build on material in Chapters 10 and 12. Particular prerequisites for reading this chapter are Sections 10.5 and 10.6 on pricing single-name credit derivatives in models with stochastic hazard rates and Section 12.3 on CDO pricing in factor copula models.

17.1 Dynamic Portfolio Credit Risk Models

17.1.1 Why Dynamic Models of Portfolio Credit Risk?

In the copula models of portfolio credit risk in Chapter 12, the joint distribution of the default times was specified directly. However, the evolution of this distribution over time, for instance in reaction to new economic information, was not modelled. While the copula approach is sufficient for computing the prices of many credit products such as index swaps or CDO tranches at a given point in time t, it is not possible to say anything about the dynamics of these prices over time. Certain important tasks in the risk management of credit derivative portfolios cannot therefore be handled properly in the copula framework, as we now discuss.

To begin with, in a copula model it is not possible to price options on credit-risky instruments such as options on a basket of corporate bonds. This is an issue of

practical relevance, since the task of pricing an option on a credit-risky instrument arises in the management of counterparty credit risk management, particularly in the computation of a so-called *credit value adjustment* (CVA) for a credit derivative. Consider, for instance, a protection buyer who has entered into a CDS contract with some protection seller. Suppose now that the protection seller defaults during the lifetime of the CDS contract and that the credit spread of the reference entity has gone up. In that case the protection buyer suffers a loss, since in order to renew his protection he has to enter into a new CDS at a higher spread. In order to account for this loss the protection buyer should make an adjustment to the value of the CDS. We will show in Section 17.2 that, roughly speaking, this adjustment takes the form of an option on the value of the future cash flows of the CDS with maturity date equal to the default time of the protection seller.

Moreover, in copula models it is not possible to derive model-based dynamic hedging strategies. For this reason risk managers often resort to sensitivity-based hedging strategies that are similar to the use of duration-based immunization for the risk management of bond portfolios. This is not entirely satisfactory, since it is known from markets for other types of derivatives that hedging strategies that lack a proper theoretical foundation may perform poorly (see Notes and Comments for references). On the theoretical side, the non-existence of model-based hedging strategies implies that pricing results in copula models are not supported by hedging arguments, so that an important insight of modern derivative asset analysis is ignored in standard CDO pricing. Of course, dynamic trading strategies are substantially more difficult to implement in credit-derivative markets than in equity markets and their performance is less robust with respect to model misspecification. However, in our view this should not serve as an excuse for neglecting the issue of dynamic modelling and dynamic hedging altogether. A number of recent papers on the hedging of portfolio credit derivatives are listed in Notes and Comments.

17.1.2 Classes of Reduced-Form Models of Portfolio Credit Risk

The properties of a reduced-form model of portfolio credit risk are essentially determined by the *intensities* of the default times. Different models can be classified according to the mathematical structure given to these intensities. The concept of the intensity of a random default time was encountered in Definition 10.15 and will be explained in more detail in Section 17.3.1.

The simplest reduced-form portfolio models are models with *conditionally independent defaults* (see Section 17.3). These models are a straightforward multi-firm extension of models with doubly stochastic default times. Default times are modelled as conditionally independent given the realization of some observable economic background process $(\boldsymbol{\Psi}_t)$. The default intensities are adapted to the filtration generated by $(\boldsymbol{\Psi}_t)$, and dependence between defaults is generated by the mutual dependence of the default intensities on the common background process. An important special case is the class of models with Markov-modulated intensities where $\lambda_{t,i} = \lambda_i(\boldsymbol{\Psi}_t)$ and $(\boldsymbol{\Psi}_t)$ is Markovian.

In Section 17.4 we consider reduced-form models with *incomplete information*. We consider a set-up where the default times are independent given a common factor. In the simplest case this factor is simply a random variable V, as in the factor copula models studied in Section 12.2.2. Assume for the moment that V is observable. Then defaults are independent and the default intensities have the simple form $\tilde{\lambda}_{t,i} = \gamma_i(V, t)$ for suitable functions γ_i. We assume, however, that the factor is not directly observable. Instead investors are only able to observe the default history and, at most, an auxiliary process representing noisy observations of V. In this context it will be shown that the default intensities are computed by projection. Denoting the default intensity of firm i at time t by $\lambda_{t,i}$ and the information available to investors at t by the σ-field $\mathcal{G}_t$, we have that

$$\lambda_{t,i} = E(\gamma_i(V, t) \mid \mathcal{G}_t). \tag{17.1}$$

Moreover, prices of many credit derivatives are given by similar conditional expectations. We will see that Bayesian updating and, more generally, stochastic filtering techniques can be employed in the evaluation of expressions of the form (17.1) and in the analysis of the model in general.

A further important model class comprises models with *interacting intensities*. We will not present these in any detail for reasons of space but important contributions to the literature are listed in Notes and Comments. In these models the impact of the default of one firm on the default intensities of surviving firms is specified explicitly. For instance, we might assume that, after a default event, default intensities increase by 10% of their pre-default value. This interaction among intensities provides an alternative mechanism for creating dependence between default events. In formal terms, models with interacting intensities are constructed as Markov chains on the set of all possible default states of the portfolio. The theory of Markov chains therefore plays an important role in their analysis.

A common feature of models with interacting default intensities and models with incomplete information is the presence of *default contagion*. This means that the default intensity of a surviving firm jumps (usually upwards) given the information that some other firm has defaulted. As a consequence, the credit spread of surviving firms increases when default events occur in the portfolio. Default contagion can arise via different channels. On the one hand, contagious effects might be due to direct economic links between firms, such as a close business relationship or a strong borrower–lender relationship. For instance, the default probability of a bank is likely to increase if one of its major borrowers defaults. Broadly speaking, this channel of default interaction is linked to counterparty risk. On the other hand, changes in the conditional default probability of non-defaulted firms can be caused by information effects; investors might revise their estimate of the financial health of non-defaulted firms in light of the news that a particular firm has defaulted. This is known as *information-based default contagion*. Note that models with incomplete information generate information-based default contagion by design. At a default event the conditional distribution of the factor V given the investor information is updated; by (17.1) this leads to a jump in the default intensity $\lambda_{t,i}$. There is in

fact substantial evidence for contagion effects. A good example is provided by the default of the investment bank Lehman Brothers in autumn 2008; the default event combined with the general nervousness caused by the worsening financial crisis sent credit spreads to unprecedentedly high levels.

We can also distinguish between *bottom-up* and *top-down* models of portfolio credit risk. This distinction relates to the quantities that are modelled and cuts through all types of reduced-form portfolio credit risk models including the copula models of Chapter 12.

The fundamental objects that are modelled in a bottom-up model are the default indicator processes of the individual firms in the portfolio under consideration; the dynamics of the portfolio loss are then derived from these. In this approach it is possible to price portfolio products consistently with observed single-name CDS spreads and to derive hedging strategies for portfolio products that use single-name CDSs as hedging instruments. These are obvious advantages of this model class. However, there are also some drawbacks related to tractability: in the bottom-up approach we have to keep track of all default-indicator processes and possibly also background processes driving the model. This typically leads to substantial computational challenges in pricing and model calibration, particularly if the portfolio size is fairly large.

In top-down models, on the other hand, the portfolio loss process is modelled directly, without reference to the individual default indicator processes. This obviously drastically reduces the dimensionality of the resulting models. It can be argued that top-down models are sufficient for the pricing of index derivatives, since the payoff of these contracts depends only on the portfolio loss. However, in this model class the information contained in single-name spreads is neglected for pricing purposes, and it is not possible to study the hedging of portfolio derivatives with single-name CDSs.

There is no obvious and universally valid answer to the question of which model class should be preferred; in Notes and Comments we provide a few references in which this issue is discussed further. In our own analysis we concentrate on bottom-up models.

Notes and Comments

The limitations of static copula models are discussed in a number of research papers; a particularly readable contribution is Shreve (2009). An interesting collection of papers that deal with portfolio credit risk models "after copulas" is found in Lipton and Rennie (2008). Dynamic hedging strategies for portfolio credit derivatives are studied by Frey and Backhaus (2010), Laurent, Cousin and Fermanian (2011) and Cont and Kan (2011), among others; an earlier contribution is Bielecki, Jeanblanc and Rutkowski (2004). A detailed mathematical analysis of hedging errors for equity and currency derivatives is given in El Karoui, Jeanblanc-Picqué and Shreve (1998).

There is a rich literature on models with interacting intensities. Bottom-up models are considered by Davis and Lo (2001), Jarrow and Yu (2001), Yu (2007), Frey and Backhaus (2008) and Herbertsson (2008). Top-down models with interacting

intensities include the contributions by Arnsdorf and Halperin (2009) and Cont and Minca (2013). Moreover, there are top-down models where the dynamics of the whole "surface" of CDO tranche spreads—that is the dynamics of CDO spreads for all maturities and attachment points—are modelled directly: see, for example, Ehlers and Schönbucher (2009), Sidenius, Piterbarg and Andersen (2008) and Filipović, Overbeck and Schmidt (2011). The modelling philosophy of these three papers is akin to the well-known HJM models for the term structure of interest rates. A general discussion of the pros and cons of bottom-up and top-down models can be found in Bielecki, Crépey, and Jeanblanc (2010) (see also Giesecke, Goldberg and Ding 2011).

Credit risk models with explicitly specified interactions between default intensities are conceptually close to network models and to models for interacting particle systems developed in statistical physics. Föllmer (1994) contains an inspiring discussion of the relevance of these ideas to financial modelling; the link to credit risk is explored by Giesecke and Weber (2004, 2006) and Horst (2007). Network models are frequently used for the study of systemic risk in financial networks, an issue that has become highly relevant in the aftermath of the financial crisis of 2007–9. Interesting contributions in this rapidly growing field include the papers by Eisenberg and Noe (2001), Elsinger, Lehar and Summer (2006), Gai and Kapadia (2010), Upper (2011) and Amini, Cont and Minca (2012).

There are some empirical papers on default contagion. Jarrow and Yu (2001) provide anecdotal evidence for counterparty-risk-related contagion in small portfolios. In contrast, Lando and Nielsen (2010) find no strong empirical evidence for default contagion in historical default patterns.

Other work has tested the impact of the default or spread widening of a given firm on the credit spreads or stock returns of other firms: see Collin-Dufresne, Goldstein and Helwege (2010) or Lang and Stulz (1992). The evidence in favour of this type of default contagion is quite strong. For instance, Collin-Dufresne, Goldstein and Helwege (2010) found that, even after controlling for other macroeconomic variables influencing bond returns, the returns of large corporate bond indices in months where one or more large firms experienced a significant widening in credit spreads (above 200 basis points) were significantly lower than the returns of these indices in other months; this is clear evidence supporting contagion.

17.2 Counterparty Credit Risk Management

A substantial proportion of all derivative transactions are carried out OTC, so that counterparty credit risk is a key issue for financial institutions. The management of counterparty risk poses a number of challenges. To begin with, a financial institution needs to measure (in close to real time) its counterparty risk exposure to its various trading partners. Moreover, counterparty risk needs to be taken into account in the pricing of derivative contracts, which leads to the issue of computing credit value adjustments. Finally, financial institutions and major corporations apply various risk-mitigation strategies in order to control and reduce their counterparty risk exposure. In particular, many OTC derivative transactions are now *collateralized*.

Consider a derivative transaction such as a CDS contract between two parties—the protection seller S and the protection buyer B—and suppose that the deal is collateralized. If the value of the CDS is negative for, say, S, then S passes cash or securities (the collateral) to B. If S defaults before the maturity of the underlying CDS and if the value of the CDS at the default time τ_S of S is positive for B, the protection buyer is permitted to liquidate the collateral in order to reduce the loss due to the default of S; excess collateral must be returned to S. Most collateralization agreements are symmetric so that the roles of S and B can be exchanged when the value of the underlying CDS changes its sign.

In this section we study quantitative aspects of counterparty risk management. In Section 17.2.1 we introduce the general form of credit value adjustments for uncollateralized derivative transactions and we discuss various simplifications that are used in practice. In Section 17.2.2 we consider the case of collateralized transactions. For concreteness, we discuss value adjustments and collateralization strategies for a single-name CDS, but our arguments apply to other contracts as well.

17.2.1 Uncollateralized Value Adjustments for a CDS

We begin with an analysis of the form of credit value adjustments for an uncollateralized single-name CDS contract on some reference entity R. We work on a filtered probability space $(\Omega, \mathcal{G}, (\mathcal{G}_t), Q)$, where Q denotes the risk-neutral measure used for pricing derivatives and where the filtration $(\mathcal{G}_t)$ represents the information available to investors. Our notation is as follows: the default times of the protection seller S, the protection buyer B and the reference entity R are denoted by the $(\mathcal{G}_t)$ stopping times τ_S, τ_B and τ_R; δ^R, δ^S and δ^B are the losses given default (LGDs) of the contracting parties; $T_1 = \min\{\tau_R, \tau_S, \tau_B\}$ denotes the first default time; $\xi_1 \in \{R, S, B\}$ gives the identity of the firm that defaults first. We assume that δ^R, δ^S and δ^B are constant; for a discussion of the calibration of these parameters in the context of counterparty credit risk, we refer to Gregory (2012).

The CDS contract referencing R has premium payment dates $t_1 < \cdots < t_N = T$, where t_1 is greater than the current time t and a fixed spread x. The default-free short rate is given by the $(\mathcal{G}_t)$-adapted process (r_t). In discounting future cash flows it will be convenient to use the abbreviation

$$D(s_1, s_2) = \exp\left(-\int_{s_1}^{s_2} r_u \, \mathrm{d}u\right), \quad 0 \leqslant s_1 \leqslant s_2 \leqslant T.$$

The promised cash flow of a protection buyer position in the CDS between two time points $s_1 < s_2$, discounted back to time s_1, will be denoted by $\Pi(s_1, s_2)$. Ignoring for simplicity accrued premium payments, we therefore have

$$\Pi(s_1, s_2) = \int_{s_1}^{s_2} D(s_1, u)\delta^R \, \mathrm{d}Y_{R,u} - x \sum_{t_n \in (s_1, s_2]} D(s_1, t_n)(1 - Y_{R,t_n}). \tag{17.2}$$

The first term on the right-hand side of (17.2) represents the discounted default payment, and the second term corresponds to the discounted premium payment.

The value at some stopping time $\tau \geqslant t$ of the promised cash-flow stream for B is then given by

$$V_\tau := E^Q(\Pi(\tau, T) \mid \mathcal{G}_\tau); \tag{17.3}$$

sometimes we will call (V_t) the counterparty-risk-free CDS price. The discounted cash flows that are made or received by B over the period $(s_1, s_2]$ (the real cash flows) are denoted by $\Pi^{\text{real}}(s_1, s_2)$. Note that Π and Π^{real} are in general different as S or B might default before the maturity date T of the transaction.

In order to describe Π^{real} we distinguish the following scenarios.

- If $T_1 > T$ or if $T_1 \leqslant T$ and $\xi_1 = R$, that is if both S and B survive until the maturity date of the CDS, the actual and promised cash-flow streams coincide, so that $\Pi^{\text{real}}(\cdot, T) = \Pi(\cdot, T)$.
- Consider next the scenario where $T_1 < T$ and $\xi_1 = S$, that is the protection seller defaults first and this occurs before the maturity date of the CDS. In that case, prior to T_1, actual and promised cash flows coincide. At T_1, if the counterparty-risk-free CDS price V_{T_1} is positive, B is entitled to charge a *close-out amount* to S in order to settle the contract. Following the literature we assume that this close-out amount is given by V_{T_1}. However, S is typically unable to pay the close-out amount in full, and B receives only a recovery payment of size $(1 - \delta^S)V_{T_1}$. If, on the other hand, V_{T_1} is negative, B has to pay the full amount $|V_{T_1}|$ to S. Using the notation $x^+ = \max\{x, 0\}$ and $x^- = -\min\{x, 0\}$, the actual cash flows on the set $\{T_1 < T\} \cap \{\xi_1 = S\}$ are given by

$$\Pi^{\text{real}}(t, T) = \Pi(t, T_1) + D(t, T_1)((1 - \delta^S)V_{T_1}^+ - V_{T_1}^-). \tag{17.4}$$

- Finally, consider the scenario where $T_1 < T$ and $\xi_1 = B$. If $V_{T_1} > 0$, S has to pay the full amount V_{T_1} to B; if $V_{T_1} < 0$, the protection buyer makes a recovery payment of size $(1 - \delta^B)|V_{T_1}|$ to S. Thus, on the set $\{T_1 < T\} \cap \{\xi_1 = B\}$ we have

$$\Pi^{\text{real}}(t, T) = \Pi(t, T_1) + D(t, T_1)(V_{T_1}^+ - (1 - \delta^B)V_{T_1}^-). \tag{17.5}$$

The correct value of the protection-buyer position in the CDS in the presence of counterparty risk is given by $E^Q(\Pi^{\text{real}}(t, T) \mid \mathcal{G}_t)$. For $t < T_1$ the difference

$$\text{BCVA}_t := E^Q(\Pi(t, T) \mid \mathcal{G}_t) - E^Q(\Pi^{\text{real}}(t, T) \mid \mathcal{G}_t) \tag{17.6}$$

is known as the *bilateral credit value adjustment* (BCVA) at time t. Note that $E^Q(\Pi^{\text{real}}(t, T) \mid \mathcal{G}_t) = V_t - \text{BCVA}_t$: that is, BCVA_t is the adjustment that needs to be made to the counterparty-risk-free CDS price in order to obtain the value of the cash-flow stream Π^{real}. The term *bilateral* refers to the fact that the value adjustment takes the possibility of the default of both contracting parties, B and S, into account. By definition, the bilateral value adjustment is symmetric in the sense that the value adjustment computed from the viewpoint of the protection seller at time

t is given by $-\mathrm{BCVA}_t$; this is obvious since the cash-flow stream received by S is exactly the negative of the cash-flow stream received by B. This contrasts with so-called unilateral value adjustments where each party neglects the possibility of its own default in computing the adjustment to the value of the CDS.

The next proposition gives a more succinct expression for the BCVA.

Proposition 17.1. *For $t < T_1$ we have that* $\mathrm{BCVA}_t = \mathrm{CVA}_t - \mathrm{DVA}_t$, *where*

$$\mathrm{CVA}_t = E^Q(I_{\{T_1 \leqslant T\}} I_{\{\xi_1 = S\}} D(t, T_1) \delta^S V_{T_1}^+ \mid \mathcal{G}_t), \tag{17.7}$$

$$\mathrm{DVA}_t = E^Q(I_{\{T_1 \leqslant T\}} I_{\{\xi_1 = B\}} D(t, T_1) \delta^B V_{T_1}^- \mid \mathcal{G}_t). \tag{17.8}$$

Comments. The CVA in (17.7) reflects the potential loss incurred by B due to a premature default of S; the *debt value adjustment*, or DVA, in (17.8) reflects the potential loss incurred by S due to a premature default of B. A similar formula obviously holds for other products; the only part that needs to be adapted is the definition of the promised cash-flow stream in (17.2).

Accounting rules require that both CVA and DVA have to be taken into account if an instrument is valued via mark-to-market accounting techniques. Note, however, that the use of DVA is somewhat controversial for the following reason: a decrease in the credit quality of B leads to an increase in the probability that B defaults first and hence to a larger DVA term. If both CVA and DVA are taken into account in the valuation of the CDS, this would be reported as a profit for B. It is not clear, however, how B could turn this accounting profit into an actual cash flow for its shareholders.

Proposition 17.1 shows that the problem of computing the BCVA amounts to computing the price of a call option and a put option on (V_t) with strike $K = 0$ and random maturity date T_1. The computation of the value adjustments is therefore more involved than the pricing of the CDS itself, and a dynamic portfolio credit model is needed to compute the value adjustment in a consistent way. The actual computation of value adjustments depends on the structure of the underlying credit model. For further information we refer to Sections 17.3.3 and 17.4.4.

Proof of Proposition 17.1. For $t < T_1 \leqslant s$ it holds that

$$D(t, s) = D(t, T_1) D(T_1, s).$$

Hence, on the set $\{T_1 \leqslant T\}$ we may write $\Pi(t, T) = \Pi(t, T_1) + D(t, T_1)\Pi(T_1, T)$. This yields

$$\Pi(t, T) - \Pi^{\mathrm{real}}(t, T) = I_{\{T_1 \leqslant T\}} D(t, T_1)(\Pi(T_1, T) - I_{\{\xi_1 = S\}}((1 - \delta^S)V_{T_1}^+ - V_{T_1}^-) - I_{\{\xi_1 = B\}}(V_{T_1}^+ - (1 - \delta^B)V_{T_1}^-)).$$

By iterated conditional expectations it follows that

$$E^Q(\Pi(t, T) - \Pi^{\mathrm{real}}(t, T)) = E^Q(E^Q(\Pi(t, T) - \Pi^{\mathrm{real}}(t, T) \mid \mathcal{G}_{T_1})). \tag{17.9}$$

We concentrate on the inner conditional expectation. Since $D(t,T_1)$ and the events $\{T_1 \leqslant T\}$, $\{\xi_1 = S\}$ and $\{\xi_1 = B\}$ are $\mathcal{G}_{T_1}$-measurable, we obtain

$$E^Q(\Pi(t,T) - \Pi^{\text{real}}(t,T) \mid \mathcal{G}_{T_1}) \tag{17.10a}$$

$$= I_{\{T_1 \leqslant T\}} I_{\{\xi_1 = S\}} D(t,T_1) E^Q(\Pi(T_1,T) - ((1-\delta^S)V_{T_1}^+ - V_{T_1}^-) \mid \mathcal{G}_{T_1}) \tag{17.10b}$$

$$+ I_{\{T_1 \leqslant T\}} I_{\{\xi_1 = B\}} D(t,T_1) E^Q(\Pi(T_1,T) - (V_{T_1}^+ - (1-\delta^B)V_{T_1}^-) \mid \mathcal{G}_{T_1}). \tag{17.10c}$$

Now, by the definition of (V_t) we have that $E^Q(\Pi(T_1,T) \mid \mathcal{G}_{T_1}) = V_{T_1}$. Moreover, we use the decomposition $V_{T_1} = V_{T_1}^+ - V_{T_1}^-$, where $V_{T_1}^+$ and $V_{T_1}^-$ are $\mathcal{G}_{T_1}$ measurable. Hence (17.10 *b*) equals $I_{\{T_1 \leqslant T\}} I_{\{\xi_1 = S\}} D(t,T_1)\delta^S V_{T_1}^+$ and, similarly, (17.10 *c*) equals $-I_{\{T_1 \leqslant T\}} I_{\{\xi_1 = B\}} \delta^B V_{T_1}^-$. Putting these together, (17.10 *a*) is equal to

$$I_{\{T_1 \leqslant T\}} D(t,T_1)(I_{\{\xi_1 = S\}}\delta^S V_{T_1}^+ - I_{\{\xi_1 = B\}} D(t,T_1)\delta^B V_{T_1}^-),$$

and substituting this into (17.9) gives the result. □

Simplified value adjustments and wrong-way risk. In order to simplify the computation of value adjustments, it is often assumed that $(Y_{t,S})$, $(Y_{t,B})$ and the counterparty-risk free CDS price (V_t) are independent stochastic processes and that the risk-free interest rate is deterministic. We now explain how the value adjustment formulas (17.7) and (17.8) simplify under this independence assumption. For simplicity we consider the case $t = 0$. Denote by $\bar{F}_S(t)$, $\bar{F}_B(t)$, $f_S(t)$ and $f_B(t)$ the survival functions and densities of τ_S and τ_B. Since $\{\xi_1 = S\} = \{\tau_S < \tau_B\} \cap \{\tau_S < \tau_R\}$ and since $V_{T_1} = V_{\tau_S}$ on $\{\xi_1 = S\}$, we obtain

$$\begin{aligned}\text{CVA} = \text{CVA}_0 &= E^Q(I_{\{\tau_S \leqslant T\}} I_{\{\tau_S < \tau_B\}} I_{\{\tau_S < \tau_R\}} D(0,\tau_S)\delta^S V_{\tau_S}^+) \\ &= \delta^S \int_0^T E^Q(I_{\{\tau_S < \tau_B\}} D(0,\tau_S) V_{\tau_S}^+ \mid \tau_S = t) f_S(t)\,\mathrm{d}t.\end{aligned}$$

In the last line we have used the fact that δ^S is deterministic and we have used the identity $V_s \equiv 0$ on $\{\tau_R \leqslant s\}$, which allows the indicator $I_{\{\tau_S < \tau_R\}}$ to be dropped. The independence of the processes $(Y_{t,S})$, $(Y_{t,B})$, (V_t) and the fact that interest rates are deterministic imply that

$$\begin{aligned}E^Q(I_{\{\tau_S < \tau_B\}} D(0,\tau_S) V_{\tau_S}^+ \mid \tau_S = t) &= E^Q(I_{\{t < \tau_B\}} D(0,t) V_t^+) \\ &= \bar{F}_B(t) D(0,t) E^Q(V_t^+).\end{aligned}$$

Hence $\text{CVA}^{\text{indep}}$, the credit value adjustment at $t = 0$ under the independence assumption, is given by

$$\text{CVA}^{\text{indep}} = \text{CVA}_0^{\text{indep}} = \delta^S \int_0^T \bar{F}_B(t) D(0,t) E^Q(V_t^+) f_S(t)\,\mathrm{d}t, \tag{17.11}$$

and, by a similar argument, the debt value adjustment under independence is

$$\text{DVA}^{\text{indep}} = \text{DVA}_0^{\text{indep}} = \delta^B \int_0^T \bar{F}_S(t) D(0,t) E^Q(V_t^-) f_B(t)\,\mathrm{d}t. \tag{17.12}$$

Note that formulas (17.11) and (17.12) are much easier to evaluate than the "correct" expressions (17.7) and (17.8). In particular, we only need to determine the marginal distribution of τ_S and τ_B and the so-called *expected exposures* $E^Q(V_t^+)$ and $E^Q(V_t^-)$. On the other hand, the assumption that (V_t), $(Y_{t,S})$ and $(Y_{t,B})$ are independent is often difficult to justify, and the simplified adjustments can be misleading. Consider, for instance, the case where S, R and B are financial institutions and suppose that $T_1 < T$ and $\xi_1 = S$. In that case it is quite likely that τ_S falls in a time period where financial institutions face adverse conditions so that the credit spread of the reference entity at τ_S and, hence, the market value V_{τ_S} of the CDS referencing R are comparatively high. We therefore expect that

$$E^Q(V_{\tau_S}^+ \mid \tau_S = t) > E^Q(V_t^+),$$

so that $\text{CVA} > \text{CVA}^{\text{indep}}$. Similarly, we expect that $E^Q(V_{\tau_B}^- \mid \tau_B = t) < E^Q(V_t^-)$, so that $\text{DVA} < \text{DVA}^{\text{indep}}$. The aggregate effect would be that

$$\text{BCVA} > \text{BCVA}^{\text{indep}}$$

in that case. Some numerical results that support this intuition are given in Section 17.4.4. The phenomenon whereby the conditional expected exposure given the default of the counterparty is higher than the unconditional expected exposure is a typical example of an unfavourable dependence between the size of an exposure and the credit quality of the counterparty. In counterparty risk management, such an unfavourable dependence is known as *wrong-way risk* (since the exposure to a counterparty and the credit quality of that party evolve in the "wrong way"). Wrong-way risk is an important issue in counterparty risk management (see, for example, Chapter 15 of Gregory (2012)).

Unilateral credit value adjustments. In a unilateral value adjustment each party neglects the possibility of its own default. The unilateral value adjustment for the protection buyer B is therefore obtained from the formula for the bilateral value adjustment by assuming that $Q(\xi_1 = B) = 0$. This gives

$$\text{UCVA}_t = E^Q(I_{\{\tau_S < T\}} D(t, \tau_S) \delta^S V_{\tau_S}^+ \mid \mathcal{G}_t).$$

An analogous formula holds for the unilateral value adjustment of the protection seller. Unilateral value adjustments avoid the problem that a worsening credit spread of a financial institution leads to an accounting profit. On the other hand, if B and S use unilateral adjustments, they might come to different conclusions about the value of a given deal.

Netting. A further issue that arises in practice is *netting*. Under a legally enforceable netting agreement the market value of all CDS transactions between B and S at T_1 is computed and only the aggregate value is subject to bankruptcy procedures. In particular, perfectly offsetting transactions cancel each other out. Netting can reduce counterparty risk substantially, so netting agreements are widely used in practice. On the other hand, netting substantially increases the computational complexity of CVA and DVA computations, as we now explain. Suppose that there are N transactions

between B and S that fall under a netting agreement and let these be indexed by $n \in \{1, \dots, N\}$. Denote by $(V_{t,n})$ the market value from the point of view of B of the nth transaction. A similar argument to the one used in the proof of Proposition 17.1 implies that

$$\mathrm{CVA}_t = E^Q\left(I_{\{T_1<T\}} I_{\{\xi_1=S\}} D(t, T_1)\delta^S\left(\sum_{n=1}^{N} V_{T_1,n} \right)^+ \,\middle|\, \mathcal{G}_t \right),$$

and a similar formula applies to the debt value adjustment. Hence in the presence of netting agreements the computation of value adjustments amounts to the pricing of an option on the sum of the market value of all transactions covered by the netting agreement. In the case of CDS contracts each would typically refer to a different reference entity, so we have to consider $n + 2$ different default times. This is in general a much more difficult problem than pricing the options individually.

17.2.2 Collateralized Value Adjustments for a CDS

In this section we introduce popular collateralization strategies and analyse qualitatively the impact of collateralization on credit value adjustments for a CDS. To keep things simple we assume that the collateral is posted in the form of cash and earns the risk-free rate of interest. Many collateralization arrangements used in practice are of this form, but arrangements where other securities are used as collateral can also be found.

Details of the collateralization arrangement for an OTC CDS transaction are fixed in the credit support annex of the transaction. Roughly speaking, the procedure works as follows. At $t_0 = 0$ a collateral account is opened. Let C_t denote the cash balance in the account at time t. Here $C_t > 0$ means that S has posted the collateral and that B is the collateral taker, whereas $C_t < 0$ means that B has posted the collateral and that S is the collateral taker. The collateral position is updated at discrete time points $t_1, \dots, t_N \leqslant T$. At t_1 the collateral taker pays interest on the collateral, and the cash balance C_{t_1} is adjusted in reaction to changes in the price of the underlying CDS over $(t_0, t_1]$. This procedure continues up to the maturity of the CDS or until the first default occurs. If $T_1 > T$ or if $T_1 < T$ and $\xi_1 = R$, the collateral account is closed at the "natural end" of the contract, so $C_t \equiv 0$ for $t \geqslant T_1 \wedge T$. If there is an early default of B or S—that is, if $T_1 \leqslant T$ and $\xi_1 \in \{B, S\}$—the collateral is used to reduce the loss of the collateral taker and any remaining collateral is returned.

An issue arising in this context is *rehypothecation*. The collateral taker typically has unrestricted access to the posted collateral; in particular, the funds can be used as collateral in other OTC derivative transactions. It is therefore possible to have a situation in which the collateral taker defaults and a part of the collateral that should be returned is missing. To keep things simple we ignore this issue and assume that, in the case of a default of the collateral taker, the collateral is always returned in full to the other party. We refer to Notes and Comments for references regarding credit value adjustments in the presence of rehypothecation.

Collateralization strategies. We describe the cash balance in the collateral account by a $(\mathcal{G}_t)$-adapted process (C_t), the so-called collateralization strategy. For convenience we allow for strategies where the collateral account is changed continuously and not just at predetermined rebalancing dates. Recall that the counterparty-risk-free CDS price is denoted by (V_t). In practice, most collateralization arrangements take the form of a threshold collateralization strategy. Formally, a *threshold collateralization strategy* with *thresholds* $M_1, M_2 \geqslant 0$, labelled $(C_t^{M_1,M_2})$ for $0 \leqslant t \leqslant T_1 \wedge T$, is given by

$$C_t^{\gamma,M_1,M_2} = (V_t^+ - M_1)I_{\{V_t^+ > M_1\}} - (V_t^- - M_2)I_{\{V_t^- > M_2\}}. \tag{17.13}$$

Under this strategy collateral is posted if V_t^+ (the exposure of B) exceeds the threshold M_1 or if V_t^- (the exposure of S) exceeds the threshold M_2. A threshold strategy is used if B and S want to protect themselves against severe counterparty-risk-related losses, while accepting the possibility of smaller losses in order to simplify the practical management of the collateralization process. For $M_1 = M_2 = 0$ we obtain the special case of *market-value collateralization* with

$$C_t^{\text{market}} = V_t, \quad 0 \leqslant t \leqslant T_1 \wedge T. \tag{17.14}$$

Collateralized value adjustment. Value adjustments for collateralized CDS contracts are largely analogous to the uncollateralized case, so we keep our presentation short. The bilateral collateralized value adjustment BCCVA is the difference between the collateralized credit value adjustment CCVA and the collateralized debt value adjustment CDVA. As before, the CCVA gives the value of the potential loss for B due to an early default of S, whereas the CDVA gives the value of the potential loss for S due to an early default of B.

In order to describe these potential losses we have to consider the payments at an early default. Note that no additional collateral is posted at or after the default of B or S. The amount of collateral available for the settlement of the contract is therefore given by C_{T_1-} (the amount of collateral that has been posted immediately prior to T_1). This distinction matters if the close-out amount (V_t) jumps at T_1, for instance due to contagion effects, or if there is some delay between the last adjustment of the collateral account and the settlement of the positions. We begin with the scenario where the protection seller defaults first. We have to distinguish the cases $V_{T_1} > 0$ and $V_{T_1} < 0$.

- Suppose that $V_{T_1} \geqslant 0$ and that the protection buyer is the collateral taker, that is $C_{T_1-} \geqslant 0$. In that case the collateral is used to reduce the loss of the protection buyer and excess collateral is returned. If C_{T_1-} is smaller than V_{T_1}, the protection buyer incurs a counterparty-risk-related loss of size $\delta^S(V_{T_1} - C_{T_1-})$; if $C_{T_1-} \geqslant V_{T_1}$, the amount of collateral is sufficient to protect B from losses due to counterparty risk. If S is the collateral taker, i.e. if $C_{T_1-} \leqslant 0$, there is no available collateral to protect B and he suffers a loss of size $\delta^S V_{T_1}$.

- Suppose that $V_{T_1} \leqslant 0$. In that case B has no exposure to S, so he does not suffer a loss related to counterparty risk (the fact that he has incurred a loss due to the decrease in the counterparty-risk-free CDS price is irrelevant for the computation of value adjustments for counterparty risk).

Summarizing, the counterparty-risk-related loss of B is given by

$$I_{\{T_1<T\}} I_{\{\xi_1=S\}} \delta^S (V_{T_1}^+ - C_{T_1-}^+)^+.$$

Similarly, S suffers a loss in the scenario where $V_{T_1} \leqslant 0$ and where there is insufficient collateral to settle the contract in full, that is, for $V_{T_1}^- > C_{T_1}^-$. The counterparty-risk-related loss of S is thus given by

$$I_{\{T_1<T\}} I_{\{\xi_1=B\}} \delta^B (V_{T_1}^- - C_{T_1-}^-)^+.$$

Thus BCCVA_t, the *bilateral collateralized credit value adjustment* at time t, is given by

$$\mathrm{BCCVA}_t = \mathrm{CCVA}_t - \mathrm{CDVA}_t, \tag{17.15}$$

where

$$\begin{aligned}
\mathrm{CCVA}_t &:= E(I_{\{T_1<T\}} I_{\{\xi_1=S\}} D(t,T_1) \delta^S (V_{T_1}^+ - C_{T_1-}^+)^+ \mid \mathcal{G}_t), \\
\mathrm{CDVA}_t &:= E(I_{\{T_1<T\}} I_{\{\xi_1=B\}} D(t,T_1) \delta^B (V_{T_1}^- - C_{T_1-}^-)^+ \mid \mathcal{G}_t).
\end{aligned}$$

Without collateralization, i.e. for $C_t \equiv 0$, formula (17.15) reduces to the simpler result of Proposition 17.1.

Performance of market-value collateralization. The sum $\mathrm{CCVA}_t + \mathrm{CDVA}_t$ gives the value in t of the entire counterparty-risk-related loss, and it can therefore be viewed as a measure of the performance of a given collateralization strategy. Here we make the following immediate observation: suppose that market-value collateralization with $C_t^{\mathrm{market}} = V_t$ is used and that the market value of the CDS does not jump at T_1, that is, $V_{T_1} = V_{T_1-}$ almost surely. In that case the formulas for CCVA_t and CDVA_t in (17.15) show that

$$\mathrm{CCVA}_t = \mathrm{CDVA}_t = 0, \quad t \leqslant T_1,$$

so that market-value collateralization works perfectly. If, on the other hand, $|\Delta V_{T_1}| = |V_{T_1} - V_{T_1-}|$ is comparatively large, the performance of market-value collateralization will be not so good. Some numerical results supporting this observation will be presented in Section 17.4.4.

Notes and Comments

The literature on counterparty risk management is growing rapidly, leading to a proliferation of valuation-adjustment acronyms (CVA, DVA, FVA and others). A detailed introduction can be found in the textbooks by Gregory (2012), Cesari et al. (2009) and Brigo, Morini and Pallavicini (2013). A non-technical introduction to the computation of value adjustments for counterparty risk is given in the papers by Hull and White (2012, 2013).

The derivation of the bilateral credit value adjustments in Propositions 17.1 and (17.15) is based on the papers by Brigo and Chourdakis (2009) and Brigo, Capponi and Pallavicini (2014) (see also Frey and Rösler 2014). The last two papers also consider the case of rehypothecation and discuss the actual computation of value adjustments in various portfolio credit risk models. Credit value adjustments in structural credit risk models are studied in Lipton and Sepp (2009). A very general technical analysis of value adjustments is given in Crépey (2012a,b).

17.3 Conditionally Independent Default Times

In this section we discuss models with conditionally independent default times. We begin with a discussion of general mathematical properties; applications and specific examples from the literature are considered in Sections 17.3.2 and 17.3.3.

17.3.1 Definition and Mathematical Properties

Throughout we consider a portfolio of m obligors with default times τ_i and default indicators $Y_{t,i} = I_{\{\tau_i \leqslant t\}}$, $1 \leqslant i \leqslant m$, on a probability space $(\Omega, \mathcal{F}, P)$. The ordered default times are denoted by $0 = T_0 < T_1 < \cdots < T_m$, and $\xi_n \in \{1, \ldots, m\}$ gives the identity of the firm defaulting at time T_n. We introduce the filtrations $(\mathcal{H}_t^i)$, $1 \leqslant i \leqslant m$, and $(\mathcal{H}_t)$ defined by

$$\mathcal{H}_t^i = \sigma(\{Y_{s,i} : s \leqslant t\}) \quad \text{and} \quad \mathcal{H}_t = \mathcal{H}_t^i \vee \cdots \vee \mathcal{H}_t^m. \tag{17.16}$$

$(\mathcal{H}_t^i)$ is the filtration generated by the default observation for obligor i alone; $(\mathcal{H}_t)$ is the filtration generated by default observations for all obligors. Often $(\mathcal{H}_t)$ is called the *default history* of the portfolio or the *internal filtration* generated by the default times $\tau_1, \ldots, \tau_m$. The definition of conditionally independent default times is a straightforward multivariate extension of the notion of doubly stochastic default times from Section 10.5.1. In particular, the distribution of the default times is affected by additional information on top of the default history $(\mathcal{H}_t)$. Formally, we represent this information by a filtration $(\mathcal{F}_t)$ on the underlying probability space. Typically, $(\mathcal{F}_t)$ is generated by some observable background process. The information available to investors is given by the filtration $(\mathcal{G}_t) = (\mathcal{F}_t) \vee (\mathcal{H}_t)$ (see also (10.46)).

Definition 17.2. The default times $\tau_1, \ldots, \tau_m$ are conditionally independent doubly stochastic random times if there are positive, $(\mathcal{F}_t)$-adapted processes $(\gamma_{t,i})$, $1 \leqslant i \leqslant m$, with $\Gamma_{t,i} = \int_0^t \gamma_{s,i}\, \mathrm{d}s$ strictly increasing and finite for every $t > 0$, such that

$$P(\tau_1 > t_1, \ldots, \tau_m > t_m \mid \mathcal{F}_\infty) = \prod_{i=1}^m \exp\left(-\int_0^{t_i} \gamma_{s,i}\, \mathrm{d}s\right). \tag{17.17}$$

Note that the definition implies that each of the τ_i is a doubly stochastic random time with conditional hazard process $(\gamma_{t,i})$ in the sense of Definition 10.10 and that the rvs $\tau_1, \ldots, \tau_m$ are conditionally independent given $\mathcal{F}_\infty$.

Construction and simulation via thresholds. Lemma 17.3 extends Lemma 10.11.

Lemma 17.3. *Let $(\gamma_{t,1}), \dots, (\gamma_{t,m})$ be positive, $(\mathcal{F}_t)$-adapted processes such that $\Gamma_{t,i} := \int_0^t \gamma_{s,i}\,\mathrm{d}s$ is strictly increasing and finite for any $t > 0$. Let $\boldsymbol{X} = (X_1, \dots, X_m)'$ be a vector of independent, standard exponentially distributed rvs independent of $\mathcal{F}_\infty$. Define*

$$\tau_i = \Gamma_i^{\leftarrow}(X_i) = \inf\{t \geqslant 0 \colon \Gamma_{t,i} \geqslant X_i\}.$$

Then $\tau_1, \dots, \tau_m$ are conditionally independent doubly stochastic random times with hazard processes $(\gamma_{t,i})$, $1 \leqslant i \leqslant m$.

Proof. By the definition of τ_i we have $\tau_i > t \iff X_i > \Gamma_{t,i}$. The rvs $\Gamma_{t_i,i}$ are now measurable with respect to $\mathcal{F}_\infty$, whereas the X_i are mutually independent, independent of $\mathcal{F}_\infty$ and standard exponentially distributed. We therefore infer that

$$\begin{aligned} P(\tau_1 > t_1, \dots, \tau_m > t_m \mid \mathcal{F}_\infty) &= P(X_1 > \Gamma_{t_1,1}, \dots, X_m > \Gamma_{t_m,m} \mid \mathcal{F}_\infty) \\ &= \prod_{i=1}^m P(X_i > \Gamma_{t_i,i} \mid \mathcal{F}_\infty) \\ &= \prod_{i=1}^m \mathrm{e}^{-\Gamma_{t_i,i}}, \end{aligned} \tag{17.18}$$

which shows that the τ_i satisfy the conditions of Definition 17.2. □

Lemma 17.3 is the basis for the following simulation algorithm.

Algorithm 17.4 (multivariate threshold simulation).

(1) Generate trajectories for the hazard processes $(\gamma_{t,i})$ for $i = 1, \dots, m$. The same techniques as in the univariate case can be used here.

(2) Generate a vector $\boldsymbol{X}$ of independent standard exponentially distributed rvs (the threshold vector) and set $\tau_i = \Gamma_i^{\leftarrow}(X_i)$, $1 \leqslant i \leqslant m$.

As in the univariate case (see Lemma 10.12), Lemma 17.3 has a converse.

Lemma 17.5. *Let $\tau_1, \dots, \tau_m$ be conditionally independent doubly stochastic random times with $(\mathcal{F}_t)$-conditional hazard processes $(\gamma_{t,i})$. Define a random vector $\boldsymbol{X}$ by setting $X_i = \Gamma_i(\tau_i)$, $1 \leqslant i \leqslant m$. Then $\boldsymbol{X}$ is a vector of independent, standard exponentially distributed rvs that is independent of $\mathcal{F}_\infty$, and $\tau_i = \Gamma_i^{\leftarrow}(X_i)$ almost surely.*

Proof. For $t_1, \dots, t_m \geqslant 0$ the conditional independence of the τ_i implies that

$$P(\Gamma_1(\tau_1) \leqslant t_1, \dots, \Gamma_m(\tau_m) \leqslant t_m \mid \mathcal{F}_\infty) = \prod_{i=1}^m P(\Gamma_i(\tau_i) \leqslant t_i \mid \mathcal{F}_\infty).$$

Moreover, similar reasoning to the univariate case implies that

$$P(\Gamma_i(\tau_i) \leqslant t_i \mid \mathcal{F}_\infty) = P(\tau_i \leqslant \Gamma_i^{\leftarrow}(t_i) \mid \mathcal{F}_\infty) = 1 - \mathrm{e}^{-t_i},$$

which proves that $\boldsymbol{X}$ has the claimed properties. □

Recursive default time simulation. We now describe a second recursive algorithm for simulating conditionally independent default times, which can be more efficient than multivariate threshold simulation. We need the following lemma, which gives properties of the first default time T_1.

Lemma 17.6. *Let $\tau_1, \dots, \tau_m$ be conditionally independent doubly stochastic random times with hazard processes $(\gamma_{t,1}), \dots, (\gamma_{t,m})$. Then T_1 is a doubly stochastic random time with $(\mathcal{F}_t)$-conditional hazard process $\bar{\gamma}_t := \sum_{i=1}^m \gamma_{t,i}$, $t \geqslant 0$.*

Proof. Using the conditional independence of the τ_i we infer that

$$P(T_1 > t \mid \mathcal{F}_\infty) = P(\tau_1 > t, \dots, \tau_m > t \mid \mathcal{F}_\infty) = \prod_{i=1}^m \exp\left(-\int_0^t \gamma_{s,i}\,\mathrm{d}s\right),$$

which is obviously equal to $\exp(-\int_0^t \bar{\gamma}_s\,\mathrm{d}s)$. As this expression is $\mathcal{F}_t$-measurable, the result follows. □

Next we compute the conditional probability of the event $\{\xi_1 = i\}$ given the time T_1 of the first default and full information about the background filtration.

Proposition 17.7. *Under the assumptions of Lemma 17.6 we have*

$$P(\xi_1 = i \mid \mathcal{F}_\infty \vee \sigma(T_1)) = \gamma_i(T_1)/\bar{\gamma}(T_1), \quad i \in \{1, \dots, m\}.$$

Proof. Conditional on $\mathcal{F}_\infty$ the τ_i are independent with deterministic hazard functions $\gamma_i(t)$, so it is sufficient to prove the proposition for independent random times with deterministic hazard functions. Fix some $t > 0$ and note that, since the random vector $(\tau_1, \dots, \tau_m)$ has a joint density, the probability of having more than one default in the interval $(t-h, t]$ is of order $o(h)$. Hence,

$$\begin{aligned} P(\{\xi_1 = i\} \cap \{T_1 \in (t-h, t]\}) & \\ &= P(\{\tau_i \in (t-h, t]\} \cap \{\tau_j > t \text{ for all } j \neq i\}) + o(h) \\ &= P(\tau_i \in (t-h, t]) \prod_{j \neq i} P(\tau_j > t) + o(h) \end{aligned}$$

by the independence of the τ_i. Since $P(\tau_i > t) = \exp(-\int_0^t \gamma_i(s)\,\mathrm{d}s)$, $1 \leqslant i \leqslant m$, the above expression equates to

$$\exp\left(-\int_0^{t-h} \gamma_i(s)\,\mathrm{d}s\right)\left(1 - \exp\left(-\int_{t-h}^t \gamma_i(s)\,\mathrm{d}s\right)\right) \times \prod_{j \neq i} \exp\left(-\int_0^t \gamma_j(s)\,\mathrm{d}s\right) + o(h).$$

Hence,

$$\lim_{h \to 0+} \frac{1}{h} P(\{\xi_1 = i\} \cap \{T_1 \in (t-h, t]\}) = \gamma_i(t) \exp\left(-\int_0^t \bar{\gamma}(s)\,\mathrm{d}s\right).$$

Moreover, by Lemma 17.6,

$$\lim_{h\to 0+} \frac{1}{h} P(T_1 \in (t-h, t]) = \bar{\gamma}(t) \exp\left(-\int_0^t \bar{\gamma}(s)\,\mathrm{d}s\right),$$

so the claim follows from the definition of elementary conditional probability. □

Algorithm 17.8 (recursive default time simulation). This algorithm simulates a realization of the sequence (T_n, ξ_n) up to some maturity date T. In order to formulate the algorithm we introduce the notation A_n for the set of non-defaulted firms immediately after T_n. Formally, we set

$$A_0 := \{1, \dots, m\}, \quad A_n := \{1 \leqslant i \leqslant m \colon Y_i(T_n) = 0\}, \quad n \geqslant 1. \tag{17.19}$$

Moreover we define, for $n \geqslant 0$,

$$\bar{\gamma}_t^{(n)} := \sum_{i \in A_n} \gamma_{t,i}, \quad 0 \leqslant n \leqslant m.$$

The algorithm proceeds in the following steps.

(1) Generate trajectories of the hazard processes $(\gamma_{t,i})$ up to time T and set $n = 0$, $T_0 = 0$ and $\xi_0 = 0$.

(2) Generate the waiting time $T_{n+1} - T_n$ by standard univariate threshold simulation using Algorithm 10.13. For this we use a generalization of Lemma 17.6 that states that for conditionally independent defaults and $n < m$ we have

$$P(T_{n+1} - T_n > t \mid (T_1, \xi_1), \dots, (T_n, \xi_n), \mathcal{F}_\infty) = \exp\left(-\int_{T_n}^{T_n+t} \bar{\gamma}_s^{(n)}\,\mathrm{d}s\right).$$

(3) If $T_{n+1} \geqslant T$, stop. Otherwise use Proposition 17.7 and determine ξ_{n+1} as a realization of a multinomial rv ξ with

$$P(\xi = i) = \frac{\gamma_i(T_{n+1})}{\bar{\gamma}^{(n)}(T_{n+1})}, \quad i \in A_n.$$

(4) If $n + 1 = m$, stop. Otherwise increase n by 1 and return to step (2).

Recursive default time simulation is particularly efficient if we only need to simulate defaults occurring before some maturity date T and if defaults are rare. In that case, $T_n > T$ for n relatively small, so only a few ordered default times need to be simulated. With multivariate threshold simulation, on the other hand, we need to simulate the default times of all firms in the portfolio.

Default intensities. We begin with a general definition of default intensities in reduced-form portfolio credit risk models.

Definition 17.9 (default intensity). Consider a generic filtration $(\mathcal{G}_t)$ such that the default indicator process $(Y_{t,i})$ is $(\mathcal{G}_t)$-adapted. Then a non-negative $(\mathcal{G}_t)$-adapted

right-continuous process $(\lambda_{t,i})$ with $\int_0^t \lambda_{s,i}\,\mathrm{d}s < \infty$ almost surely for all t is called the $(\mathcal{G}_t)$ default intensity of firm i if the process

$$M_{t,i} := Y_{t,i} - \int_0^{t\wedge\tau_i} \lambda_{s,i}\,\mathrm{d}s = Y_{t,i} - \int_0^t (1 - Y_{s,i})\lambda_{s,i}\,\mathrm{d}s$$

is a $(\mathcal{G}_t)$-martingale.

The filtration $(\mathcal{G}_t)$ is an integral part of the definition of $(\lambda_{t,i})$. If the choice of the underlying filtration is clear from the context, we will simply speak of the default intensity of firm i. It is well known from stochastic calculus that the compensator of $(Y_{t,i})$ (the continuous, $(\mathcal{G}_t)$-adapted process $(\Lambda_{t,i})$ such that $Y_{t,i} - \Lambda_{t,i}$ is a $(\mathcal{G}_t)$-martingale) is unique. Since we assumed that (λ_t) is right continuous, it follows that the default intensity $\lambda_{t,i}$ is uniquely defined for $t < \tau_i$.

In the next result we link Definition 17.9 to the informal interpretation of default intensities as instantaneous default probabilities.

Lemma 17.10. *Let $(\lambda_{t,i})$ be the $(\mathcal{G}_t)$ default intensity of firm i and suppose that the process $(\lambda_{t,i})$ is bounded. We then have the equality*

$$I_{\{\tau_i>t\}}\lambda_{t,i} = I_{\{\tau_i>t\}} \lim_{h\to 0^+} \frac{1}{h} P(\tau_i \leqslant t + h \mid \mathcal{G}_t).$$

Proof. Since the firm under consideration is fixed we will simply write τ, Y_t and λ_t for the default time, the default indicator and the default intensity. By the definition of Y_t we have

$$P(\tau \leqslant t + h \mid \mathcal{G}_t) = E(Y_{t+h} \mid \mathcal{G}_t) = Y_t + E(Y_{t+h} - Y_t \mid \mathcal{G}_t).$$

Since $Y_t - \int_0^t (1 - Y_s)\lambda_s\,\mathrm{d}s$ is a $(\mathcal{G}_t)$-martingale, it follows that $E(Y_{t+h} - Y_t \mid \mathcal{G}_t) = E(\int_t^{t+h}(1 - Y_s)\lambda_s\,\mathrm{d}s \mid \mathcal{G}_t)$, and therefore, since $I_{\{\tau>t\}}Y_t = 0$,

$$I_{\{\tau>t\}}P(\tau \leqslant t + h \mid \mathcal{G}_t) = I_{\{\tau>t\}}E\left(\int_t^{t+h} (1 - Y_s)\lambda_s\,\mathrm{d}s \;\middle|\; \mathcal{G}_t \right). \tag{17.20}$$

The right-continuity of the process $((1 - Y_s)\lambda_s)$ implies that

$$\lim_{h\to 0^+} \frac{1}{h} \int_t^{t+h} (1 - Y_s)\lambda_s\,\mathrm{d}s = (1 - Y_t)\lambda_t \quad \text{a.s.}$$

Recalling that (λ_t) is bounded by assumption, we can use (17.20) and the bounded convergence theorem for conditional expectations to deduce that

$$I_{\{\tau>t\}} \lim_{h\to 0^+} \frac{1}{h} P(\tau \leqslant t + h \mid \mathcal{G}_t) = I_{\{\tau>t\}}(1 - Y_t)\lambda_t = I_{\{\tau>t\}}\lambda_t,$$

which proves the claim. □

Remark 17.11. The converse statement is also true; if the derivative

$$\lim_{h\to 0^+} \frac{1}{h} P(\tau \leqslant t + h \mid \mathcal{G}_t)$$

exists almost surely, then τ admits an intensity (under some technical conditions); references are given in Notes and Comments.

Finally, we return to the special case of models with conditionally independent defaults as introduced in Definition 17.2. The following proposition shows that in this model class a default intensity of any given firm i is given by the conditional hazard process of the default time τ_i.

Proposition 17.12. *Let $\tau_1, \dots, \tau_m$ be conditionally independent doubly stochastic random times with $(\mathcal{F}_t)$-conditional hazard processes $(\gamma_{t,1}), \dots, (\gamma_{t,m})$. The processes*

$$M_{t,i} := Y_{t,i} - \int_0^{t\wedge\tau_i} \gamma_{s,i}\, \mathrm{d}s$$

are then $(\mathcal{G}_t)$-martingales.

Proof. The result follows immediately from Proposition 10.14 (which gives the compensator of a doubly stochastic default time in the single-firm case) if we can show that τ_i is a doubly stochastic default time with $(\mathcal{G}_t^{-i})$-conditional hazard process $(\gamma_{t,i})$, where $(\mathcal{G}_t^{-i})$ is the artificial background filtration defined by

$$\mathcal{G}_t^{-i} = \mathcal{F}_t \vee \mathcal{H}_t^1 \cdots \vee \mathcal{H}_t^{i-1} \vee \mathcal{H}_t^{i+1} \vee \cdots \vee \mathcal{H}_t^m, \quad t \geqslant 0. \tag{17.21}$$

According to Definition 10.10 we have to show that

$$P(\tau_i > t \mid \mathcal{G}_\infty^{-i}) = \exp\left(-\int_0^t \gamma_{s,i}\, \mathrm{d}s\right), \quad t \geqslant 0. \tag{17.22}$$

This relationship is quite intuitive; since the τ_i are independent given $\mathcal{F}_\infty$, the default history of obligor $j \neq i$ that is contained in $\mathcal{G}_\infty^{-i}$ but not in $\mathcal{F}_\infty$ has no impact on the conditional default probability of obligor i.

A formal argument is as follows. Using Lemma 17.5, we may assume that there is a vector $\boldsymbol{X}$ of independent, standard exponential rvs, independent of $\mathcal{F}_\infty$, such that for all $1 \leqslant j \leqslant m$ we have $\tau_j = \Gamma_j^{\leftarrow}(X_j)$. Obviously, τ_i is independent of X_j for $j \neq i$, so

$$P(\tau_i > t \mid \mathcal{F}_\infty \vee \sigma(\{X_j : j \neq i\}) = P(\tau_i > t \mid \mathcal{F}_\infty) = \exp\left(-\int_0^t \gamma_{s,i}\, \mathrm{d}s\right). \tag{17.23}$$

On the other hand, if we know X_j and the trajectory $(\gamma_{t,j})_{0\leqslant t\leqslant\infty}$, we can determine the trajectory $(Y_{t,j})_{0\leqslant t\leqslant\infty}$, so $\mathcal{G}_\infty^{-i} \subseteq \mathcal{F}_\infty \vee \sigma(\{X_j : j \neq i\})$ (in fact, the two σ-algebras are equal). Since the right-hand side of (17.23) is measurable with respect to $\mathcal{G}_\infty^{-i}$, equation (17.22) follows from (17.23). □

Remark 17.13 (redundant default information). Suppose that $\tau_1, \dots, \tau_m$ are conditionally independent doubly stochastic random times. Consider a financial product with maturity T whose discounted pay-off H depends only on the evolution of default-free security prices (which we assume to be $(\mathcal{F}_t)$-adapted) and on the default history of a subset $A = \{i_1, \dots, i_k\}$ of the firms in the portfolio and is therefore measurable with respect to the σ-algebra

$$\mathcal{G}_T^A := \mathcal{F}_T \vee \mathcal{H}_T^{i_1} \vee \cdots \vee \mathcal{H}_T^{i_k} \subset \mathcal{G}_T.$$

A typical example would be the default and the premium payment leg of a single-name CDS on a given firm i. In this context an argument similar to the one used to establish (17.22) above shows that

$$E(H \mid \mathcal{G}_t) = E(H \mid \mathcal{G}_t^A);$$

in particular, the default history of the firms that do not belong to A is redundant for computing the price of H. Such a relationship does not hold for more general portfolio credit risk models where the default times are not conditionally independent.

17.3.2 Examples and Applications

In many models from the financial literature with conditionally independent defaults, the hazard rates are modelled as linear combinations of independent affine diffusions, possibly with jumps. A typical model takes the form

$$\gamma_{t,i} = \gamma_{i0} + \sum_{j=1}^{p} \gamma_{ij} \Psi_{t,j}^{\text{syst}} + \Psi_{t,i}^{\text{id}}, \quad 1 \leqslant i \leqslant m, \tag{17.24}$$

where $(\Psi_{t,j}^{\text{syst}})$, $1 \leqslant j \leqslant p$, and $(\Psi_{t,i}^{\text{id}})$, $1 \leqslant i \leqslant m$, are independent CIR square-root diffusions as in (10.68) or, more generally, basic affine jump diffusions as in (10.75); the factor weights γ_{ij}, $0 \leqslant j \leqslant p$, are non-negative constants.

Writing $\boldsymbol{\Psi}_t^{\text{syst}} = (\Psi_{t,1}^{\text{syst}}, \ldots, \Psi_{t,p}^{\text{syst}})'$, it is obvious that $(\boldsymbol{\Psi}_t^{\text{syst}})$ represents common or systematic factors, while the $(\Psi_{t,i}^{\text{id}})$ processes represent idiosyncratic factors affecting only the hazard rate of obligor i. Note that the weight attached to the idiosyncratic factor can be incorporated into the parameters of the dynamics of $(\Psi_{t,i}^{\text{id}})$ and does not need to feature explicitly. Throughout this section we assume that the background filtration is generated by $(\boldsymbol{\Psi}_t^{\text{syst}})$ and $(\Psi_{t,i}^{\text{id}})$, $1 \leqslant i \leqslant m$. In practical applications of the model, the current value of these processes is derived from observed prices of defaultable bonds.

We now consider some examples proposed in the literature. Duffee (1999) has estimated a model of the form (17.24) with $p = 2$; in his model all factor processes are assumed to follow CIR square-root diffusions, so that their dynamics are characterized by the parameter triplet $(\kappa, \bar{\theta}, \sigma)$ (see equation (10.68)). In Duffee's model, $(\boldsymbol{\Psi}_t^{\text{syst}})$ represents factors driving the default-free short rate; the parameters of these processes are estimated from treasury data. The factor weights γ_{ij} and the parameters of (Ψ_t^{id}), on the other hand, are estimated from corporate bond prices.

In their influential case study on CDO pricing, Duffie and Gârleanu (2001) use basic affine jump diffusion processes of the form (10.75) to model the factors driving the hazard rates. Jumps in (γ_t) represent shocks that increase the default probability of a firm. They consider a homogeneous model with one systematic factor where $\gamma_{t,i} = \Psi_t^{\text{syst}} + \Psi_{t,i}^{\text{id}}$, $1 \leqslant i \leqslant m$, and they assume that the speed of mean reversion κ, the volatility σ and the mean jump size μ are identical for (Ψ_t^{syst}) and $(\Psi_{t,i}^{\text{id}})$. It is straightforward to show that this implies that the sum $\gamma_{t,i} = \Psi_t^{\text{syst}} + \Psi_{t,i}^{\text{id}}$ also follows a basic affine jump diffusion with parameters κ, $\bar{\theta}^{\text{syst}} + \bar{\theta}^{\text{id}}$, σ, $(l^0)^{\text{syst}} + (l^0)^{\text{id}}$ and μ; the parameters of $(\gamma_{t,i})$ used in Duffie and Gârleanu (2001) can be found in the row labelled "base case" in Table 17.1.

Table 17.1. Parameter sets for the Duffie–Gârleanu model used in Figure 17.1.

Parameter set	κ	$\bar{\theta}$	σ	l^0	μ
Pure diffusion	0.6	0.0505	0.141	0	0
Base case	0.6	0.02	0.141	0.2	0.1
High jump intensity	0.6	0.0018	0.141	0.32	0.1

Pricing single-name credit products. As discussed in Remark 17.13, in the framework of conditionally independent defaults the pricing formulas obtained in a single-firm model remain valid in the portfolio context. Moreover, when the hazard processes are specified as in (17.24), most computations can be reduced to one-dimensional problems involving affine processes, to which the results of Section 10.6 apply. As a simple example we consider the computation of the conditional survival probability of obligor i. By Remark 17.13 and Theorem 10.19 it follows that

$$P(\tau_i > T \mid \mathcal{G}_t) = P(\tau_i > T \mid \mathcal{G}_t^i) = I_{\{\tau_i > t\}} E\left(\exp\left(-\int_t^T \gamma_{s,i} \, \mathrm{d}s \right) \,\middle|\, \mathcal{F}_t \right).$$

For hazard processes of the form (17.24) this equals

$$I_{\{\tau_i > t\}} \mathrm{e}^{-\gamma_{i0}(T-t)} E\left(\exp\left(-\int_t^T \Psi_{s,i}^{\mathrm{id}} \, \mathrm{d}s \right) \,\middle|\, \mathcal{F}_t \right) \times \prod_{j=1}^{p} E\left(\exp\left(-\int_t^T \Psi_{s,j}^{\mathrm{syst}} \, \mathrm{d}s \right) \,\middle|\, \mathcal{F}_t \right). \quad (17.25)$$

Each of the conditional expectations in (17.25) can now be computed using the results for one-dimensional affine models from Section 10.6. More general models, where hazard rates are given by a general multivariate affine process (and not simply by a linear combination of independent one-dimensional affine processes) can be dealt with using the general affine-model technology developed by Duffie, Pan and Singleton (2000).

The implied one-period model and computation of the loss distribution. The conditional independence assumption and the factor structure (17.24) of the hazard processes have interesting implications for the distribution of the default indicators at a fixed time point.

For simplicity we suppose that there are no idiosyncratic factors $(\Psi_{t,i}^{\mathrm{id}})$ in the model. We fix $T > 0$ and consider the random vector $\boldsymbol{Y}_T = (Y_{T,1}, \ldots, Y_{T,m})'$. For $\boldsymbol{y} \in \{0, 1\}^m$ we can compute that

$$\begin{aligned} P(\boldsymbol{Y}_T = \boldsymbol{y}) &= E(P(\boldsymbol{Y}_T = \boldsymbol{y} \mid \mathcal{F}_\infty)) \\ &= E\Bigg(\prod_{j:\, y_j = 1} P(\tau_j \leqslant T \mid \mathcal{F}_\infty) \prod_{j:\, y_j = 0} P(\tau_j > T \mid \mathcal{F}_\infty) \Bigg). \end{aligned}$$

From (17.24) we also know that

$$P(\tau_i \leqslant T \mid \mathcal{F}_\infty) = 1 - \exp\left(-\gamma_{i0}T - \sum_{j=1}^{p} \gamma_{ij} \int_0^T \Psi_{s,j}^{\text{syst}}\, \mathrm{d}s\right). \tag{17.26}$$

This argument shows that $\boldsymbol{Y}_T$ follows a Bernoulli mixture model with p-factor structure as in Definition 11.5. The factor vector is given by

$$\boldsymbol{\Psi} := \left(\int_0^T \Psi_{s,1}^{\text{syst}}\, \mathrm{d}s, \ldots, \int_0^T \Psi_{s,p}^{\text{syst}}\, \mathrm{d}s\right)'$$

and the conditional default probabilities $p_i(\boldsymbol{\Psi})$ are given in (17.26).

For practical purposes, such as CDO pricing, we need to be able to evaluate the distribution of the portfolio loss $L_T = \sum_{i=1}^m \delta_i Y_{T,i}$ in this model. As explained in Section 11.2.1 it is useful to be able to compute the Laplace–Stieltjes transform $\hat{F}_{L_T}(\lambda) = E(\mathrm{e}^{-\lambda L_T})$ since the distribution of L_T can then be determined using inversion techniques for transforms.

In Section 11.2.5 we noted that it is quite common in practice to approximate Bernoulli mixture models with Poisson mixture models, which are often more tractable. We will derive the Laplace–Stieltjes transform for an approximating Poisson mixture model and see that it consists of terms that can computed using results for affine processes in Section 10.6.3. More precisely, we replace L_T by the rv $\tilde{L}_T := \sum_{i=1}^m \delta_i \tilde{Y}_{T,i}$, and we assume that, conditional on $\mathcal{F}_\infty$, the rvs $\tilde{Y}_{T,1}, \ldots, \tilde{Y}_{T,m}$ are independent, Poisson-distributed rvs with parameters

$$\Gamma_{T,i} = \gamma_{i0}T + \sum_{j=1}^{p} \gamma_{ij} \int_0^T \Psi_{s,j}^{\text{syst}}\, \mathrm{d}s, \quad 1 \leqslant i \leqslant m. \tag{17.27}$$

We also assume that the losses given default $\delta_1, \ldots, \delta_m$ are deterministic and we use the fact that the Laplace–Stieltjes transform of a generic Poisson rv $N \sim \mathrm{Poi}(\Gamma)$ is given by $\hat{F}_N(\lambda) = E(\mathrm{e}^{-\lambda N}) = \exp(\Gamma(\mathrm{e}^{-\lambda} - 1))$. Using the conditional independence of $\tilde{Y}_{T,1}, \ldots, \tilde{Y}_{T,m}$ it follows that

$$\begin{aligned} E(\mathrm{e}^{-\lambda \tilde{L}_T} \mid \mathcal{F}_\infty) &= \prod_{i=1}^m E(\mathrm{e}^{-\lambda\delta_i \tilde{Y}_{T,i}} \mid \mathcal{F}_\infty) = \prod_{i=1}^m \exp(\Gamma_{T,i}(\mathrm{e}^{-\lambda\delta_i} - 1)) \\ &= \exp\left(\sum_{i=1}^m (\mathrm{e}^{-\lambda\delta_i} - 1)\Gamma_{T,i}\right). \end{aligned}$$

Using the definition of $\Gamma_{T,i}$ in (17.27) and the independence of the systematic risk-factor processes $(\Psi_{t,1}^{\text{syst}}), \ldots, (\Psi_{t,p}^{\text{syst}})$, we obtain

$$\begin{aligned} \hat{F}_{\tilde{L}_T}(\lambda) &= E(E(\mathrm{e}^{-\lambda \tilde{L}_T} \mid \mathcal{F}_\infty)) \\ &= \exp\left\{\sum_{i=1}^m (\mathrm{e}^{-\lambda\delta_i} - 1)\gamma_{i0}T\right\} \\ &\quad \times \prod_{j=1}^p E\left(\exp\left\{\left(\sum_{i=1}^m (\mathrm{e}^{-\lambda\delta_i} - 1)\gamma_{ij}\right)\int_0^T \Psi_{s,j}^{\text{syst}}\, \mathrm{d}s\right\}\right). \end{aligned} \tag{17.28}$$

Expression (17.28) can be computed using the results for one-dimensional affine models in Section 10.6.3. For further refinements and references to Laplace inversion methods we refer to Di Graziano and Rogers (2009).

Default correlation. As we have seen in Chapter 10, default correlations (defined as the correlations $\rho(Y_{T,i}, Y_{T,j})$, $i \neq j$, between default indicators) are closely related to the heaviness of the tail of the credit loss distribution. In computing default correlations in models with conditionally independent defaults it is more convenient to work with the *survival indicator* $1 - Y_{T,i}$. By the definition of linear correlation we have

$$\begin{aligned}\rho(Y_{T,i}, Y_{T,j}) &= \rho(1 - Y_{T,i}, 1 - Y_{T,j}) \\ &= \frac{P(\tau_i > T, \tau_j > T) - \bar{F}_i(T)\bar{F}_j(T)}{(\bar{F}_i(T)(1 - \bar{F}_i(T)))^{1/2}(\bar{F}_j(T)(1 - \bar{F}_j(T)))^{1/2}}. \end{aligned} \tag{17.29}$$

For models with hazard rates as in (17.24), the computation of the survival probabilities $\bar{F}_i(T)$ has been discussed above. For the joint survival probability we obtain, using conditional independence,

$$\begin{aligned}P(\tau_i > T, \tau_j > T) &= E(P(\tau_i > T, \tau_j > T \mid \mathcal{F}_\infty)) \\ &= E(P(\tau_i > T \mid \mathcal{F}_\infty)P(\tau_j > T \mid \mathcal{F}_\infty)) \\ &= E\left(\exp\left(-\int_0^T (\gamma_{s,i} + \gamma_{s,j})\,\mathrm{d}s\right)\right). \end{aligned} \tag{17.30}$$

For hazard rates of the form (17.24), expression (17.30) can be decomposed—using a similar approach to the decomposition in (17.25)—into terms that can be evaluated using our results on one-dimensional affine models.

It is often claimed that the default correlation values that can be attained in models with conditionally independent defaults are too low compared with empirical default correlations (see, for example, Hull and White 2001; Schönbucher and Schubert 2001). Since default correlations do have a significant impact on the loss distribution generated by a model, we discuss this issue further. As a concrete example we use the Duffie–Gârleanu model and assume that there are no idiosyncratic factor processes (Ψ_t^{id}) and that all risk is systematic. As discussed above, in that case the default indicator vector $\boldsymbol{Y}_T$ follows an exchangeable Bernoulli mixture model with mixing variable $\tilde{Q}$ given by $1 - \exp(-\int_0^T \Psi_s^{\mathrm{syst}}\,\mathrm{d}s)$.

We have seen in Section 11.2.2 that in exchangeable Bernoulli mixture models every default correlation $\rho \in (0, 1)$ can be obtained by choosing the variance of the mixing variable sufficiently high. It follows that in the Duffie–Gârleanu model high levels of default correlation can be obtained if the variance of the rv $\Gamma_T := \int_0^T \Psi_s^{\mathrm{syst}}\,\mathrm{d}s$ is sufficiently high. A high variance of Γ_T can be obtained by choosing a high value for the volatility σ of the diffusion part of $(\boldsymbol{\Psi}_t^{\mathrm{syst}})$ or by choosing a high value for the mean of the jump-size distribution μ or for the jump intensity l^0. A high value for σ translates into very volatile day-to-day fluctuations of credit spreads, which might contradict the behaviour of real bond-price data. This

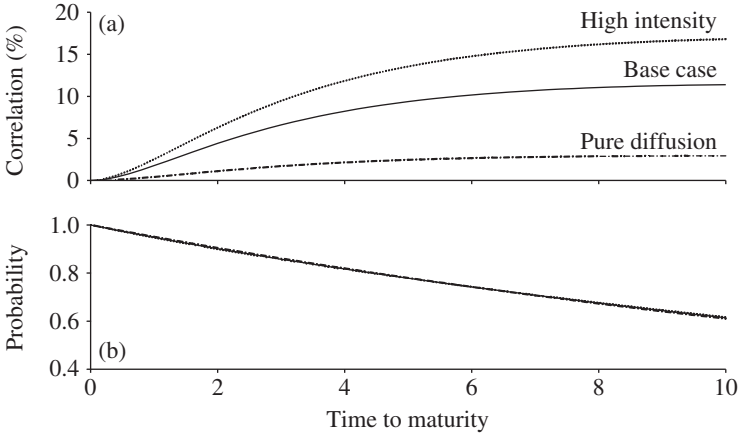

Figure 17.1. Both plots relate to the Duffie–Gârleanu model with $(\Psi_t^{\mathrm{id}}) \equiv 0$ and different parameter sets for (Ψ_t^{syst}), which are given in Table 17.1. (a) The default correlations for varying time to maturity. We see that by increasing the intensity of jumps in (Z_t), the default correlation is increased substantially. Part (b) shows that the survival probabilities for the three parameter sets are essentially equal, so that the differences in default correlations are solely due to the impact of the dynamics of (Ψ_t^{syst}) on the dependence structure of the default times.

shows that it might be difficult to generate very high levels of default correlation in models where hazard rates follow pure diffusion processes.

Instead, in the Duffie–Gârleanu model we can increase the frequency or size of the jumps in the hazard process by increasing l^0 or μ. This is a very effective mechanism for generating default correlation, as is shown in Figure 17.1. In fact, this additional flexibility in modelling default correlations is an important motivation for considering affine jump diffusions instead of the simpler CIR diffusion models.

These qualitative findings obviously carry over to other models with conditionally independent defaults. Summing up, we conclude that it is certainly possible to generate high levels of default correlation in models with conditionally independent defaults; however, the required models for the hazard processes can become quite complex.

17.3.3 Credit Value Adjustments

Finally, we turn to the computation of uncollateralized credit value adjustments for credit default swaps in models with conditionally independent default times. We first recall the notation of Section 17.2.1: τ_S, τ_B and τ_R are the default times of the protection seller, the protection buyer and the reference entity in a given CDS contract; τ_S, τ_B and τ_R are conditionally independent under the risk-neutral pricing measure Q with $(\mathcal{F}_t)$-conditional hazard processes γ^S, γ^B and γ^R; the short rate is given by some $(\mathcal{F}_t)$-adapted process (r_t); recovery rates are deterministic; $T_1 := \tau_S \wedge \tau_B \wedge \tau_R$ is the time of the first default; $D(0,t) = \exp(-\int_0^t r_u \, \mathrm{d}u)$.

As in Section 17.2.1 we denote by $V_t = E^Q(\Pi(t,T) \mid \mathcal{G}_t)$ the net present value of the promised cash flows of the protection-buyer position in the CDS. It was shown in Section 10.5.3 that $V_t = I_{\{\tau_R > t\}} \tilde{V}_t$, where the $(\mathcal{F}_t)$-adapted process $\tilde{V}_t$ is given

by

$$\tilde{V}_t = \delta^R E^Q\bigg(\int_t^T \gamma_s^R \exp\bigg(-\int_t^s (r_u + \gamma_u^R)\,\mathrm{d}u \bigg)\,\mathrm{d}s \;\bigg|\; \mathcal{F}_t \bigg) - x E^Q\bigg(\sum_{\{t_n > t\}} (t_n - t_{n-1}) \exp\bigg(-\int_t^s (r_u + \gamma_u^R)\,\mathrm{d}u \bigg) \;\bigg|\; \mathcal{F}_t \bigg), \quad (17.31)$$

and the time points t_n represent the premium payment dates.

A general formula. For simplicity we consider the case $t = 0$. According to Proposition 17.1, the bilateral uncollateralized value adjustment (for the protection buyer) is given by

$$\begin{aligned} \mathrm{BCVA} &= \mathrm{CVA} - \mathrm{DVA} \\ &= E^Q(I_{\{T_1<T\}} I_{\{\xi=S\}} D(0,T_1)\delta^S(\tilde{V}_{T_1})^+) \\ &\quad - E^Q(I_{\{T_1<T\}} I_{\{\xi=B\}} D(0,T_1)\delta^B(\tilde{V}_{T_1})^-). \end{aligned} \quad (17.32)$$

Here we have used the fact that $V_{T_1} = \tilde{V}_{T_1}$ on $\{\xi = S\}$ or $\{\xi = B\}$. We concentrate on the CVA term in (17.32). Recall from Proposition 17.7 that

$$Q(\xi = S \mid \sigma(T_1) \vee \mathcal{F}_\infty) = \frac{\gamma_{T_1}^S}{\gamma_{T_1}^S + \gamma_{T_1}^B + \gamma_{T_1}^R}.$$

By double conditioning we therefore obtain that

$$\begin{aligned} \mathrm{CVA} &= E^Q(E^Q(I_{\{T_1<T\}} I_{\{\xi=S\}} D(0,T_1)\delta^S(\tilde{V}_{T_1})^+ \mid \sigma(T_1) \vee \mathcal{F}_\infty)) \\ &= E^Q(I_{\{T_1<T\}} D(0,T_1)\delta^S(\tilde{V}_{T_1})^+ Q(\xi = S \mid \sigma(T_1) \vee \mathcal{F}_\infty)) \\ &= E^Q\bigg(I_{\{T_1<T\}} D(0,T_1)\delta^S(\tilde{V}_{T_1})^+ \frac{\gamma_{T_1}^S}{\gamma_{T_1}^S + \gamma_{T_1}^B + \gamma_{T_1}^R} \bigg). \end{aligned}$$

In the terminology of Section 10.5.2 this is a typical payment-at-default claim with payment at the stopping time T_1. By Lemma 17.6, T_1 is a doubly stochastic random time with hazard rate $\gamma_t^S + \gamma_t^B + \gamma_t^R$. Applying the third identity in Theorem 10.19 gives

$$\mathrm{CVA} = E^Q\bigg(\int_0^T \gamma_t^S \delta^S(\tilde{V}_t)^+ \exp\bigg(-\int_0^t (r_s + \gamma_s^S + \gamma_s^B + \gamma_s^R)\,\mathrm{d}s \bigg)\,\mathrm{d}t \bigg). \quad (17.33)$$

A similar computation for the DVA term in (17.32) yields

$$\mathrm{DVA} = E^Q\bigg(\int_0^T \gamma_t^B \delta^B(\tilde{V}_t)^- \exp\bigg(-\int_0^t (r_s + \gamma_s^S + \gamma_s^B + \gamma_s^R)\,\mathrm{d}s \bigg)\,\mathrm{d}t \bigg). \quad (17.34)$$

More explicit results. In order to evaluate (17.33) and (17.34) we need to make more specific assumptions about the form of the hazard rates. Here we derive a PDE for the value adjustments in a one-factor CIR model. More precisely, we assume that γ^S, γ^B, γ^R and the short rate r are affine functions of a one-dimensional CIR process Ψ with parameters κ, θ and σ. Therefore,

$$\gamma_t^R = \gamma^R(t, \Psi_t) = a^R \Psi_t + b^R \tag{17.35}$$

for $a^R > 0, b^R \geqslant 0$, and similarly for γ^B, γ^S and r. In that case, the pre-default value $\tilde{V}_t$ of the CDS is given by a function $\tilde{v}(t, \Psi)$ that can be easily computed using the affine-model techniques introduced in Section 10.6. Moreover, the Feynman–Kac formula (see Lemma 10.21) gives

$$E^Q\left(\int_t^T \gamma_s^S \delta^S \tilde{V}_s^+ \exp\left(-\int_t^T (\gamma_u^S + \gamma_u^B + \gamma_u^R + r_u)\, \mathrm{d}u \right) \,\middle|\, \mathcal{F}_t \right) = h^S(t, \Psi_t),$$

where the function h^S solves the PDE

$$h_t^S + (\kappa - \theta\psi) h_\psi^S + \tfrac{1}{2}\sigma^2 \psi h_{\psi\psi}^S + \gamma^S \delta^S \tilde{v}^+ = (r + \gamma^S + \gamma^B + \gamma^R) h^S. \tag{17.36}$$

Here the arguments (t, ψ) have been omitted for ease of notation. The corresponding PDE for the buyer adjustment is

$$h_t^B + (\kappa - \theta\psi) h_\psi^B + \tfrac{1}{2}\sigma^2 \psi h_{\psi\psi}^B + \gamma^B \delta^B \tilde{v}^- = (r + \gamma^S + \gamma^B + \gamma^R) h^B, \tag{17.37}$$

again with terminal value $h^B(T, \psi) = 0$. Both PDEs can be solved numerically.

Summarizing we obtain, in the special case of the one-factor CIR model, that

$$\mathrm{BCVA} = h^S(t, \Psi_t) - h^B(t, \Psi_t). \tag{17.38}$$

Notes and Comments

For an alternative textbook-level treatment of models with conditionally independent defaults, see, for example, Chapter 9 of Bielecki and Rutkowski (2002). The simulation of conditionally independent default times is discussed in Duffie and Singleton (2003). The existence of default intensities is discussed in Aven (1985) and in Section 2 of Janson, M'Baye and Protter (2011). Both sources contain results that can be viewed as partial converses to Lemma 17.10. A model with conditionally independent defaults where the factor Ψ follows a finite-state Markov chain is considered by Di Graziano and Rogers (2009).

Further empirical work on affine models for credit portfolios includes that of Duffee (1999) and Driessen (2005). We refer the reader to Chapter 10 of Filipović (2009) for further information on affine processes. The empirical estimation of models with conditionally independent defaults and various related issues are discussed in Duffie (2011).

17.4 Credit Risk Models with Incomplete Information

In this section we are concerned with reduced-form portfolio credit risk models under incomplete information. We consider models where the default times are independent given the realization of some common factor; in principle, this factor could be a stochastic process (Ψ_t), as in the models with conditionally independent defaults considered in Section 17.3, but for expository purposes we concentrate on the case where the factor is simply a one-dimensional random variable V. We assume that V is not directly observable by investors. Rather, their information set consists of the default history and—in the more advanced versions of the model—additional noisy observations of the factor represented by an auxiliary σ-algebra.

It will turn out that portfolio credit risk models with incomplete information have a number of attractive features. To begin with, they are able to generate rich dynamics of credit spreads incorporating both default contagion and credit-spread volatility effects. Moreover, the pricing of typical credit derivatives can be carried out using a natural and fairly efficient two-step procedure.

Our presentation starts with a general discussion of credit risk and incomplete information in Section 17.4.1. In Section 17.4.2 we consider simple models in which investors observe only the default history; the extension to a richer information set and applications to counterparty risk management are discussed in Sections 17.4.3 and 17.4.4.

17.4.1 Credit Risk and Incomplete Information

Throughout this section we use the following set-up. We work on a filtered probability space $(\Omega, \mathcal{F}, (\tilde{\mathcal{G}}_t), Q)$, where Q is the risk-neutral measure used for pricing derivatives and $(\tilde{\mathcal{G}}_t)$ is the global filtration, so all processes introduced will be $(\tilde{\mathcal{G}}_t)$-adapted. We consider a portfolio of m obligors with default times $\tau_1, \dots, \tau_m$ and default indicator processes $(Y_{t,i})$ given by $Y_{t,i} = I_{\{\tau_i \leqslant t\}}$, $1 \leqslant i \leqslant m$. We assume that there is a random variable V on $(\Omega, \mathcal{F}, Q)$ such that the rvs $\tau_1, \dots, \tau_m$ are conditionally independent given V, with conditional survival functions of the form

$$\bar{F}_{\tau_i|V}(t \mid v) = \exp\left(-\int_0^t \gamma_i(v, s)\, \mathrm{d}s\right) \tag{17.39}$$

for continuous functions $\gamma_i : \mathbb{R} \times [0, \infty) \to (0, \infty)$, such that $\int_0^t \gamma_i(v, s)\, \mathrm{d}s < \infty$ for all t and v. Note that this implies that $\gamma_i(v, t)$ is the conditional hazard function, and the conditional density $f_{\tau_i|V}(t \mid v)$ satisfies

$$f_{\tau_i|V}(t \mid v) = \gamma_i(v, t) \exp\left(-\int_0^t \gamma_i(v, s)\, \mathrm{d}s\right). \tag{17.40}$$

We consider two cases for the distribution of the factor V. In the first case V is an absolutely continuous rv with density $g_V(v)$, as in the factor copula models considered in Section 12.2.2. In the second case V is a discrete random variable with values in the set $S^V := \{v_1, \dots, v_K\}$ and probability mass function $\pi_k = Q(V = v_k)$, $1 \leqslant k \leqslant K$, as in the class of implied copula models considered in

Section 12.3.3. Finally, we always take the loss given default δ_i of the firms in the portfolio and the default-free interest rate $r(t)$ to be deterministic.

The full-information filtration is given by $\tilde{\mathcal{G}}_t = \mathcal{H}_t \vee \sigma(V) \vee \sigma(B_s : s \leqslant t)$, $t > 0$. Here, $(\mathcal{H}_t)$ is the default history of the portfolio and (B_t) is a Brownian motion independent of V and of the τ_i. (The process (B_t) is used in Section 17.4.3 as a building block in modelling the noisy information available to investors.) If considered under the full-information filtration $(\tilde{\mathcal{G}}_t)$, the model is therefore a model with conditionally independent defaults as discussed in Section 17.3. However, as mentioned before, the rv V is not observable by investors. We therefore describe the information available to investors by a filtration $(\mathcal{G}_t)$ with $\mathcal{G}_t \subset \tilde{\mathcal{G}}_t$, $t \geqslant 0$, and refer to $(\mathcal{G}_t)$ as the *investor filtration*. We assume throughout that investors are able to observe the default history of the portfolio, i.e. $\mathcal{H}_t \subset \mathcal{G}_t$ for all $t \geqslant 0$. On the other hand, V is not $\mathcal{G}_t$ measurable for all $t \geqslant 0$.

Pricing of credit derivatives. Consider a single-name CDS, a CDS index or a CDO tranche on the firms in the portfolio (or a subgroup thereof), and denote by $\Pi(t, T)$ the corresponding stream of discounted future cash flows. In the setting of this section, the price of this cash-flow stream at time t is given by $V_t = E^Q(\Pi(t, T) \mid \mathcal{G}_t)$. Note that V_t is given by a conditional expectation with respect to $\mathcal{G}_t$, the information actually available to investors at time t.

The price V_t can be computed in two steps. By iterated conditional expectations we obtain that

$$V_t = E^Q(E^Q(\Pi(t, T) \mid \tilde{\mathcal{G}}_t) \mid \mathcal{G}_t). \tag{17.41}$$

In the first step we consider the inner conditional expectation. In order to evaluate this expectation we need to determine the $\tilde{\mathcal{G}}_t$-conditional distribution of the non-defaulted firms. Denote by $\{i_1, \ldots, i_\ell\} \subseteq \{1, \ldots, m\}$ the identity of these firms. Since the Brownian motion (B_t) is independent of V and the default times τ_i, we find, for arbitrary $t_1, \ldots, t_\ell \geqslant t$, that

$$Q(\tau_{i_1} > t_1, \ldots, \tau_{i_\ell} > t_\ell \mid \tilde{\mathcal{G}}_t) = Q(\tau_{i_1} > t_1, \ldots, \tau_{i_\ell} > t_\ell \mid \mathcal{H}_t \vee \sigma(V)).$$

Since the τ_i are independent given V, it follows that

$$\begin{aligned} Q(\tau_{i_1} > t_1, \ldots, \tau_{i_\ell} > t_\ell \mid \mathcal{H}_t \vee \sigma(V)) &= \prod_{l=1}^{\ell} Q(\tau_{i_l} > t_l \mid \mathcal{H}_t^{i_l} \vee \sigma(V)) \\ &= \prod_{l=1}^{\ell} I_{\{\tau_{i_l} > t\}} \frac{Q(\tau_{i_l} > t_l \mid \sigma(V))}{Q(\tau_{i_l} > t \mid \sigma(V))} \\ &= \prod_{l=1}^{\ell} I_{\{\tau_{i_l} > t\}} \exp\left(- \int_t^{t_l} \gamma_{i_l}(V, s)\, \mathrm{d}s \right). \end{aligned}$$

The firms in the portfolio are therefore conditionally independent given $\tilde{\mathcal{G}}_t$, with conditional survival probabilities

$$Q(\tau_i > s \mid \tilde{\mathcal{G}}_t) = I_{\{\tau_i > t\}} \exp\left(- \int_t^s \gamma_i(V, s)\, \mathrm{d}s \right), \quad s \geqslant t. \tag{17.42}$$

As in Section 12.3.1 it follows that the value $E^Q(\Pi(t,T) \mid \tilde{\mathcal{G}}_t)$ (the price with respect to the larger σ-algebra $\tilde{\mathcal{G}}_t$) depends only on the time t, the realized value of the factor V and the realized value of the default indicator vector $\boldsymbol{Y}_t = (Y_{t,1}, \ldots, Y_{t,m})'$ at time t. In summary, this means that $E^Q(\Pi(t,T) \mid \tilde{\mathcal{G}}_t) = h(t, V, \boldsymbol{Y}_t)$ for a suitable function $h(\cdot)$. For instance, in the case of a protection-buyer position in a single-name CDS on firm i with premium payment dates $t_1, \ldots, t_N > t$ and spread x, we have that

$$h(t,v,\boldsymbol{y}) = (1-y_i)\bigg(\delta_i \int_t^{t_N} \gamma_i(v,s) \exp\bigg(-\int_t^s r(u) + \gamma_i(v,u)\,\mathrm{d}u\bigg)\,\mathrm{d}s \\ - x\sum_{n=1}^{N}(t_n - t_{n-1}) \exp\bigg(-\int_t^{t_n} r(s) + \gamma_i(v,s)\,\mathrm{d}s\bigg)\bigg).$$

For CDO tranches, the function $h(\cdot)$ can be computed using the analytics for factor credit risk models developed in Section 12.3.1.

Using (17.41) we infer that the price of the cash-flow stream at time t is given by

$$V_t = E^Q(h(t,V,\boldsymbol{Y}_t) \mid \mathcal{G}_t) = \int_{\mathbb{R}} h(t,v,\boldsymbol{Y}_t) g_{V|\mathcal{G}_t}(v)\,\mathrm{d}v, \tag{17.43}$$

where $g_{V|\mathcal{G}_t}(v)$ represents the conditional density of V given $\mathcal{G}_t$. If V is discrete, the last integral is replaced by the sum

$$\sum_{k=1}^{K} h(t, v_k, \boldsymbol{Y}_t) Q(V = v_k \mid \mathcal{G}_t).$$

The pricing formulas therefore have a similar structure to those in the static factor models considered in Section 12.3, but the unconditional density $g_V(v)$ and the unconditional probability mass function $\boldsymbol{\pi}$ have to be replaced by the conditional density $g_{V|\mathcal{G}_t}(v)$ and the conditional probabilities $\pi_t^k := Q(V = v_k \mid \mathcal{G}_t)$, $1 \leqslant k \leqslant K$. The computation of these quantities is discussed below.

Default intensities. A similar two-step procedure applies to the computation of default intensities. Since defaults are conditionally independent given V, we can apply Proposition 17.12 to infer that the default intensity of firm i with respect to the global filtration $(\tilde{\mathcal{G}}_t)$ satisfies $\tilde{\lambda}_{t,i} = \gamma_i(V,t)$. In order to compute the intensity of τ_i with respect to the investor filtration $(\mathcal{G}_t)$, we use the following general result.

Lemma 17.14. *Consider two filtrations $(\mathcal{G}_t)$ and $(\tilde{\mathcal{G}}_t)$ such that $\mathcal{G}_t \subset \tilde{G}_t$ for all $t \geqslant 0$ and some random time τ such that the corresponding indicator process $Y_t = I_{\{\tau \leqslant t\}}$ is $(\mathcal{G}_t)$-adapted. Suppose that τ admits the $(\tilde{G}_t)$-intensity $\tilde{\lambda}_t$. Then the $(\mathcal{G}_t)$ default intensity is given by the right-continuous version of the process $\lambda_t = E(\tilde{\lambda}_t \mid \mathcal{G}_t)$.*

Lemma 17.14 is a special case of Theorem 14 in Chapter 2 of Brémaud (1981) and we refer to that source for a proof. Applying Lemma 17.14 we conclude that $(\lambda_{t,i})$, the $(\mathcal{G}_t)$ default intensity of firm i, is given by

$$\lambda_{t,i} = E(\gamma_i(V,t) \mid \mathcal{G}_t) = \int_{\mathbb{R}} \gamma_i(v,t) g_{V|\mathcal{G}_t}(v)\,\mathrm{d}v; \tag{17.44}$$

in the case where V is discrete, the last integral is replaced by the sum

$$\sum_{k=1}^{K} \gamma_i(v_k, t) Q(V = v_k \mid \mathcal{G}_t).$$

17.4.2 Pure Default Information

In this section we consider the case where the only information available to investors is the default history of the portfolio. This is modelled by setting $(\mathcal{G}_t) = (\mathcal{H}_t)$.

Recursive computation of $g_{V|\mathcal{H}_t}$. We concentrate on the case where V has a density; the discrete case can be handled by analogous arguments. We will compute the conditional density $g_{V|\mathcal{H}_t}$ on the sets $\{t < T_1\}$ (before the first default) and $\{T_1 \leqslant t < T_2\}$ (after the first default and before the second).

Let $A \subseteq \mathbb{R}$ be a measurable set. We first note that on the set $\{t < T_1\}$ we have $Q(V \in A \mid \mathcal{H}_t) = Q(V \in A \mid T_1 > t)$. Now

$$Q(V \in A \mid T_1 > t) = \frac{Q(\{V \in A\} \cap \{T_1 > t\})}{Q(T_1 > t)} \tag{17.45}$$

and, using the conditional independence of defaults given V, the numerator may be calculated to be

$$\begin{aligned} Q(\{V \in A\} \cap \{T_1 > t\}) &= \int_A Q(T_1 > t \mid V = v) g_V(v)\, \mathrm{d}v \\ &= \int_A \prod_{i=1}^{m} Q(\tau_i > t \mid V = v) g_V(v)\, \mathrm{d}v \\ &= \int_A \exp\left(-\int_0^t \sum_{i=1}^{m} \gamma_i(v, s)\, \mathrm{d}s\right) g_V(v)\, \mathrm{d}v. \end{aligned}$$

It follows from (17.45) that on $\{t < T_1\}$ the conditional df $G_{V|\mathcal{H}_t}(v)$ is given by

$$G_V(v \mid \mathcal{H}_t) = \frac{1}{Q(T_1 > t)} \int_{-\infty}^{v} \exp\left(-\int_0^t \sum_{i=1}^{m} \gamma_i(w, s)\, \mathrm{d}s\right) g_V(w)\, \mathrm{d}w,$$

and hence the conditional density $g_{V|\mathcal{H}_t}(v)$ satisfies

$$g_{V|\mathcal{H}_t}(v) \propto \exp\left(-\int_0^t \sum_{i=1}^{m} \gamma_i(v, s)\, \mathrm{d}s\right) g_V(v),$$

where $\propto$ stands for "is proportional to". The constant of proportionality is of course given by $(Q(T_1 > t))^{-1}$, but this quantity is irrelevant for our subsequent analysis. To simplify notation we define the quantity

$$\bar{\gamma}(v, s) = \sum_{i=1}^{m} (1 - Y_{s,i}) \gamma_i(v, s), \tag{17.46}$$

and with this notation we have that on $\{t < T_1\}$

$$g_{V|\mathcal{H}_t}(v) \propto \exp\left(-\int_0^t \bar{\gamma}(v, s)\, \mathrm{d}s\right) g_V(v). \tag{17.47}$$

Next we calculate $g_{V|\mathcal{H}_t}(v)$ on the set $\{T_1 \leqslant t < T_2\}$. As an intermediate step we derive $g_{V|\tau_j}(v \mid t)$, the conditional density of V given the default time τ_j of an arbitrary firm j in the portfolio. Applying the conditional density formula (see, for instance, (6.2)) to the conditional density $f_{\tau_j|V}(t \mid v)$ given in (17.40), we obtain

$$g_{V|\tau_j}(v \mid t) \propto f_{\tau_j|V}(t \mid v) g_V(v) = \gamma_j(v,t) \exp\left(-\int_0^t \gamma_j(v,s)\,\mathrm{d}s\right) g_V(v). \tag{17.48}$$

The available information at t consists of the rvs T_1 and ξ_1 (the default time and identity ξ_1 of the firm defaulting first) and the event $B_t := \{\tau_i > t,\ i \neq \xi_1\}$ (the knowledge that the other firms have not yet defaulted at t). We can therefore write $Q(V \in A \mid \mathcal{H}_t) = Q(V \in A \mid B_t, T_1, \xi_1)$ on $\{T_1 \leqslant t < T_2\}$ and we note that

$$Q(V \in A \mid B_t, T_1, \xi_1) = \frac{Q(\{V \in A\} \cap B_t \mid T_1, \xi_1)}{Q(B_t \mid T_1, \xi_1)} \propto Q(\{V \in A\} \cap B_t \mid T_1, \xi_1).$$

Now we calculate that

$$Q(\{V \in A\} \cap B_t \mid T_1, \xi_1) = \int_A \prod_{i \neq \xi_1} \exp\left(-\int_0^t \gamma_i(v,s)\,\mathrm{d}s\right) g_{V|\tau_{\xi_1}}(v \mid T_1)\,\mathrm{d}v,$$

and we can use (17.48) to infer that this probability is proportional to

$$\int_A \gamma_{\xi_1}(v, T_1) \exp\left(-\int_0^{T_1} \sum_{i=1}^m \gamma(v,s)\,\mathrm{d}s\right) \exp\left(-\int_{T_1}^t \sum_{i \neq \xi_1} \gamma(v,s)\,\mathrm{d}s\right) g_V(v)\,\mathrm{d}v.$$

Using the notation (17.46) we find that on $\{T_1 \leqslant t < T_2\}$

$$g_{V|\mathcal{H}_t}(v) \propto \gamma_{\xi_1}(v, T_1) \exp\left(-\int_0^t \bar{\gamma}(v,s)\,\mathrm{d}s\right) g_V(v). \tag{17.49}$$

If we compare (17.47) and (17.49), we can see the impact of the additional default information at T_1 on the conditional distribution of V. At T_1 the "a posteriori density" $g_{V|\mathcal{H}_{T_1}}(v)$ is proportional to the product of the hazard rate $\gamma_{\xi_1}(v, T_1)$ and the "a priori density" $g_{V|\mathcal{H}_t}(v)$ just before T_1. In the following two examples this result will be used to derive explicit expressions for information-driven contagion effects.

The above analysis is a relatively simple example of a *stochastic filtering problem*. In general, in a filtering problem, we consider a stochastic process (Ψ_t) (a random variable is a special case) and a filtration $(\mathcal{G}_t)$ such that (Ψ_t) is not $(\mathcal{G}_t)$-adapted. We attempt to estimate the conditional distribution of Ψ_t for $t \geqslant 0$ given the σ-algebra $\mathcal{G}_t$ in a recursive way.

Example 17.15 (Clayton copula model). For factor copula models with a Clayton survival copula, the conditional density $g_{V|\mathcal{H}_t}(v)$ and the default intensities $(\lambda_{t,i})$ can be computed explicitly. The Clayton copula model with parameter $\theta > 0$ is an LT-Archimedean copula model (see Example 12.5) with $V \sim \mathrm{Ga}(1/\theta, 1)$. We denote the Laplace–Stieltjes transform of the distribution of V and its functional inverse by $\hat{G}_V$ and $\hat{G}_V^{-1}$, respectively, and we recall that the density $g(v; \alpha, \beta)$ of the $\mathrm{Ga}(\alpha, \beta)$ distribution satisfies $g(v; \alpha, \beta) \propto v^{\alpha-1}\mathrm{e}^{-\beta v}$ and has mean α/β (see Appendix A.2.4).

As shown in Example 12.5, in threshold copula models with an LT-Archimedean copula the conditional survival function is given by

$$\bar{F}_{\tau_i|V}(t \mid v) = \exp(-v\hat{G}_V^{-1}(\bar{F}_i(t))),$$

where $\bar{F}_i(t)$ denotes the marginal survival function. The conditional hazard function $\gamma_i(v,t)$ is therefore given by

$$\gamma_i(v,t) = -\frac{\mathrm{d}}{\mathrm{d}t}\ln \bar{F}_{\tau_i|V}(t \mid v) = -v\frac{\mathrm{d}}{\mathrm{d}t}\hat{G}_V^{-1}(\bar{F}_i(t)). \tag{17.50}$$

Since $Q(T_1 > t \mid V = v) = \exp(-v\sum_{i=1}^m \hat{G}_V^{-1}(\bar{F}_i(t)))$, we can use (17.47) to obtain

$$g_{V|\mathcal{H}_t}(v) \propto v^{(1/\theta)-1}\exp\left(-v\left(1+\sum_{i=1}^m \hat{G}_V^{-1}(\bar{F}_i(t))\right)\right), \quad t < T_1. \tag{17.51}$$

Hence, for $t < T_1$, the conditional distribution of V given $\mathcal{H}_t$ is again a gamma distribution but now with parameters $\alpha = 1/\theta$ and $\beta = 1 + \sum_{i=1}^m \hat{G}_V^{-1}(\bar{F}_i(t))$. Since the mean of this distribution is α/β, we see that the conditional mean of V given $T_1 > t$ is lower than the unconditional mean of V. This is in line with economic intuition; indeed, the fact that $T_1 > t$ is "good news" for the portfolio.

Next we turn to the updating at $t = T_1$. Using (17.49), (17.50) and (17.51) we find that

$$g_{V|\mathcal{H}_{T_1}} \propto v\left\{v^{(1/\theta)-1}\exp\left(-v\left(1+\sum_{i=1}^m \hat{G}_V^{-1}(\bar{F}_i(T_1))\right)\right)\right\}, \tag{17.52}$$

so that given $\mathcal{H}_{T_1}$, V follows a gamma distribution with parameters $\alpha = 1 + 1/\theta$ and $\beta = 1 + \sum_{i=1}^m \hat{G}_V^{-1}(\bar{F}_i(T_1))$. At T_1 the conditional mean $\mu_{V|\mathcal{H}_t}$ of V jumps upwards. We have the formulas

$$\lim_{t\to T_1^-}\mu_{V|\mathcal{H}_t} = \frac{1/\theta}{1+\sum_{j=1}^m \hat{G}_V^{-1}(\bar{F}_j(T_1))}, \quad \mu_{V|\mathcal{H}_{T_1}} = \frac{1+1/\theta}{1+\sum_{j=1}^m \hat{G}_V^{-1}(\bar{F}_j(T_1))}.$$

Readers familiar with Bayesian statistics may note that the explicit form of the updating rules is due to the fact that the gamma family is a conjugate family for the exponential distribution.

Finally, we consider the $(\mathcal{H}_t)$ default intensity process $(\lambda_{t,i})$. Since $\gamma_i(v,t) \propto v$, we use (17.44) to infer that $\lambda_{t,i}$ is proportional to $E(V \mid \mathcal{H}_t)$, the conditional mean of V given $\mathcal{H}_t$. In particular, $\lambda_{t,i}$ jumps upwards at each successive default time and decreases gradually between defaults. This is illustrated in Figure 17.2, where we consider an explicit example in which, for all i, the marginal survival functions are given by $\bar{F}_i(t) = \mathrm{e}^{-\gamma t}$ for a fixed default rate γ. The parameter θ is chosen to obtain a desired level of default correlation.

Example 17.16 (implied copula model). In this example we take a closer look at the case where V is discrete with state space $S^V = \{v_1, \ldots, v_K\}$ and probability mass function $\boldsymbol{\pi}$. In this way we obtain a dynamic version of the implied copula model from Section 12.3.3.

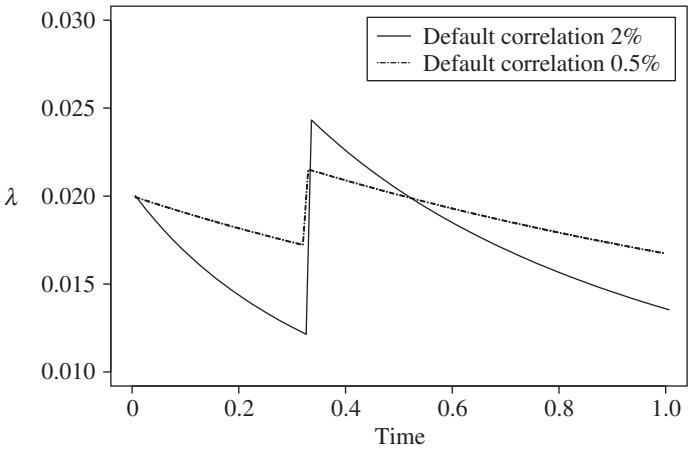

Figure 17.2. Paths of the default intensity (λ_t) in the Clayton copula model, assuming that the first default time T_1 is four months. The parameters are as follows: portfolio size $m = 100$; marginal default rate $\gamma = 0.02$; θ is chosen to give one-year default correlations of 2% and 0.5%. As we expect, a higher default correlation implies a stronger contagion effect.

Using similar arguments to those used to derive (17.47) and (17.49) gives, for $t < T_1$, that

$$Q(V = v_k \mid \mathcal{H}_t) = \frac{\exp(-\int_0^t \bar{\gamma}(v_k, s)\,\mathrm{d}s)\pi_k}{\sum_{l=1}^K \exp(-\int_0^t \bar{\gamma}(v_l, s)\,\mathrm{d}s)\pi_l}, \quad t < T_1; \tag{17.53}$$

at T_1 we have

$$Q(V = v_k \mid \mathcal{H}_{T_1}) \propto \gamma_{\xi_1}(v_k, T_1) \exp\left(-\int_0^{T_1} \bar{\gamma}(v_k, s)\,\mathrm{d}s\right)\pi_k. \tag{17.54}$$

Set $\pi_t^k := Q(V = v_k \mid \mathcal{H}_t)$, $1 \leqslant k \leqslant K$. We now use relations (17.53) and (17.54) to derive a system of stochastic differential equations for the process $\boldsymbol{\pi}_t = (\pi_t^1, \ldots, \pi_t^K)'$, which is driven by time and the default indicator process $(\boldsymbol{Y}_t)$. This helps to understand the dynamics of the default intensities and of credit derivatives.

First we consider the dynamics of the process $(\boldsymbol{\pi}_t)$ between default times. Define $E_t^k := \exp(-\int_0^t \bar{\gamma}(v_k, s)\,\mathrm{d}s)$. With this abbreviation we get, for $t < T_1$, the relationship $\pi_t^k = (\pi_k E_t^k)/(\sum_{l=1}^K \pi_l E_t^l)$, and hence

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\pi_t^k &= \frac{-\pi_k \bar{\gamma}(v_k, t) E_t^k \big(\sum_{l=1}^K \pi_l E_t^l\big) + \pi_k E_t^k \big(\sum_{l=1}^K \pi_l \bar{\gamma}(v_l, t) E_t^l\big)}{(\sum_{l=1}^K \pi_l E_t^l)^2} \\
&= \pi_t^k \left(-\bar{\gamma}(v_k, t) + \frac{\sum_{l=1}^K \pi_l \bar{\gamma}(v_l, t) E_t^l}{\sum_{l=1}^K \pi_l E_t^l}\right) \\
&= \pi_t^k \left(-\bar{\gamma}(v_k, t) + \sum_{l=1}^K \pi_t^l \bar{\gamma}(v_l, t)\right), \quad 1 \leqslant k \leqslant K.
\end{aligned} \tag{17.55}$$

Note that (17.55) is a system of K ordinary differential equations for $(\boldsymbol{\pi}_t)$, so the process $(\boldsymbol{\pi}_t)$ evolves in a deterministic manner up to time T_1. Next we consider the

case $t = T_1$. We will use the general notation $\pi^k_{t-} := \lim_{t \to t^-} \pi^k_t$. From (17.54) we get that

$$\pi^k_{T_1} = \frac{\gamma_{\xi_1}(v_k, T_1) E^k_{T_1} \pi_k}{\sum_{l=1}^K \gamma_{\xi_1}(v_l, T_1) E^l_{T_1} \pi_l} = \frac{\gamma_{\xi_1}(v_k, T_1) \pi^k_{T_1^-}}{\sum_{l=1}^K \gamma_{\xi_1}(v_l, T_1) \pi^l_{T_1^-}}, \tag{17.56}$$

where the second equality follows since $\pi^k_{T_1^-} = \pi_k E^k_{T_1} / (\sum_l \pi_l E^l_{T_1})$. Observe that at $t = T_1$ the process (π^k_t) has a jump of size

$$\pi^k_{t-} \left(\frac{\gamma_{\xi_1}(v_k, t) \pi^k_{t-}}{\sum_{l=1}^K \gamma_{\xi_1}(v_l, t) \pi^l_{t-}} - 1 \right).$$

Combining (17.55) and (17.56) we therefore obtain the following K-dimensional system of stochastic differential equations for the dynamics of $(\boldsymbol{\pi}_t)$ for all t:

$$\begin{aligned} \mathrm{d}\pi^k_t = {} & \pi^k_{t-} \left(-\bar{\gamma}(v_k, t) + \sum_{l=1}^K \pi^l_{t-} \bar{\gamma}(v_l, t) \right) \mathrm{d}t \\ & + \sum_{i=1}^m \pi^k_{t-} \left(\frac{\gamma_i(v_k, t) \pi^k_{t-}}{\sum_{l=1}^K \gamma_i(v_l, t) \pi^l_{t-}} - 1 \right) \mathrm{d}Y_{t,i}, \quad 1 \leqslant k \leqslant K. \end{aligned} \tag{17.57}$$

Recall that $\lambda_{t,i} = \sum_{k=1}^K \pi^k_t \gamma_i(v_k, t)$ is the $(\mathcal{H}_t)$ default intensity of firm i. Define the compensated default indicator processes by

$$M_{t,i} = Y_{t,i} - \int_0^t (1 - Y_{s-,i}) \lambda_{s-,i} \, \mathrm{d}s = Y_{t,i} - \int_0^t (1 - Y_{s-,i}) \sum_{k=1}^K \pi^k_{s-} \gamma_i(v_k, s) \, \mathrm{d}s. \tag{17.58}$$

Next we show that the system (17.57) can be simplified by using the processes $(M_{t,1}), \ldots, (M_{t,m})$ as drivers. This representation will be useful to see the link between our analysis and the more general filtering model considered in Section 17.4.3. To this end, note that the $\mathrm{d}t$ term in (17.57) is equal to

$$\pi^k_{t-} \sum_{i=1}^m \left((1 - Y_{t,i}) \left(\frac{\gamma_i(v_k, t)}{\sum_{l=1}^K \pi^l_{t-} \gamma_i(v_l, t)} - 1 \right) \left(- \sum_{l=1}^K \pi^l_{t-} \gamma_i(v_l, t) \right) \right).$$

Using the compensated default indicator processes as drivers, the dynamics of $(\boldsymbol{\pi}_t)$ can therefore be rewritten in the following simpler form:

$$\mathrm{d}\pi^k_t = \sum_{i=1}^m \pi^k_{t-} \left(\frac{\gamma_i(v_k, t)}{\sum_{l=1}^K \gamma_i(v_l, t) \pi^l_{t-}} - 1 \right) \mathrm{d}M_{t,i}. \tag{17.59}$$

Default intensities and contagion. Equation (17.59) determines the dynamics of the $(\mathcal{H}_t)$ default intensities. As in the Clayton copula model, $(\lambda_{t,i})$ evolves deterministically between default events and jumps at the random default times in the portfolio. It is possible to give an explicit expression for the size of this jump and hence for the contagion effects induced by incomplete information, as we now show.

Consider two firms $i \neq j$. It follows from (17.59) that the jump in the default intensity of firm i at the default time τ_j of firm j is given by

$$\begin{aligned}\lambda_{\tau_j,i} - \lambda_{\tau_j-,i} &= \sum_{k=1}^{K} \gamma_i(v_k, \tau_j)\pi^k_{\tau_j-}\left(\frac{\gamma_j(v_k, \tau_j)}{\sum_{l=1}^{K} \gamma_j(v_l, \tau_j)\pi^l_{\tau_j-}} - 1\right) \\ &= \frac{\mathrm{cov}^{\boldsymbol{\pi}_{\tau_j-}}(\gamma_i, \gamma_j)}{\lambda_{\tau_j-,j}}. \end{aligned} \tag{17.60}$$

Here, $\mathrm{cov}^{\boldsymbol{\pi}}$ denotes the covariance with respect to the probability measure $\boldsymbol{\pi}$ on S^V, and γ_i is shorthand for the random variable $v \mapsto \gamma_i(v, \tau_j)$.

Formula (17.60) makes two very intuitive predictions about default contagion. First, all other things being equal, default contagion increases with increasing correlation of the random variables γ_i and γ_j under $\boldsymbol{\pi}_{\tau_j-}$, i.e. contagion effects are strongest for obligors with similar characteristics. Second, contagion effects are inversely proportional to $\lambda_{\tau_j-,j}$, the default intensity of the defaulting entity. In particular, the default of an entity j with high credit quality and, hence, a low value of $\lambda_{\tau_j-,j}$ has a comparatively large impact on the market, perhaps because the default comes as a surprise to market participants.

17.4.3 Additional Information

In the dynamic version of the implied copula model studied in Example 17.16, default intensities (and hence credit spreads) evolve deterministically between defaults. This unrealistic behaviour is due to the assumption that the investor filtration is the default history $(\mathcal{H}_t)$. With this choice of filtration, significant new information enters the model only at default times. In this section we discuss an extension of the model with a richer investor filtration $(\mathcal{G}_t)$ that is due to Frey and Schmidt (2012).

The set-up. As in Example 17.16 we assume that the factor is a discrete rv V taking values in the set $S^V = \{v_1, \ldots, v_K\}$. The investor filtration $(\mathcal{G}_t)$ is generated by the default history of the firms in the portfolio and—this is the new part—an additional process (Z_t) representing observations of V with additive noise. Formally, we set $(\mathcal{G}_t) = (\mathcal{H}_t) \vee (\mathcal{F}^Z_t)$, where (Z_t) is of the form

$$Z_t = \int_0^t a(V, s)\,\mathrm{d}s + \sigma B_t. \tag{17.61}$$

Here, (B_t) is a Brownian motion, independent of V and the default indicator process $(\boldsymbol{Y}_t)$, $a(\cdot)$ is a function from $S^V \times [0, \infty)$ to $\mathbb{R}$, and σ is a scaling parameter that modulates the effect of the noise.

Equation (17.61) is the standard way of incorporating noisy information about V into a continuous-time stochastic filtering model. To develop more intuition for the model we show how (17.61) arises as the limit of a simpler discrete-time model. Suppose that investors receive noisy information about V at discrete time points $t_k = k\Delta, k = 1, 2, \ldots$, and that this information takes the form $X_{t_k} = a(V, t_k) + \epsilon_k$ for an iid sequence of noise variables ϵ_k, independent of V, with mean 0 and variance

$\sigma_\epsilon^2 > 0$. Now define the scaled cumulative observation process $Z_t^\Delta := \Delta \sum_{t_k \leqslant t} X_{t_k}$ and let $\sigma = \sqrt{\Delta \sigma_\epsilon^2}$. For Δ small we have the approximation

$$Z_t^\Delta = \sum_{t_k \leqslant t} \Delta a(V, t_k) + \Delta \sum_{t_k \leqslant t} \epsilon_k \approx \int_0^t a(V_s, s)\, \mathrm{d}s + \sigma B_t. \tag{17.62}$$

Without loss of generality we set $\sigma = 1$; other values of σ could be incorporated by rescaling the function $a(\cdot)$. In the numerical examples considered below we assume $a(v_k, t) = c \ln \bar{\gamma}(v_k, t)$, where $\bar{\gamma}(v_k, t)$ is as defined in (17.46) and the constant $c \geqslant 0$ models the information content of (Z_t); for $c = 0$, (Z_t) carries no information, whereas, for c large, the state of V can be observed with high precision. The process (Z_t) is an abstract model-building device that represents information contained in security prices; it is not directly linked to observable economic quantities. We come back to this point when we discuss calibration strategies for the model.

Dynamics of $(\boldsymbol{\pi}_t)$. As before we use the notation $\pi_t^k = Q(V = v_k \mid \mathcal{G}_t)$, $1 \leqslant k \leqslant K$, and $\boldsymbol{\pi}_t = (\pi_t^1, \ldots, \pi_t^K)'$. A crucial part of the analysis of Frey and Schmidt (2012) is the derivation of a stochastic differential equation for the dynamics of the process $(\boldsymbol{\pi}_t)$. We begin by introducing the processes that drive this equation. According to Lemma 17.14, the $(\mathcal{G}_t)$ default intensity of firm i is given by $\lambda_{t,i} = \sum_{l=1}^K \pi_t^k \gamma_i(v_k, t)$, and the compensated default indicator processes are given by the martingales $(M_{t,i})$ introduced in (17.58). Moreover, we define the process

$$W_t = Z_t - \int_0^t \sum_{k=1}^K \pi_s^k a(v_k, s)\, \mathrm{d}s, \quad t \geqslant 0. \tag{17.63}$$

In intuitive terms we have the relationship

$$E(\mathrm{d}Z_t \mid \mathcal{G}_t) = E(a(v_k, t)\, \mathrm{d}t \mid \mathcal{G}_t) = \sum_{k=1}^K \pi_t^k a(v_k, t)\, \mathrm{d}t.$$

Hence, $\mathrm{d}W_t = \mathrm{d}Z_t - E(\mathrm{d}Z_t \mid \mathcal{G}_t)\, \mathrm{d}t$, so the increment $\mathrm{d}W_t$ represents the unpredictable part of the new information $\mathrm{d}Z_t$. For this reason, (W_t) is called the *innovations process* in the literature on stochastic filtering. It is well known that (W_t) is a $(\mathcal{G}_t)$-Brownian motion (for a formal proof see, for example, Bain and Crisan (2009) or Davis and Marcus (1981)).

The following result from Frey and Schmidt (2012) generalizes equation (17.59).

Proposition 17.17. *The process* $(\boldsymbol{\pi}_t) = (\pi_t^1, \ldots, \pi_t^K)'$ *solves the SDE system*

$$\mathrm{d}\pi_t^k = \sum_{i=1}^m \pi_{t-}^k \left(\frac{\gamma_i(v_k, t)}{\sum_{l=1}^K \gamma_i(v_l, t) \pi_{t-}^l} - 1 \right) \mathrm{d}M_{t,i} + \pi_{t-}^k \left(a(v_k, t) - \sum_{l=1}^K \pi_t^l a(v_l, t) \right) \mathrm{d}W_t, \tag{17.64}$$

$1 \leqslant k \leqslant K$, *with initial condition* $\boldsymbol{\pi}_0 = \boldsymbol{\pi}$.

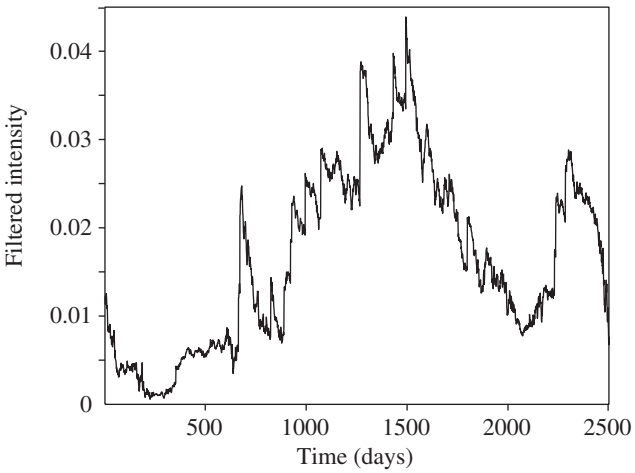

Figure 17.3. A simulated path for the default intensity in the model with incomplete information where additional information is modelled by the noisy observation process (Z_t). The graph was created using the set-up of Example 12.7.

In stochastic filtering, equation (17.64) is known as the *Kushner–Stratonovich equation*. We omit the formal proof of the proposition but try to give some intuitive explanation for equation (17.64). The jump part of this equation has the same form as in equation (17.59). This term represents the impact of default information on $\boldsymbol{\pi}_t$ and it is responsible for contagion effects due to defaults. Next we consider the diffusion part. Define random variables $\tilde{a}$ and $I_k \colon S^V \to \mathbb{R}$ by $\tilde{a}(v) = a(v,t)$ and $I_k(v) = I_{\{v=v_k\}}$. The coefficient of the diffusion part can then be written in the form

$$E^{\boldsymbol{\pi}}(I_k, \tilde{a}) - E^{\boldsymbol{\pi}}(I_k)E^{\boldsymbol{\pi}}(\tilde{a}) = \operatorname{cov}^{\boldsymbol{\pi}}(I_k, \tilde{a}).$$

It follows that a positive increment $\mathrm{d}W_t$ of the innovations process leads to an increase in π_t^k if the rvs I_k and $\tilde{a}$ are positively correlated under the measure $\boldsymbol{\pi}_t$.

The Kushner–Stratonovich equation (17.64) lends itself to simulation. In particular, the equation can be used to generate trajectories of $(\boldsymbol{\pi}_t)$ and of the default intensities $\lambda_{t,i} = \sum_{l=1}^{K} \pi_t^k \gamma_i(v_k, t)$, $1 \leqslant i \leqslant m$. Details are given in Frey and Schmidt (2011).

The extension of the investor information to the larger information set $(\mathcal{G}_t)$ leads to rich and realistic dynamics for default intensities incorporating both random fluctuations of credit spreads between defaults and default contagion. This is illustrated in Figure 17.3, where we plot a typical simulated trajectory of the default intensity. The fluctuation of the intensity between defaults as well as the contagion effects at default times (e.g. around $t = 600$) can be clearly seen.

Pricing of credit derivatives. All commonly encountered credit-risky instruments, credit derivatives and credit value adjustments can be categorized into the following two classes.

Basic credit products: this class comprises products where the cash-flow stream depends on the default history of the underlying portfolio and is therefore $(\mathcal{H}_t)$-adapted. Examples are corporate bonds, CDSs and CDOs.

Options on traded credit products: this class contains derivatives whose pay-off depends on the future *market value* of basic credit products. Examples include options on corporate bonds, options on CDS indices, and credit value adjustments for credit derivatives, as discussed in Section 17.2.

The pricing methodology for these product classes differs and we discuss them separately. We begin with basic credit products. Denote the associated stream of discounted future cash flows by the $\mathcal{H}_T$-measurable rv $\Pi(t, T)$. Then, using risk-neutral pricing the price at time t of the basic credit product is given by $V_t = E^Q(\Pi(t, T) \mid \mathcal{G}_t)$. Recall that $\mathcal{G}_t = \mathcal{H}_t \vee \mathcal{F}_t^Z$. Using a similar approach to the derivation of (17.43) we obtain

$$V_t = \sum_{k=1}^{K} h(t, v_k, \boldsymbol{Y}_t)\pi_t^k, \tag{17.65}$$

where $h(t, V, \boldsymbol{Y}_t) = E^Q(\Pi(t, T) \mid \tilde{\mathcal{G}}_t)$, the hypothetical value of the claim for known V. Note that V_t depends only on $\boldsymbol{Y}_t$, $\boldsymbol{\pi}_t$ and the function $h(\cdot)$; the precise form of the function $a(\cdot)$ in (17.61) is irrelevant for the pricing of basic credit products.

Next we consider options on traded credit products. Assume that N basic credit products—such as CDSs, CDO tranches or index swaps—are traded on the market, and denote their ex-dividend prices at time t by $V_{t,1}, \ldots, V_{t,N}$. The pay-off of an option on a traded credit product then takes the form $I_{\{\tilde{T} \leqslant T\}} g(\boldsymbol{Y}_{\tilde{T}}, V_{\tilde{T},1}, \ldots, V_{\tilde{T},N})$, where $\tilde{T}$ is a $(\mathcal{G}_t)$ stopping time at which the pay-out is made and T is the maturity of the contract. A prime example is the credit value adjustment for a CDS. Denote by (V_t^{CDS}) the market value of the counterparty-risk-free CDS. It was shown in Proposition 17.1 that the credit value adjustment is an option on the CDS with $\tilde{T} = T_1$ and pay-off

$$g(\boldsymbol{Y}_{T_1}, V_{T_1}^{\mathrm{CDS}}) = (1 - Y_{T_1}^R)(1 - Y_{T_1}^B)(V_{T_1}^{\mathrm{CDS}})^+.$$

Another example of an option on a traded credit product would be an option on a CDS index, as discussed in Frey and Schmidt (2011).

It can be shown that the process $(\boldsymbol{Y}_t, \boldsymbol{\pi}_t)$ is a Markov process in the investor filtration so that the price at time t of an option on a traded credit product is a function of Y_t and π_t of the form

$$E\left(\exp\left(-\int_t^{\tilde{T}} r(s)\,\mathrm{d}s\right) I_{\{\tilde{T} \leqslant T\}} g(\boldsymbol{Y}_{\tilde{T}}, V_{\tilde{T},1}, \ldots, V_{\tilde{T},N}) \,\middle|\, \mathcal{G}_t\right) = I_{\{\tilde{T} > t\}} g(t, \boldsymbol{Y}_t, \boldsymbol{\pi}_t). \tag{17.66}$$

The actual evaluation of the function g is usually based on Monte Carlo methods. Note that for an option on traded credit products the function g will in general depend on the entire dynamics of the process $(\boldsymbol{Y}_t, \boldsymbol{\pi}_t)$, as we will see in the analysis of credit value adjustments in the next section.

Calibration. Calibration methodologies are a crucial part of the model of Frey and Schmidt (2012) for the following reason. Recall that we view the process (Z_t) generating the filtration $\mathbb{G}$ as an abstract model-building device that is not directly linked to observable quantities. Consequently, the process $(\boldsymbol{\pi}_t)$ is not directly observable by investors. On the other hand, pricing formulas in our model depend on the values of $\boldsymbol{Y}_t$ and $\boldsymbol{\pi}_t$. Since pricing formulas need to be evaluated in terms of publicly available information, a key point in the application of the model is therefore to determine the realization of $\boldsymbol{\pi}_t$ at time t from prices of basic traded credit products observed at that date.

Suppose that N basic credit products are traded at time t at market prices p_n^*, $1 \leqslant n \leqslant N$, and denote by $h_n(t, V, \boldsymbol{Y}_t)$ the value of contract n in the artificial model where V is known. We then need to find some probability vector $\boldsymbol{\pi} = (\pi^1, \ldots, \pi^K)'$ such that the model prices $V_t^n(\boldsymbol{\pi}) = \sum_{k=1}^K \pi^k h_n(t, v_k, \boldsymbol{Y}_t)$ are close to the observed prices p_n^* for all n. The reader will note that this problem is similar to the calibration problem arising in the implied copula framework of Section 12.3.3, and we refer to that section for details of algorithms and numerical results.

17.4.4 Collateralized Credit Value Adjustments and Contagion Effects

We now study the impact of different price dynamics on the size of credit value adjustments and on the performance of collateralization strategies for a single-name CDS. We are particularly interested in the influence of contagion. To see that contagion might be relevant for the performance of collateralization strategies, consider the scenario in which the protection seller defaults before the maturity of the CDS. In such a case contagion might lead to a substantial increase in the credit spread of the reference entity (the firm on which the CDS is written) and hence in turn to a much higher replacement value for the CDS. In standard collateralization strategies this is taken into account in a fairly crude way, and the amount of collateral posted before the default may well be insufficient to replace the CDS.

Our exposition is based on Frey and Rösler (2014). We use the model with incomplete information from the previous section for our analysis. Slightly extending the set-up of that section, we assume that the factor is a finite-state Markov chain (Ψ_t) (details are not relevant for our discussion here). We consider two versions of the model that differ with respect to the amount of information that is available to investors. In the version with full information it is assumed that the process (Ψ_t) is observable, so there are no contagion effects. In the version with incomplete information investors observe only the process $Z_t = \int_0^t a(\Psi_s)\,\mathrm{d}s + B_t$ for $a(\psi) = c \ln \bar{\gamma}(\psi)$ (and, of course, the default history). As we have seen before, under incomplete information there is default contagion caused by the updating of the conditional distribution of (Ψ_t) at default times.

In order to compute credit value adjustments and to measure the performance of collateralization strategies, we use the bilateral collateralized credit value adjustment (BCCVA) introduced in Section 17.2.2. The actual computation of credit value adjustments is mostly carried out using Monte Carlo simulation. The numerical experiments that follow are based on a Markov chain (Ψ_t) with $K = 8$ states

Table 17.2. Results of model calibration for the case study on credit value adjustments.

State	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
π_0	0.0810	0.0000	0.2831	0.0548	0.0000	0.0000	0.0000	0.5811
γ_B	0.0000	0.0010	0.0027	0.0040	0.0050	0.0059	0.0091	0.0195
γ_R	0.0031	0.0669	0.1187	0.1482	0.1687	0.1855	0.2393	0.3668
γ_S	0.0007	0.0245	0.0482	0.0627	0.0732	0.0818	0.1108	0.1840

$v_1, \dots, v_8$, where v_1 is the best state (lowest default probabilities for all firms considered) and v_8 is the worst state. In order to calibrate the model we assume that the protection buyer B has a credit spread of 50 bp, the reference entity R has a credit spread of 1000 bp, and the protection seller S has a credit spread of 500 bp, so that B is of far better credit quality than S; moreover, the default correlations of the three firms are fixed to be $\rho_{BR} = 2.0\%$, $\rho_{BS} = 1.5\%$ and $\rho_{RS} = 5\%$. The results of the calibration exercise are given in Table 17.2. Note that the default intensities at any fixed time t are comonotonic random variables.

Now we present the results of the simulation study. We begin with an analysis of the performance of popular collateralization strategies, for which we refer the reader to Section 17.2.2. The market value of the counterparty-risk-free CDS referencing R will be denoted by (V_t^{CDS}). We compare market-value collateralization where $C_t^{\mathrm{market}} = V_t^{\mathrm{CDS}}$, $t \leqslant T$, and threshold collateralization where

$$C_t^{M_1,M_2} = ((V_t^{\mathrm{CDS}})^+ - M_1) I_{\{(V_t^{\mathrm{CDS}})^+ > M_1\}} - ((V_t^{\mathrm{CDS}})^- - M_2) I_{\{(V_t^{\mathrm{CDS}})^- > M_2\}}$$

for thresholds M_1 and M_2. Note that market-value collateralization can be viewed as threshold collateralization with $M_1 = M_2 = 0$.

Numerical results illustrating the performance of these collateralization strategies are given in Table 17.3. We see that market value collateralization is very effective in the model with complete information. The performance of threshold collateralization is also satisfactory, as can be seen by comparing the CCVA for a small positive threshold with the CCVA for the uncollateralized case. The CDVA is quite low for all collateralization strategies, since in the chosen example B is of far better credit quality than S, so $Q(\xi_1 = B)$ is very small. Under incomplete information, on the other hand, the performance of market-value and threshold collateralization is not entirely satisfactory. In particular, even for $M_1 = M_2 = 0$ the CCVA is quite high compared with the case of full information. The main reason for this is the fact that, because of the contagion effects, threshold collateralization systematically underestimates the market value of the CDS at T_1, which leads to losses for the protection buyer. Note that the joint distribution of the default times is the same in the two versions of the model, so the differences in the sizes of the value adjustments and in the performance of the collateralization strategies can be attributed to the different dynamics of credit spreads (contagion or no contagion) in the two model variants. Our case study therefore clearly shows that the dynamics of credit spreads matters in the management of counterparty credit risk.

Finally, we show that for the given parameter values there is clear evidence for *wrong-way risk*. To demonstrate this we compare the correct value adjustments to

Table 17.3. Value adjustments in the model with complete information and in the model with incomplete information under threshold collateralization and market-value collateralization ($M_1 = M_2 = 0$). The nominal of the CDS is normalized to 1; all numbers are in basis points. In the last row we also report the value adjustment corresponding to the simplified value adjustment formulas (17.11) and (17.12).

	Full information			Partial information		
Threshold	CCVA	CDVA	BCCVA	CCVA	CDVA	BCCVA
$M_1 = M_2 = 0$	0	0	0	35	0	35
$M_1 = M_2 = 0.02$	16	0	15	45	0	45
$M_1 = M_2 = 0.05$	38	1	37	60	0	60
No collateralization with						
(i) correct formula	93	1	92	83	1	82
(ii) simplified formula	68	6	62	54	4	49

the value adjustments computed via the simplified formulas (17.11) and (17.12), both in the case of no collateralization. The results are given in the last two rows of Table 17.3. It turns out that in both versions of the model the value adjustments computed via the correct formula are substantially larger than the adjustments computed from the simplified formulas. This suggests that in situations where there is a non-negligible default correlation between the protection seller and the reference entity, the simplified formulas may not be appropriate.

Notes and Comments

There is a large literature on credit risk models with incomplete information. Kusuoka (1999), Duffie and Lando (2001), Giesecke and Goldberg (2004), Jarrow and Protter (2004), Coculescu, Geman and Jeanblanc (2008), Frey and Schmidt (2009), Cetin (2012) and Frey, Rösler and Lu (2014) all consider structural models in the spirit of the Merton model, where the values of assets and/or liabilities are not directly observable. The last three of these papers use stochastic filtering techniques to study structural models under incomplete information.

Turning to reduced-form models with incomplete information, the models of Schönbucher (2004) and Collin-Dufresne, Goldstein and Helwege (2010) are similar to the model considered in Section 17.4.2. Both papers point out that the successive updating of the conditional distribution of the unobserved factor in reaction to incoming default observations generates information-driven default contagion. Duffie et al. (2009) assume that the unobservable factor (Ψ_t) is given by an Ornstein–Uhlenbeck process. Their paper contains interesting empirical results; in particular, the analysis provides strong support for the assertion that an unobservable stochastic process driving default intensities (a so-called *dynamic frailty*) is needed on top of observable covariates in order to explain the clustering of defaults in historical data. The link between stochastic filtering and reduced-form models is explored extensively in Frey and Runggaldier (2010) and Frey and Schmidt (2012). Our analysis

in Section 17.4.3 is largely based on these papers. The numerical results on the performance of collateralization strategies are due to Frey and Rösler (2014). A survey of credit risk modelling under incomplete information can be found in Frey and Schmidt (2011).

Appendix

A.1 Miscellaneous Definitions and Results

A.1.1 Type of Distribution

Definition A.1 (equality in type). Two rvs V and W (or their distributions) are said to be of the same type if there exist constants $a > 0$ and $b \in \mathbb{R}$ such that $V \stackrel{\mathrm{d}}{=} aW + b$.

In other words, distributions of the same type are obtained from one another by location and scale transformations.

A.1.2 Generalized Inverses and Quantiles

Let T be an *increasing* function, i.e. a function satisfying $y > x \implies T(y) \geqslant T(x)$, with strict inequality on the right-hand side for some pair $y > x$. An increasing function may therefore have *flat sections*; if we want to rule this out, we stipulate that T is *strictly increasing*, so $y > x \iff T(y) > T(x)$. We first note some useful facts concerning what happens when increasing transformations are applied to rvs.

Lemma A.2.

(i) *If X is an rv and T is increasing, then $\{X \leqslant x\} \subset \{T(X) \leqslant T(x)\}$ and*

$$P(T(X) \leqslant T(x)) = P(X \leqslant x) + P(T(X) = T(x),\ X > x). \qquad \text{(A.1)}$$

(ii) *If F is the df of the rv X, then $P(F(X) \leqslant F(x)) = P(X \leqslant x)$.*

The second statement follows from (A.1) by noting that, for any x, the event given by $\{F(X) = F(x),\ X > x\}$ corresponds to a flat piece of the df F and therefore has zero probability mass.

The generalized inverse of an increasing function T is defined to be $T^{\leftarrow}(y) = \inf\{x\colon T(x) \geqslant y\}$, where we use the convention $\inf \emptyset = \infty$. Strictly speaking, this generalized inverse is known as the *left-continuous* generalized inverse. The following basic properties may be verified quite easily.

Proposition A.3 (properties of the generalized inverse). *For T increasing, the following hold.*

(i) *$T^{\leftarrow}$ is an increasing, left-continuous function.*

(ii) *T is continuous $\iff$ $T^{\leftarrow}$ is strictly increasing.*

(iii) *T is strictly increasing $\iff$ $T^{\leftarrow}$ is continuous.*

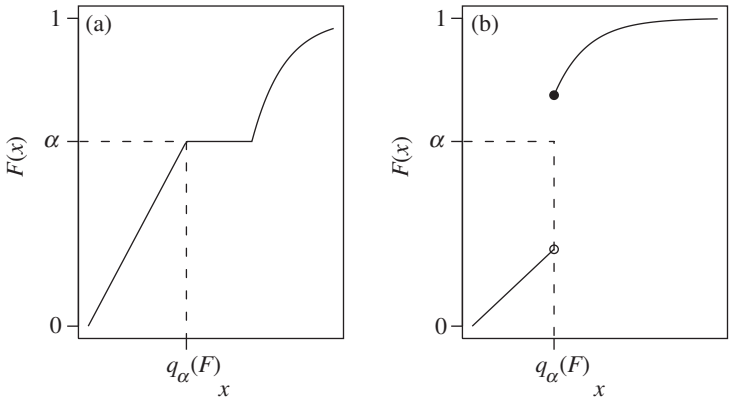

Figure A.1. Calculation of quantiles in tricky cases. The first case (a) is a continuous df, but the flat piece corresponds to an interval with zero probability mass. In the second case (b) there is an atom of probability mass such that, for X with df F, we have $P(X = q_\alpha(F)) > 0$.

For the remaining properties assume additionally that $-\infty < T^{\leftarrow}(y) < \infty$.

(iv) *If* T *is right continuous,* $T(x) \geqslant y \iff T^{\leftarrow}(y) \leqslant x$.

(v) $T^{\leftarrow} \circ T(x) \leqslant x$.

(vi) T *is right continuous* $\Longrightarrow T \circ T^{\leftarrow}(y) \geqslant y$.

(vii) T *is strictly increasing* $\Longrightarrow T^{\leftarrow} \circ T(x) = x$.

(viii) T *is continuous* $\Longrightarrow T \circ T^{\leftarrow}(y) = y$.

We apply the idea of generalized inverses to distribution functions. If F is a df, then the generalized inverse $F^{\leftarrow}$ is known as the quantile function of F. In this case, for $\alpha \in (0, 1)$, we also use the alternative notation $q_\alpha(F) = F^{\leftarrow}(\alpha)$ for the α-quantile of F. Figure A.1 illustrates the calculation of quantiles in two tricky cases.

In general, since a df need not be strictly increasing (part (a) of the figure), we have $F^{\leftarrow} \circ F(x) \leqslant x$, by Proposition A.3 (v). But the values x, where $F^{\leftarrow} \circ F(x) \neq x$, correspond to flat pieces and have zero probability mass. That is, we have the following useful fact.

Proposition A.4. *If* X *is an rv with df* F*, then* $P(F^{\leftarrow} \circ F(X) = X) = 1$.

The following proposition shows how quantiles can be computed for transformed random variables and it is used in Section 7.2.1 to prove the comonotone additivity of value-at-risk.

Proposition A.5. *For a random variable* X *and an increasing, left-continuous function* T*, the quantile function of* $T(X)$ *is given by*

$$F^{\leftarrow}_{T(X)}(\alpha) = T(F^{\leftarrow}_X(\alpha)).$$

Proof. Since $F_{T(X)}$ is a right-continuous function, Proposition A.3 (iv) tells us that for any point x we have the equivalence

$$F^{\leftarrow}_{T(X)}(\alpha) \leqslant x \iff F_{T(X)}(x) \geqslant \alpha. \tag{A.2}$$

Let $y_0(x) = \sup\{y : T(y) \leqslant x\}$ and observe that the left continuity of T implies

$$T(y) \leqslant x \iff y \leqslant y_0(x). \tag{A.3}$$

From (A.3) we conclude that the events $\{T(X) \leqslant x\}$ and $\{X \leqslant y_0(x)\}$ are identical, from which it follows that

$$F_{T(X)}(x) \geqslant \alpha \iff F_X(y_0(x)) \geqslant \alpha. \tag{A.4}$$

We also have that

$$F_X(y_0(x)) \geqslant \alpha \overset{\text{(a)}}{\iff} F_X^{\leftarrow}(\alpha) \leqslant y_0(x) \overset{\text{(b)}}{\iff} T(F_X^{\leftarrow}(\alpha)) \leqslant x, \tag{A.5}$$

where we again use Proposition A.3 (iv) to establish (a) and we use (A.3) to establish (b). The equivalences (A.2), (A.4) and (A.5) together with the fact that x is an arbitrary point prove the lemma. □

A.1.3 Distributional Transform

The distributional transform can be used to prove the general version of Sklar's Theorem (Theorem 7.3), an insight due to Rüschendorf (2009). For a random variable X with distribution function F, define the modified distribution function by

$$\tilde{F}(x, \lambda) = P(X < x) + \lambda P(X = x), \quad \lambda \in [0, 1]. \tag{A.6}$$

Obviously we have $\tilde{F}(x, \lambda) = F(x-) + \lambda(F(x) - F(x-))$, and if F is continuous at x, then $\tilde{F}(x, \lambda) = F(x)$, but in general $F(x-) \leqslant \tilde{F}(x, \lambda) \leqslant F(x)$.

The distributional transform of X is given by

$$U := \tilde{F}(X, V), \tag{A.7}$$

where $V \sim U(0, 1)$ is a uniform rv independent of X and $\tilde{F}(x, \lambda)$ is as in (A.6). The following result shows how the distributional transform generalizes the probability transform of Proposition 7.2.

Proposition A.6. *For an rv X with df F let $U = \tilde{F}(X, V)$ be the distributional transform of X. Then $U \sim U(0, 1)$ and $X = F^{\leftarrow}(U)$ almost surely.*

Proof. For given u we compute $P(U \leqslant u)$. Let $q(u) = P(X < F^{\leftarrow}(u))$ and $p(u) = P(X = F^{\leftarrow}(u))$ and observe that

$$\{\tilde{F}(X, V) \leqslant u\} = \{X < F^{\leftarrow}(u)\} \cup \{X = F^{\leftarrow}(u),\ q(u) + Vp(u) \leqslant u\}.$$

There are two cases to consider: either F is continuous at the u-quantile $F^{\leftarrow}(u)$, in which case $p(u) = 0$ and $q(u) = u$; or F has a jump at the u-quantile, in which case $p(u) > 0$. In the former case

$$P(U \leqslant u) = P(X < F^{\leftarrow}(u)) = q(u) = u,$$

and in the latter case

$$P(U \leqslant u) = q(u) + p(u)P\left(V \leqslant \frac{u - q(u)}{p(u)}\right) = u,$$

which proves that U has a uniform distribution.

To prove the second assertion fix $\omega \in \Omega$ and let $x = X(\omega)$ and $u = U(\omega) = \tilde{F}(X(\omega), V(\omega))$. Clearly we have $F(x-) \leqslant u \leqslant F(x)$. If $F(x-) < u \leqslant F(x)$ then $F^{\leftarrow}(u) = x$, but if $F(x-) = u$ then we may have that $F^{\leftarrow}(u) < x$. However, the values of ω for which the latter situation may occur comprise a null set. □

A.1.4 Karamata's Theorem

The following result for regularly varying functions is used in Chapter 5. For more details see Bingham, Goldie and Teugels (1987). Essentially, the result says that the slowly varying function can be taken outside the integral as if it were a constant. Note that the symbol "$\sim$" indicates asymptotic equality here, i.e. if we write $a(x) \sim b(x)$ as $x \to x_0$, we mean $\lim_{x \to x_0} a(x)/b(x) = 1$.

Theorem A.7 (Karamata's Theorem). *Let L be a slowly varying function that is locally bounded in $[x_0, \infty)$ for some $x_0 \geqslant 0$. Then,*

(a) *for $\kappa > -1$,* $\displaystyle\int_{x_0}^{x} t^{\kappa} L(t)\,\mathrm{d}t \sim \frac{1}{\kappa+1} x^{\kappa+1} L(x),\ x \to \infty,$

(b) *for $\kappa < -1$,* $\displaystyle\int_{x}^{\infty} t^{\kappa} L(t)\,\mathrm{d}t \sim -\frac{1}{\kappa+1} x^{\kappa+1} L(x),\ x \to \infty.$

A.1.5 Supporting and Separating Hyperplane Theorems

The following result on the existence of supporting and separating hyperplanes is needed at various points in Chapter 8. For further information we refer to Appendix B2 in Bertsekas (1999), Section 2.5 in Boyd and Vandenberghe (2004) and Chapter 11 of Rockafellar (1970).

Proposition A.8. *Consider a convex set $C \subset \mathbb{R}^n$ and some $\boldsymbol{x}_0 \in \mathbb{R}^n$.*

(a) *If $\boldsymbol{x}_0$ is not an interior point of C, there exists a supporting hyperplane for C through $\boldsymbol{x}_0$, i.e. there is some $\boldsymbol{u} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ such that $\boldsymbol{u}'\boldsymbol{x}_0 \geqslant \sup\{\boldsymbol{u}'\boldsymbol{x} : \boldsymbol{x} \in C\}$.*

(b) *If $\boldsymbol{x}_0$ does not belong to the closure $\bar{C}$ of C, one has strict separation, i.e. there is some $\boldsymbol{u} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ such that $\boldsymbol{u}'\boldsymbol{x}_0 > \sup\{\boldsymbol{u}'\boldsymbol{x} : \boldsymbol{x} \in \bar{C}\}$.*

A.2 Probability Distributions

The gamma and beta functions appear in the definitions of a number of these distributions. The gamma function is

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} \mathrm{e}^{-x}\,\mathrm{d}x, \quad \alpha > 0, \tag{A.8}$$

and it satisfies the useful recursive relationship $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$. The beta function is

$$\beta(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,\mathrm{d}x, \quad a, b > 0. \tag{A.9}$$

It is related to the gamma function by $\beta(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$.

A.2.1 Beta

The rv X has a beta distribution, written $X \sim \text{Beta}(a, b)$, if its density is

$$f(x) = \frac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1,\ a, b > 0, \tag{A.10}$$

where $\beta(a, b)$ is the beta function in (A.9). The uniform distribution $U(0, 1)$ is obtained as a special case when $a = b = 1$. The mean and variance of the distribution are, respectively, $E(X) = a/(a + b)$ and $\text{var}(X) = (ab)/((a + b + 1)(a + b)^2)$.

A.2.2 Exponential

The rv X has an exponential distribution, written $X \sim \text{Exp}(\lambda)$, if its density is

$$f(x) = \lambda \mathrm{e}^{-\lambda x}, \quad x > 0,\ \lambda > 0. \tag{A.11}$$

The mean of this distribution is $E(X) = \lambda^{-1}$ and the variance is $\text{var}(X) = \lambda^{-2}$.

A.2.3 F

The rv X has an F distribution, written $X \sim F(\nu_1, \nu_2)$, if its density is

$$f(x) = \frac{1}{\beta(\frac{1}{2}\nu_1, \frac{1}{2}\nu_2)} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \frac{x^{(\nu_1-2)/2}}{(1 + \nu_1 x/\nu_2)^{(\nu_1+\nu_2)/2}}, \quad x > 0,\ \nu_1, \nu_2 > 0. \tag{A.12}$$

The mean of this distribution is $E(X) = \nu_2/(\nu_2 - 2)$ provided that $\nu_2 > 2$. Provided that $\nu_2 > 4$, the variance is

$$\text{var}(X) = 2\left(\frac{\nu_2}{\nu_2 - 2}\right)^2 \frac{\nu_1 + \nu_2 - 2}{\nu_1(\nu_1 - 4)}.$$

A.2.4 Gamma

The rv X has a gamma distribution, written $X \sim \text{Ga}(\alpha, \beta)$, if its density is

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\beta x}, \quad x > 0,\ \alpha > 0,\ \beta > 0, \tag{A.13}$$

where $\Gamma(\alpha)$ denotes the gamma function in (A.8). Using the recursive property of the gamma function, the mean and variance of the gamma distribution are easily calculated to be $E(X) = \alpha/\beta$ and $\text{var}(X) = \alpha/\beta^2$. For fitting a multivariate t distribution using the EM approach of Section 15.1.1 it is also useful to know that $E(\ln X) = \psi(\alpha) - \ln(\beta)$, where $\psi(k) = \mathrm{d}\ln(\Gamma(k))/\mathrm{d}k$ is the digamma or psi function.

An exponential distribution is obtained in the special case when $\alpha = 1$. If $X \sim \text{Ga}(\alpha, \beta)$ and $k > 0$, then $kX \sim \text{Ga}(\alpha, \beta/k)$. For two independent gamma variates $X_1 \sim \text{Ga}(\alpha_1, \beta)$ and $X_2 \sim \text{Ga}(\alpha_2, \beta)$, we have that $X_1 + X_2 \sim \text{Ga}(\alpha_1 + \alpha_2, \beta)$. Note also that, if $X \sim \text{Ga}(\frac{1}{2}\nu, \frac{1}{2})$, then X has a chi-squared distribution with ν degrees of freedom, also written $X \sim \chi^2_\nu$.

A.2.5 Generalized Inverse Gaussian

The rv X has a generalized inverse Gaussian (GIG) distribution, written $X \sim N^-(\lambda, \chi, \psi)$, if its density is

$$f(x) = \frac{\chi^{-\lambda}(\sqrt{\chi\psi})^\lambda}{2K_\lambda(\sqrt{\chi\psi})} x^{\lambda-1} \exp(-\tfrac{1}{2}(\chi x^{-1} + \psi x)), \quad x > 0, \tag{A.14}$$

where K_λ denotes a modified Bessel function of the third kind with index λ and the parameters satisfy $\chi > 0$, $\psi \geqslant 0$ if $\lambda < 0$; $\chi > 0$, $\psi > 0$ if $\lambda = 0$; and $\chi \geqslant 0$, $\psi > 0$ if $\lambda > 0$. For more on this Bessel function see Abramowitz and Stegun (1965).

The GIG density actually contains the gamma and inverse gamma densities as special limiting cases, corresponding to $\chi = 0$ and $\psi = 0$, respectively. In these cases (A.14) must be interpreted as a limit, which can be evaluated using the asymptotic relations $K_\lambda(x) \sim \Gamma(\lambda)2^{\lambda-1}x^{-\lambda}$ as $x \to 0+$ for $\lambda > 0$ and $K_\lambda(x) \sim \Gamma(-\lambda)2^{-\lambda-1}x^{\lambda}$ as $x \to 0+$ for $\lambda < 0$. The fact that $K_\lambda(x) = K_{-\lambda}(x)$ is also useful. In this way it can be verified that, for $\lambda > 0$ and $\chi = 0$, $X \sim \mathrm{Ga}(\lambda, \frac{1}{2}\psi)$. If $\lambda < 0$ and $\psi = 0$, we have $X \sim \mathrm{Ig}(-\lambda, \frac{1}{2}\chi)$. The case $\lambda = -\frac{1}{2}$ is known as the inverse Gaussian distribution. Note that, in general, if $Y \sim N^-(\lambda, \chi, \psi)$, then $1/Y \sim N^-(-\lambda, \psi, \chi)$.

For the non-limiting case when $\chi > 0$ and $\psi > 0$ it may be calculated that

$$E(X^\alpha) = \left(\frac{\chi}{\psi}\right)^{\alpha/2} \frac{K_{\lambda+\alpha}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}, \quad \alpha \in \mathbb{R}, \tag{A.15}$$

$$E(\ln X) = \left.\frac{\mathrm{d}E(X^\alpha)}{\mathrm{d}\alpha}\right|_{\alpha=0}. \tag{A.16}$$

A.2.6 Inverse Gamma

The rv X has an inverse gamma distribution, written $X \sim \mathrm{Ig}(\alpha, \beta)$, if its density is

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} \mathrm{e}^{-\beta/x}, \quad x > 0,\ \alpha > 0,\ \beta > 0. \tag{A.17}$$

Note that if $Y \sim \mathrm{Ga}(\alpha, \beta)$, then $1/Y \sim \mathrm{Ig}(\alpha, \beta)$. Provided that $\alpha > 1$, the mean is $E(X) = \beta/(\alpha - 1)$, and provided that $\alpha > 2$, the variance is $\mathrm{var}(X) = \beta^2/((\alpha-1)^2(\alpha-2))$. Moreover, $E(\ln X) = \ln(\beta) - \psi(\alpha)$.

A.2.7 Negative Binomial

The rv N has a negative binomial distribution with parameters $\alpha > 0$ and $0 < p < 1$, written $N \sim \mathrm{NB}(\alpha, p)$, if its probability mass function is

$$P(N = k) = \binom{\alpha + k - 1}{k} p^\alpha (1-p)^k, \quad k = 0, 1, 2, \ldots, \tag{A.18}$$

where $\binom{x}{k}$ for $x \in \mathbb{R}$ and $k \in \mathbb{N}_0$ denotes an extended binomial coefficient defined by $\binom{x}{0} = 1$ and

$$\binom{x}{k} = \frac{x(x-1)\cdots(x-k+1)}{k!}, \quad k > 0.$$

The moments of this distribution are

$$E(N) = \alpha(1-p)/p \quad \text{and} \quad \operatorname{var}(N) = \alpha(1-p)/p^2.$$

For $\alpha = r \in \mathbb{N}$ the rv $N + r$ represents the waiting time until the rth success in independent Bernoulli trials with success probability p, i.e. the total number of trials that are required until we have r successes. For $\alpha = 1$ the rv $N + 1$ is said to have a geometric distribution.

A.2.8 Pareto

The rv X has a Pareto distribution, written $X \sim \mathrm{Pa}(\alpha, \kappa)$, if its df is

$$F(x) = 1 - \left(\frac{\kappa}{\kappa + x}\right)^{\alpha}, \quad \alpha, \kappa > 0,\ x \geqslant 0. \tag{A.19}$$

Provided that $\alpha > n$, the moments of this distribution are given by

$$E(X^n) = \frac{\kappa^n n!}{\prod_{i=1}^{n}(\alpha - i)}.$$

A.2.9 Stable

The rv X has an α-stable distribution, written $X \sim \mathrm{St}(\alpha, \beta, \gamma, \delta)$, if its characteristic function is

$$\begin{aligned}\phi(t) &= E\mathrm{e}^{\mathrm{i}tX} \\ &= \begin{cases} \exp(-\gamma^\alpha |t|^\alpha (1 - \mathrm{i}\beta \operatorname{sign}(t) \tan(\pi\alpha/2)) + \mathrm{i}\delta t), & \alpha \neq 1, \\ \exp(-\gamma|t|(1 + \mathrm{i}\beta \operatorname{sign}(t)(2/\pi)\ln|t|) + \mathrm{i}\delta t), & \alpha = 1, \end{cases}\end{aligned} \tag{A.20}$$

where $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\gamma > 0$ and $\delta \in \mathbb{R}$. Note that there are various alternative parametrizations of the stable distributions and we use a parametrization of Nolan (2003, Definition 1.8). The case $X \sim \mathrm{St}(\alpha, 1, \gamma, 0)$ for $\alpha < 1$ gives a distribution on the positive half-axis, which we refer to as a positive stable distribution.

A simulation algorithm for a standardized variate $Z \sim \mathrm{St}(\alpha, \beta, 1, 0)$ is given in Nolan (2003, Theorem 1.19). In the case where $\alpha \neq 1$, $X = \delta + \gamma Z$ has a $\mathrm{St}(\alpha, \beta, \gamma, \delta)$ distribution; the case $\alpha = 1$ is more complicated.

A.3 Likelihood Inference

This appendix summarizes the mechanics of performing likelihood inference, but omits theoretical details. A good starting reference for the theory is Casella and Berger (2002), which we refer to in this appendix where relevant. Other useful books include Serfling (1980), Lehmann (1983, 1986), Schervish (1995) and Stuart, Ord and Arnold (1999), all of which give details concerning the famous *regularity conditions* that are required for the asymptotic statements.

A.3.1 Maximum Likelihood Estimators

Suppose that the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)'$ has joint probability density (or mass function) in some parametric family $f_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\theta})$, indexed by a parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ in a parameter space $\boldsymbol{\Theta}$. We consider our data to be a realization of $\boldsymbol{X}$ for some unknown value of θ.

The *likelihood function* for the parameter vector $\boldsymbol{\theta}$ given the data is $L(\boldsymbol{\theta}; \boldsymbol{X}) = f_{\boldsymbol{X}}(\boldsymbol{X}; \boldsymbol{\theta})$, and the maximum likelihood estimator (MLE) $\hat{\boldsymbol{\theta}}$ is the value of $\boldsymbol{\theta}$ maximizing $L(\boldsymbol{\theta}; \boldsymbol{X})$, or equivalently the value maximizing the *log-likelihood function* $l(\boldsymbol{\theta}; \boldsymbol{X}) = \ln L(\boldsymbol{\theta}; \boldsymbol{X})$. We will also write this estimator as $\hat{\boldsymbol{\theta}}_n$ when we want to emphasize its dependence on the sample size n.

For large n we expect that the estimate $\hat{\boldsymbol{\theta}}_n$ will be *close* to the true value $\boldsymbol{\theta}$, and various well-known asymptotic results give information about the quality of the estimator in large samples. In describing these results we consider the classical situation where $\boldsymbol{X}$ is assumed to be a vector of *iid components* with univariate density f, so

$$\ln L(\boldsymbol{\theta}; \boldsymbol{X}) = \ln \prod_{i=1}^{n} f(X_i; \boldsymbol{\theta}) = \sum_{i=1}^{n} \ln L(\boldsymbol{\theta}; X_i).$$

A.3.2 Asymptotic Results: Scalar Parameter

We consider the case when $p = 1$ and we have a single parameter θ. Under suitable *regularity conditions* (see, for example, Casella and Berger 2002, p. 516), $\hat{\theta}_n$ may be shown to be a *consistent* estimator of θ (i.e. tending to θ in probability as the sample size n is increased). Notable among the regularity conditions are that θ should be an *identifiable* parameter ($\theta \neq \tilde{\theta} \Rightarrow f(x; \theta) \neq f(x; \tilde{\theta})$), that the true parameter θ should be an interior point of the parameter space Θ, and that the support of $f(x; \theta)$ should not depend on θ.

Under stronger regularity conditions (see again Casella and Berger 2002, p. 516), $\hat{\theta}_n$ may be shown to be an *asymptotically efficient* estimator of θ, so it satisfies

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathrm{d}} N(0, I(\theta)^{-1}), \tag{A.21}$$

where $I(\theta)$ denotes the *Fisher information of an observation*, defined by

$$I(\theta) = E\left(\frac{\partial}{\partial \theta} \ln L(\theta; X) \right)^2. \tag{A.22}$$

Under the regularity conditions, the Fisher information can generally also be calculated as

$$I(\theta) = -E\left(\frac{\partial^2}{\partial \theta^2} \ln L(\theta; X) \right). \tag{A.23}$$

Asymptotic efficiency entails both asymptotic normality and consistency. Moreover, it implies that, in a large enough sample, $\mathrm{var}(\hat{\theta}) \approx 1/(nI(\theta))$, where the right-hand side is the so-called Cramér–Rao lower bound, which is a lower bound for the variance of an unbiased estimator of θ constructed from an iid sample $X_1, \ldots, X_n$. The MLE is efficient in the sense that it attains this lowest possible bound asymptotically.

A.3.3 Asymptotic Results: Vector of Parameters

When $p > 1$ and we have a vector of parameters to estimate, similar results apply. The MLE $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ is asymptotically efficient in the sense that, as $n \to \infty$ and under suitable regularity conditions,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathrm{d}} N_p(\mathbf{0}, I(\boldsymbol{\theta})^{-1}), \tag{A.24}$$

where $I(\boldsymbol{\theta})$ denotes the expected Fisher information *matrix* for a single observation, given, in analogy to (A.22) and (A.23), by

$$I(\boldsymbol{\theta}) = E\left(\frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\boldsymbol{\theta}; X) \frac{\partial}{\partial \boldsymbol{\theta}'} \ln L(\boldsymbol{\theta}; X)\right) = -E\left(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \ln L(\boldsymbol{\theta}; X)\right).$$

The notation employed here should be taken to mean a matrix with components

$$I(\boldsymbol{\theta})_{ij} = E\left(\frac{\partial}{\partial \theta_i} \ln L(\boldsymbol{\theta}; X) \frac{\partial}{\partial \theta_j} \ln L(\boldsymbol{\theta}; X)\right) = -E\left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln L(\boldsymbol{\theta}; X)\right).$$

The convergence result (A.24) implies that, for n sufficiently large, we have

$$\hat{\boldsymbol{\theta}}_n \sim N_p(\boldsymbol{\theta}, n^{-1} I(\boldsymbol{\theta})^{-1}), \tag{A.25}$$

and this can be used to construct asymptotic confidence regions for $\boldsymbol{\theta}$ or intervals for any component θ_j. In practice, it is often easier to approximate $I(\boldsymbol{\theta})$ with the *observed Fisher information matrix*

$$\bar{I}(\boldsymbol{\theta}) = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \ln L(\boldsymbol{\theta}; X_i)$$

for whatever realization of $\boldsymbol{X}$ has been obtained. This should converge to the expected information matrix by the law of large numbers, and it has been suggested that in some situations this may even lead to more accurate inference (Efron and Hinkley 1978). In either case, the information matrices depend on the unknown parameters of the model and are usually estimated by taking $I(\hat{\boldsymbol{\theta}})$ or $\bar{I}(\hat{\boldsymbol{\theta}})$.

A.3.4 Wald Test and Confidence Intervals

From (A.25) we have that, for n sufficiently large,

$$Z := \frac{\hat{\theta}_j - \theta_j}{\mathrm{se}(\hat{\theta}_j)} \sim N(0, 1), \tag{A.26}$$

where $\mathrm{se}(\hat{\theta}_j)$ denotes an *asymptotic standard error* (an estimate of the asymptotic standard deviation) for $\hat{\theta}_j$, given by

$$\mathrm{se}(\hat{\theta}_j) = \sqrt{n^{-1} I(\hat{\boldsymbol{\theta}})^{-1}_{jj}} \text{ or } \sqrt{n^{-1} \bar{I}(\hat{\boldsymbol{\theta}})^{-1}_{jj}}.$$

Equation (A.26) can be used to test the null hypothesis $H_0 \colon \theta_j = \theta_{j,0}$ for some value of interest $\theta_{j,0}$ against the alternative $H_1 \colon \theta_j \neq \theta_{j,0}$. For an asymptotic test of size α we would reject H_0 if $|Z| \geqslant \Phi^{-1}(1 - \frac{1}{2}\alpha)$.

An asymptotic $100(1-\alpha)\%$ *confidence interval* for θ_j consists of those values $\theta_{j,0}$ for which the null hypothesis is not rejected and it is given by

$$(\hat{\theta}_j - \mathrm{se}(\hat{\theta}_j)\Phi^{-1}(1 - \tfrac{1}{2}\alpha), \hat{\theta}_j + \mathrm{se}(\hat{\theta}_j)\Phi^{-1}(1 - \tfrac{1}{2}\alpha)). \tag{A.27}$$

A.3.5 Likelihood Ratio Test and Confidence Intervals

Now consider testing the null hypothesis $H_0: \boldsymbol{\theta} \in \Theta_0$ against the alternative $H_1: \boldsymbol{\theta} \in \Theta_0^c$, where $\Theta_0 \subset \Theta$. We consider the *likelihood ratio test statistic*

$$\lambda(\boldsymbol{X}) = \frac{\sup_{\boldsymbol{\theta} \in \Theta_0} L(\boldsymbol{\theta}; \boldsymbol{X})}{\sup_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}; \boldsymbol{X})}$$

and assume, as before, that $X_1, \ldots, X_n$ are iid and that appropriate *regularity conditions* apply. Under the null hypothesis it can be shown that, as $n \to \infty$, $-2 \ln \lambda(\boldsymbol{X}) \sim \chi^2_\nu$, where the degrees-of-freedom parameter ν of the chi-squared distribution is essentially given by the number of free parameters specified by Θ minus the number of free parameters specified by the null hypothesis $\boldsymbol{\theta} \in \Theta_0$.

For example, suppose that we partition $\boldsymbol{\theta}$ such that $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')$, where $\boldsymbol{\theta}_1$ has dimension q and $\boldsymbol{\theta}_2$ has dimension $p - q$. We wish to test $H_0: \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{1,0}$ against $H_1: \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_{1,0}$. Writing the likelihood as $L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, the likelihood ratio test statistic satisfies

$$-2 \ln \lambda(\boldsymbol{X}) = -2(\ln L(\boldsymbol{\theta}_{1,0}, \hat{\boldsymbol{\theta}}_{2,0}; \boldsymbol{X}) - \ln L(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2; \boldsymbol{X})) \sim \chi^2_q$$

asymptotically, where $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ are the *unconstrained* MLEs of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, and $\hat{\boldsymbol{\theta}}_{2,0}$ is the *constrained* MLE of $\boldsymbol{\theta}_2$ under the null hypothesis. We would reject H_0 if $-2 \ln \lambda(\boldsymbol{X}) > c_{q,1-\alpha}$, where $c_{q,1-\alpha}$ is the $(1-\alpha)$-quantile of the χ^2_q distribution.

An asymptotic $100(1-\alpha)\%$ confidence set for $\boldsymbol{\theta}_1$ consists of the values $\boldsymbol{\theta}_{1,0}$ for which the null hypothesis $H_0: \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{1,0}$ is not rejected: that is,

$$\{\boldsymbol{\theta}_{1,0}: \ \ln L(\boldsymbol{\theta}_{1,0}, \hat{\boldsymbol{\theta}}_{2,0}; \boldsymbol{x}) \geqslant \ln L(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2; \boldsymbol{x}) - 0.5 c_{q,1-\alpha}\}.$$

In particular, if $q = 1$, so that we are interested only in θ_1, we get the confidence interval

$$\{\theta_{1,0}: \ \ln L(\theta_{1,0}, \hat{\boldsymbol{\theta}}_{2,0}; \boldsymbol{x}) \geqslant \ln L(\hat{\theta}_1, \hat{\boldsymbol{\theta}}_2; \boldsymbol{x}) - 0.5 c_{1,1-\alpha}\}. \tag{A.28}$$

Note that such an interval will, in general, be asymmetric about the MLE $\hat{\theta}_1$, in the sense that the distances from the MLE to the upper and lower bounds will be different. This is in contrast to the Wald interval in (A.27), which is rigidly symmetric.

The curve $(\theta_{1,0}, \ln L(\theta_{1,0}, \hat{\boldsymbol{\theta}}_{2,0}; \boldsymbol{x}))$ is sometimes known as the *profile log-likelihood* curve for θ_1 and it attains its maximum at $\hat{\theta}_1$.

A.3.6 Akaike Information Criterion

The likelihood ratio test is applicable to the comparison of nested models, i.e. situations where one model forms a special case of a more general model when certain parameter values are constrained. We often encounter situations where we would like to compare non-nested models with possibly quite different numbers of parameters.

Suppose we have m models $M_1, \ldots, M_m$ and that model j has k_j parameters denoted by $\boldsymbol{\theta}_j = (\theta_{j1}, \ldots, \theta_{jk_j})'$ and a likelihood function $L_j(\boldsymbol{\theta}_j; \boldsymbol{X})$. In Akaike's approach we choose the model minimizing

$$\mathrm{AIC}(M_j) = -2 \ln L_j(\hat{\boldsymbol{\theta}}_j; \boldsymbol{X}) + 2k_j,$$

where $\hat{\boldsymbol{\theta}}_j$ denotes the MLE of $\boldsymbol{\theta}_j$. The AIC number essentially imposes a penalty equal to the number of model parameters k_j on the value of the log-likelihood at the maximum. The model that is favoured is the one for which the penalized log-likelihood $\ln L_j(\hat{\boldsymbol{\theta}}_j; \boldsymbol{X}) - k_j$ is largest. There are alternatives to the AIC, such as the Bayesian information criterion (BIC) of Schwarz, which impose different penalties for the number of parameters. See Burnham and Anderson (2002) for more about model comparison using these criteria.

References

Aalen, O., Ø. Borgan and H. Gjessing. 2010. *Survival and Event History Analysis: A Process Point of View*. Springer.

Aas, K. and G. Puccetti. 2014. Bounds on total economic capital: the DNB case study. *Extremes* 17(4):693–715.

Aas, K., C. Czado, A. Frigessi and H. Bakken. 2009. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* 44(2):182–198.

Abdous, B. and B. Remillard. 1995. Relating quantiles and expectiles under weighted-symmetry. *Annals of the Institute of Statistical Mathematics* 47(2):371–384.

Abdous, B., K. Ghoudi and A. Khoudraji. 1999. Non-parametric estimation of the limit dependence function of multivariate extremes. *Extremes* 2(3):245–268.

Abraham, B. and J. Ledolter. 1983. *Statistical Methods for Forecasting*. Wiley.

Abramowitz, M. and I. A. Stegun (eds). 1965. *Handbook of Mathematical Functions*. New York: Dover.

Acerbi, C. 2002. Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking and Finance* 26(7):1505–1518.

Acerbi, C. and B. Szekely. 2014. Backtesting expected shortfall. *Risk Magazine*, December.

Acerbi, C. and D. Tasche. 2002. On the coherence of expected shortfall. *Journal of Banking and Finance* 26:1487–1503.

Acharya, B. V., V. V. Cooley, M. Richardson and I. Walter. 2009. Manufacturing tail risk: a perspective on the financial crisis of 2007–9. *Foundations and Trends in Finance* 4(4):247–325.

Adam, A., M. Houkari and J.-P. Laurent. 2008. Spectral risk measures and portfolio selection. *Journal of Banking and Finance* 32(9):1870–1882.

Admati, A. and M. Hellwig. 2013. *The Bankers' New Clothes: What's Wrong with Banking and What to Do about It*. Princeton University Press.

Admati, A., P. DeMarzo, M. Hellwig and P. Pfleiderer. 2013. Fallacies, irrelevant facts, and myths in the discussion of capital regulation: why bank equity is not socially expensive. Discussion Paper 2013/23, Max Planck Institute for Research on Collective Goods.

Albanese, C. 1997. Credit exposure, diversification risk and coherent VaR. Preprint, Department of Mathematics, University of Toronto.

Albert, A. 1962. Estimating the infinitesimal generator of a continuous time, finite state Markov process. *Annals of Mathematical Statistics* 33(2):727–753.

Alessandri, P. and M. Drehmann. 2010. An economic capital model integrating credit and interest rate risk in the banking book. *Journal of Banking and Finance* 34:730–742.

Alexander, C. 2001. *Market Models: A Guide to Financial Data Analysis*. Wiley.

——. 2009. *Market Risk Analysis*, vols 1–4. Wiley.

Alsina, C., M. J. Frank and B. Schweizer. 2006. *Associative Functions: Triangular Norms and Copulas*. World Scientific.

Altman, E. L. 1993. *Corporate Financial Distress and Bankruptcy*. Wiley.

Ames, M., T. Schuermann and H. S. Scott. 2014. Bank capital for operational risk: a tale of fragility and instability. Preprint, Oliver Wyman and Wharton Financial Institutions Center.

Amini, H., R. Cont and A. Minca. 2012. Stress testing the resilience to contagion of financial networks. *International Journal of Theoretical and Applied Finance* 15:1–20.

Andersen, L. and J. Sidenius. 2004. Extensions to the Gaussian copula: random recovery and random factor loadings. *Journal of Credit Risk* 1(1):29–70.

Ané, T. and H. Geman. 2000. Order flow, transaction clock, and normality of asset returns. *Journal of Finance* 55:2259–2284.

Arnsdorf, M. and I. Halperin. 2009. BSLP: Markovian bivariate spread-loss model for portfolio credit derivatives. *Journal of Computational Finance* 12:77–107.

Artzner, P. and F. Delbaen. 1995. Default risk insurance and incomplete markets. *Mathematical Finance* 5:187–195.

Artzner, P., F. Delbaen, J. M. Eber and D. Heath. 1997. Thinking coherently. *Risk* 10(11):68–71.

——. 1999. Coherent measures of risk. *Mathematical Finance* 9:203–228.

Artzner, P., F. Delbaen, J. M. Eber, D. Heath and H. Ku. 2007. Coherent multiperiod risk adjusted values and Bellman's principle. *Annals of Operations Research* 152:5–22.

Asmussen, S. and H. Albrecher. 2010. *Ruin Probabilities*, 2nd edn. World Scientific.

Asmussen, S. and P. Glynn. 2007. *Stochastic Simulation. Algorthms and Analysis*. Springer.

Asmussen, S., K. Binswanger and B. Højgaard. 2000. Rare events simulation for heavy-tailed distributions. *Bernoulli* 6(2):303–322.

Atkinson, A. C. 1982. The simulation of generalized inverse Gaussian and hyperbolic random variables. *SIAM Journal on Scientific Computing* 3(4):502–515.

Attanasio, O. P. 1991. Risk, time varying second moments and market efficiency. *Review of Economic Studies* 58:479–494.

Aue, F. and M. Kalkbrener. 2006. LDA at work: Deutsche Bank's approach to quantifying operational risk. *Journal of Operational Risk* 1(4):49–93.

Avellaneda, M. 1998. Minimum entropy calibration of asset-pricing models. *International Journal of Theoretical and Applied Finance* 1:447–479.

Aven, T. 1985. A theorem for determining the compensator of a counting process. *Scandinavian Journal of Statistics* 12(1):69–72.

Bailey, A. L. 1945. A generalized theory of credibility. *Proceedings of the Casualty Actuarial Society* 32:13–20.

Bain, A. and D. Crisan. 2009. *Fundamentals of Stochastic Filtering*. Springer.

Balkema, A. A. and L. de Haan. 1974. Residual life time at great age. *Annals of Probability* 2:792–804.

Balkema, A. A. and S. I. Resnick. 1977. Max-infinite divisibility. *Journal of Applied Probability* 14:309–319.

Balkema, G. and P. Embrechts. 2007. *High Risk Scenarios and Extremes: A Geometric Approach*. Zurich Lectures in Advanced Mathematics. Zurich: European Mathematical Society Publishing House.

Bangia, A., F. X. Diebold, A. Kronimus, C. Schlagen and T. Schuermann. 2002. Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking and Finance* 26:445–474.

Baringhaus, L. 1991. Testing for spherical symmetry of a multivariate distribution. *Annals of Statistics* 19(2):899–917.

Barndorff-Nielsen, O. E. 1978. Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of Statistics* 5:151–157.

——. 1997. Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics* 24:1–13.

Barndorff-Nielsen, O. E. and P. Blæsild. 1981. Hyperbolic distributions and ramifications: contributions to theory and application. In *Statistical Distributions in Scientific Work* (ed. C. Taillie, G. Patil and B. Baldessari), vol. 4, pp. 19–44. Dordrecht: Reidel.

Barone-Adesi, G., F. Bourgoin and K. Giannopoulos. 1998. Don't look back. *Risk* 11(8):100–103.

Barrieu, P. and L. Albertini (eds). 2009. *The Handbook of Insurance-Linked Securities*. Wiley.

Basak, S. and A. Shapiro. 2001. Value-at-risk based risk management: optimal policies and asset prices. *Review of Financial Studies* 14(2):371–405.

Basel Committee on Banking Supervision. 2003. The 2002 loss data collection exercise for operational risk: summary of the data collected. Risk Management Group, Bank for International Settlements. (Available from www.bis.org/bcbs/qis/ldce2002.pdf.)

——. 2004. International convergence of capital measurement and capital standards. A revised framework. Bank for International Settlements. (Available from www.bis.org/publ/bcbs107.htm.)

——. 2006. Basel II: international convergence of capital measurement and capital standards. A revised framework—comprehensive version. Bank for International Settlements. (Available from www.bis.org/publ/bcbs128.htm.)

——. 2011. Basel III: a global framework for more resilient banks and banking systems—revised version. Bank for International Settlements. (Available from www.bis.org/publ/bcbs189.htm.)

——. 2012. Fundamental review of the trading book. Bank for International Settlements. (Available from www.bis.org/publ/bcbs219.htm.)

——. 2013a. Fundamental review of the trading book: a revised market risk framework. Bank for International Settlements. (Available from www.bis.org/publ/bcbs265.htm.)

——. 2013b. The regulatory framework: balancing risk sensitivity, simplicity and comparability. Bank for International Settlements. (Available from www.bis.org/publ/bcbs258.htm.)

——. 2014. Operational risk: revisions to the simpler approaches. Consultative Document, Bank of International Settlements. (Available from www.bis.org/publ/bcbs291.htm.)

Bauwens, L., S. Laurent and J. V. K. Rombouts. 2006. Multivariate GARCH models: a survey. *Journal of Applied Econometrics* 21(1):79–109.

Baxter, M. and A. Rennie. 1996. *Financial Calculus*. Cambridge University Press.

Becherer, D. 2004. Utility indifference hedging and valuation via reaction–diffusion systems. *Proceedings of the Royal Society of London* A 460:27–51.

Becherer, D. and M. Schweizer. 2005. Classical solutions to reaction diffusion systems for hedging problems with interacting Itô and point processes. *Annals of Applied Probability* 15:1111–1144.

Bedford, T. and R. Cooke. 2001. *Probabilistic Risk Analysis: Foundations and Methods*. Cambridge University Press.

Beirlant, J., Y. Goegebeur, T. Segers and J. L. Teugels. 2004. *Statistics of Extremes: Theory and Applications*. Wiley.

Bélanger, A., S. E. Shreve and D. Wong. 2004. A general framework for pricing credit risk. *Mathematical Finance* 14:317–350.

Bellini, F. and V. Bignozzi. 2013. Elicitable risk measures. *Quantitative Finance*, forthcoming.

Bénéplanc, G. and J.-C. Rochet. 2011. *Risk Management in Turbulent Times*. Oxford University Press.

Bening, V. E. and V. Y. Korolev. 2002. *Generalized Poisson Models and Their Applications in Insurance and Finance*. Utrecht: VSP.

Benzschawel, T., J. Dagraca and H. Fok. 2010. Loan valuation. In *Encyclopedia of Quantitative Finance* (ed. R. Cont), vol. III, pp. 1071–1075. Wiley.

Beran, J. 1979. Testing for ellipsoidal symmetry of a multivariate density. *Annals of Statistics* 7(1):150–162.

Beran, J., Y. Feng, S. Ghosh and R. Kulik. 2013. *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Springer.

Berk, J. B. 1997. Necessary conditions for the CAPM. *Journal of Economic Theory* 73:245–257.

Berkes, I., L. Horváth and P. Kokoszka. 2003. GARCH processes: structure and estimation. *Bernoulli* 9:201–228.

Berkowitz, J. 2000. A coherent framework for stress testing. *Journal of Risk* 2(2):1–11.

——. 2001. Testing the accuracy of density forecasts, applications to risk management. *Journal of Business and Economic Statistics* 19(4):465–474.

——. 2002. Testing distributions. *Risk* 15(6):77–80.

Berkowitz, J. and J. O'Brien. 2002. How accurate are value-at-risk models at commercial banks? *Journal of Finance* 57:1093–1112.

Bernard, C., X. Jiang and R. Wang. 2014. Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics* 54:93–108.

Bernard, C., L. Rüschendorf and S. Vanduffel. 2013. Value-at-risk bounds with variance constraints. Preprint, University of Freiburg.

Bernard, C., Y. Liu, N. MacGillivray and J. Zhang. 2013. Bounds on capital requirements for bivariate risk with given marginals and partial information on the dependence. *Dependence Modeling* 1:37–53.

Berndt, A., R. Douglas, D. Duffie, F. Ferguson and D. Schranz. 2008. Measuring default risk premia from default swap rates and EDFs. Preprint, Tepper School of Business, Carnegie Mellon University.

Berndt, E. K., B. H. Hall, R. E. Hall and J. A. Hausman. 1974. Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement* 3:653–665.

Bernstein, P. L. 1998. *Against the Gods: The Remarkable Story of Risk*. Wiley.

Bertsekas, D. P. 1999. *Nonlinear Programming*, 2nd edn. Belmont, MA: Athena Scientific.

Bertsimas, D. and J. N. Tsitsiklis. 1997. *Introduction to Linear Optimization*. Belmont, MA: Athena Scientific.

Bibby, B. M. and M. Sørensen. 2003. Hyperbolic processes in finance. In *Handbook of Heavy Tailed Distributions in Finance* (ed. S. T. Rachev), pp. 211–248. Amsterdam: Elsevier.

Bickel, P. J. and K. A. Doksum. 2001. *Mathematical Statistics: Basic Ideas and Selected Topics*, 2nd edn, vol. 1. Upper Saddle River, NJ: Prentice Hall.

Bielecki, T. R. and M. Rutkowski. 2002. *Credit Risk: Modeling, Valuation, and Hedging*. Springer.

Bielecki, T. R., S. Crépey and M. Jeanblanc. 2010. Up and down credit risk. *Quantitative Finance* 10(10):1137–1151.

Bielecki, T. R., M. Jeanblanc and M. Rutkowski. 2004. Hedging of defaultable claims. In *Paris–Princeton Lectures on Mathematical Finance 2003* (ed. R. A. Carmona et al.), pp. 1–132. Springer.

——. 2007. Hedging of basket credit derivatives in the credit default swap market. *Journal of Credit Risk* 3(1):91–132.

Billingsley, P. 1995. *Probability and Measure*, 3rd edn. Wiley.

Bingham, N. H. and R. Kiesel. 2002. Semi-parametric modelling in finance: theoretical foundations. *Quantitative Finance* 2:241–250.

——. 2004. *Risk-Neutral Valuation*, 2nd edn. Springer.

Bingham, N. H., C. M. Goldie and J. L. Teugels. 1987. *Regular Variation*. Cambridge University Press.

Björk, T. 2004. *Arbitrage Theory in Continuous Time*, 2nd edn. Oxford University Press.

Black, F. and J. C. Cox. 1976. Valuing corporate securities: some effects of bond indenture provisions. *Journal of Finance* 31(2):351–367.

Black, F. and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81(3):637–654.

Blæsild, P. 1981. The two-dimensional hyperbolic distribution and related distributions, with an application to Johannsen's bean data. *Biometrika* 68(1):251–263.

Blæsild, P. and J. L. Jensen. 1981. Multivariate distributions of hyperbolic type. In *Statistical Distributions in Scientific Work* (ed. C. Taillie, G. Patil and B. Baldessari), vol. 4, pp. 45–66. Dordrecht: Reidel.

Blanchet-Scalliet, C. and M. Jeanblanc. 2004. Hazard rate for credit risk and hedging defaultable contingent claims. *Finance and Stochastics* 8:145–159.

Bluhm, C. and L. Overbeck. 2007. *Structured Credit Portfolio Analysis, Baskets and CDOs*. Financial Mathematics Series. Boca Raton, FL: Chapman & Hall.

Blum, P. 2005. On Some Mathematical Aspects of Dynamic Financial Analysis. PhD thesis, ETH Zurich.

Blum, P. and M. Dacorogna. 2004. DFA—dynamic financial analysis. In *Encyclopedia of Actuarial Science* (ed. J. L. Teugels and B. Sundt), pp. 505–519. Wiley.

Bodnar, G. M., G. S. Hayt and R. C. Marston. 1998. 1998 Wharton survey of risk management by US non-financial firms. *Financial Management* 27(4):70–91.

Bohn, J. R. 2000. An empirical assessment of a simple contingent-claims model for the valuation of risky debt. *Journal of Risk Finance* 1(4):55–77.

Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31:307–327.

——. 1990. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach. *Review of Economics and Statistics* 72:498–505.

Bollerslev, T., T. Chou and K. Kroner. 1992. ARCH modelling in finance: a selective review of theory and empirical evidence. *Journal of Econometrics* 52:5–59.

Bollerslev, T., R. F. Engle and D. B. Nelson. 1994. ARCH models. In *Handbook of Econometrics* (ed. R. F. Engle and D. L. McFadden), vol. 4, pp. 2959–3038. Amsterdam: North-Holland.

Bollerslev, T., R. F. Engle and J. M. Wooldridge. 1988. A capial-asset pricing model with time-varying covariances. *Journal of Political Economy* 96:116–131.

Bougerol, P. and N. Picard. 1992. Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52:115–127.

Bowers, N. L., H. U. Gerber, J. C. Hickman, D. A. Jones and C. J. Nesbitt. 1986. *Actuarial Mathematics*. Itasca, IL: Society of Actuaries.

Bowsher, C. G. 2002. Modelling security market events in continuous time: intensity based, multivariate point process models. Economics Discussion Paper 2002-W22, Nuffield College, Oxford.

——. 2007. Modelling security market events in continuous time: intensity based, multivariate point process models. *Journal of Econometrics* 141(2):876–912.

Box, G. E. P. and G. M. Jenkins. 1970. *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day.

Box, G. E. P. and D. A. Pierce. 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* 65:1509–1526.

Boyd, S. and L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press.

Boyer, B. H., M. S. Gibson and M. Loretan. 1999. Pitfalls in tests for changes in correlations. Technical Report 597R, Federal Reserve Board.

Boyle, P. P. and F. Boyle. 2001. *Derivatives: The Tools that Changed Finance*. London: Risk Books.

Brand, L. and R. Bahr. 2001. Ratings performance 2000: default, transition, recovery, and spreads. New York: Standard & Poor's.

Brandt, A. 1986. The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability* 18:211–220.

Brealey, R. and S. Myers. 2000. *Principles of Corporate Finance*, 6th edn. New York: McGraw-Hill.

Breiman, L. 1965. On some limit theorems similar to the arc-sin law. *Theory of Probability and Its Applications* 10:323–331.

Brémaud, P. 1981. *Point Processes and Queues: Martingale Dynamics*. Springer.

Brennan, M. J. and L. Trigeorgis (eds). 1999. *Project Flexibility, Agency, and Competition*. Oxford University Press.

Breslow, N. and D. G. Clayton. 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88:9–25.

Breuer, T., M. Jandačka, K. Rheinberger and M. Summer. 2009. How to find plausible, severe, and useful stress scenarios. *International Journal of Central Banking* 5:205–224.

——. 2010. Does adding up of capital for market and credit risk amount to conservative risk assessment? *Journal of Banking and Finance* 34(4):703–712.

Breymann, W., A. Dias and P. Embrechts. 2003. Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance* 3:1–14.

Brigo, D. and K. Chourdakis. 2009. Counterparty risk for credit default swaps: impact of spread volatility and default correlation. *International Journal of Theoretical and Applied Finance* 12:1007–1026.

Brigo, D. and F. Mercurio. 2006. *Interest Rate Models: Theory and Practice—With Smile, Inflation and Credit*, 2nd edn. Springer.

Brigo, D., A. Capponi and A. Pallavicini. 2014. Arbitrage-free bilateral counterparty risk valuation under collaterization and application to credit default swaps. *Mathematical Finance* 24:125–146.

Brigo, D., M. Morini and A. Pallavicini. 2013. *Counterparty Credit Risk, Collateral and Funding*. Wiley.

Brigo, D., A. Pallavicini and R. Torresetti. 2010. *Credit Models and the Crisis: A Journey into CDOs, Copulas, Correlations and Dynamic Models*. Wiley.

Briys, E. and F. de Varenne. 2001. *Insurance—From Underwriting to Derivatives: Asset Liabilitiy Management in Insurance Companies*. Wiley.

Brockwell, P. J. and R. A. Davis. 1991. *Time Series: Theory and Methods*, 2nd edn. Springer.

——. 2002. *Introduction to Time Series and Forecasting*, 2nd edn. Springer.

Brooks, C., S. P. Burke and G. Persand. 2001. Benchmarks and the accuracy of GARCH model estimation. *International Journal of Forecasting* 17:45–56.

Brown, A. 2012. *Red-Blooded Risk: The Secret History of Wall Street*. Wiley.

Brummelhuis, R. G. M. and R. Kaufmann. 2007. Time scaling for value at risk in GARCH(1,1) and AR(1)–GARCH(1,1) processes. *Journal of Risk* 9(4):39–94.

Brunnermeier, M. 2009. Deciphering the liquidity and credit crunch 2007–2008. *Journal of Economic Perspectives* 23:77–100.

Brunnermeier, M. and L. H. Pedersen. 2009. Market liquidity and funding liquidity. *Review of Financial Studies* 22:2201–2238.

Brunnermeier, M., L. Clerc and M. Scheicher. 2013. Assessing contagion risks in the CDS markets. Technical Report 17, Banque de France, Financial Stability Review.

Bühlmann, H. 1967. Experience rating and credibility. *ASTIN Bulletin* 4:199–207.

——. 1969. Experience rating and credibility. *ASTIN Bulletin* 5:157–165.

——. 1970. *Mathematical Methods in Risk Theory*. Springer.

——. 1971. Credibility procedures. In *Proceedings of the 6th Berkeley Symposium on Probability and Statistics*, vol. 1, pp. 515–525. Berkeley, CA: University of California Press.

——. 1984. Numerical evaluation of the compound Poisson distribution: recursion or the fast Fourier transform? *Scandinavian Actuarial Journal* 1984(2):116–126.

Bühlmann, H. and A. Gisler. 2005. *A Course in Credibility Theory and Its Applications*. Springer.

Bühlmann, P. and A. J. McNeil. 2002. An algorithm for nonparametric GARCH modelling. *Computational Statistics and Data Analysis* 40:665–683.

Burnham, K. P. and D. R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer.

Cambanis, S., S. Huang and G. Simons. 1981. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis* 11:368–385.

Campbell, J. Y., A. W. Lo and A. C. MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton University Press.

Carmona, R. and M. Tehranchi. 2006. *Interest Rate Models: An Infinite Dimensional Stochastic Analysis Perspective*. Springer.

Casella, G. and R. L. Berger. 2002. *Statistical Inference*. Pacific Grove, CA: Duxbury.

CEIOPS. 2006. Draft advice to the European Commission in the Framework of the Solvency II project on Pillar I issues: further advice—Committee of European Insurance and Occupational Pensions Supervisors. Technical Report, Committee of European Insurance and Occupational Pensions Supervisors.

——. 2009. Draft CEIOPS advice for Level 2 implementing measures on Solvency II: SCR Standard Formula, Article 109(1c), Correlations. Consultation Paper 74 (Technical Report), Committee of European Insurance and Occupational Pensions Supervisors.

Cesari, G., J. Aquilina, N. Charpillon, Z. Filipović, G. Lee and I. Manda. 2009. *Modelling, Pricing and Hedging Counterparty Credit Exposure: A Technical Guide*. Springer.

Cetin, U. 2012. On absolutely continuous compensators and nonlinear filtering equations in default risk models. *Stochastic Processes and Their Applications* 122:3619–3647.

Chamberlain, G. 1983. A characterization of the distributions that imply mean–variance utility functions. *Journal of Economic Theory* 29:185–201.

Chatfield, C. 2003. *The Analysis of Time Series: An Introduction*, 6th edn. London: Chapman & Hall/CRC.

Chavez-Demoulin, V. and P. Embrechts. 2004. Smooth extremal models in insurance and finance. *Journal of Risk and Insurance* 71(2):183–199.

Chavez-Demoulin, V. and J. McGill. 2012. High-frequency financial data modeling using Hawkes processes. *Journal of Banking and Finance* 36(12):3415–3426.

Chavez-Demoulin, V., A. C. Davison and A. J. McNeil. 2005. Estimating value-at-risk: a point process approach. *Quantitative Finance* 5(2):227–234.

Chavez-Demoulin, V., P. Embrechts and M. Hofert. 2014. An extreme value approach for modeling operational risk losses depending on covariates. *Journal of Risk and Insurance*, forthcoming.

Chavez-Demoulin, V., P. Embrechts and S. Sardy. 2014. Extreme-quantile tracking for financial time series. *Journal of Econometrics* 181(1):44–52.

Chekhlov, H., S. Uryasev and M. Zabarankin. 2005. Drawdown measure in portfolio optimization. *International Journal of Theoretical and Applied Finance* 8(1):13–58.

Chen, X. and Y. Fan. 2005. Pseudo-likelihood ratio tests for model selection in semiparametric multivariate copula models. *Canadian Journal of Statistics* 33:389–414.

——. 2006. Estimation of copula-based semiparametric time series models. *Journal of Econometrics* 130:307–335.

Cheridito, P., F. Delbaen and M. Kupper. 2005. Coherent and convex monetary risk measures for bounded càdlàg processes. *Finance and Stochastics* 9:1713–1732.

Cherubini, U., E. Luciano and W. Vecchiato. 2004. *Copula Methods in Finance*. Wiley.

Chicago Mercantile Exchange. 2010. CME SPAN: standard portfolio analysis of risk. (Available from www.cmegroup.com/clearing/files/span-methodology.pdf.)

Choudhry, M. 2012. *The Principles of Banking*. Wiley.

Christensen, J. H. E., F. X. Diebold and G. D. Rudebusch. 2011. The affine arbitrage-free class of Nelson–Siegel term structure models. *Journal of Econometrics* 164:4–20.

Christoffersen, P. F. and D. Pelletier. 2004. Backtesting value-at-risk: a duration-based approach. *Journal of Financial Econometrics* 2(1):84–108.

Christoffersen, P. F., J. Hahn and A. Inoue 2001. Testing and comparing value-at-risk measures. *Journal of Empirical Finance* 8(3):325–342.

Clayton, D. G. 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 65:141–151.

——. 1996. Generalized linear mixed models. In *Markov Chain Monte Carlo in Practice* (ed. W. R. Gilks, S. Richardson and D. J. Spiegelhalter), pp. 275–301. London: Chapman & Hall.

Coates, J. 2012. *The Hour between Dog and Wolf: Risk Taking, Gut Feelings and the Biology of Boom and Bust*. New York: The Penguin Press.

Coculescu, D., H. Geman and M. Jeanblanc. 2008. Valuation of default sensitive claims under imperfect information. *Finance and Stochastics* 12:195–218.

Coleman, R. 2002. Op risk modelling for extremes. Part 1. Small sample modelling. *Operational Risk* 3(12):8–11.

——. 2003. Op risk modelling for extremes. Part 2. Statistical methods. *Operational Risk* 4(1):6–9.

Coles, S. G. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer.

Coles, S. G. and J. A. Tawn. 1991. Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 53(2):377–392.

Coles, S. G., J. Heffernan and J. A. Tawn. 1999. Dependence measures for extreme value analyses. *Extremes* 2(4):339–365.

Collin-Dufresne, P., R. S. Goldstein and J. Helwege. 2010. Is credit event risk priced? Modeling contagion via the updating of beliefs. Technical Report, National Bureau of Economic Research.

Collin-Dufresne, P., R. S. Goldstein and J. Hugonnier. 2004. A general formula for valuing defaultable securities. *Econometrica* 72(5):1377–1407.

Comte, F. and O. Lieberman. 2003. Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* 84:61–84.

Connor, G. 1995. The three types of factor models: a comparison of their explanatory power. *Financial Analysts Journal* 51(3):42–46.

Conover, W. J. 1999. *Practical Nonparametric Statistics*. Wiley.

Cont, C. 2010. Credit default swaps. In *Encyclopedia of Quantitative Finance* (ed. R. Cont), pp. 422–424. Wiley.

Cont, R. 2006. Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance* 16(3):519–542.

Cont, R. and Y. H. Kan. 2011. Dynamic hedging of portfolio credit derivatives. *SIAM Journal for Financial Mathematics* 2:112–140.

Cont, R. and A. Minca. 2013. Recovering portfolio default intensities implied by CDO quotes. *Mathematical Finance* 23:94–121.

Cope, E. and G. Antonini. 2008. Observed correlations and dependencies among operational losses in the ORX Consortium databse. *Journal of Operational Risk* 3(4):47–76.

Cope, E. and A. Labbi. 2008. Operational loss scaling by exposure indicators: evidence from the ORX database. *Journal of Operational Risk* 3(4):25–46.

Copeland, T. and V. Antikarov. 2001. *Real Options*. New York: Texere.

Costanzino, N. and M. Curran. 2014. Backtesting general spectral risk measures with application to expected shortfall. Preprint, RiskLab, University of Toronto.

Couderc, F. and C. Finger. 2010. CDS indices. In *Encyclopedia of Quantitative Finance* (ed. R. Cont), pp. 422–424. Wiley.

Cox, D. R. and V. Isham. 1980. *Point Processes*. London: Chapman & Hall.

Cox, D. R. and D. Oakes. 1984. *Analysis of Survival Data*. London: Chapman & Hall.

Cox, J. C. and M. Rubinstein. 1985. *Options Markets*. Englewood Cliffs, NJ: Prentice Hall.

Cox, J. C., J. E. Ingersoll and S. A. Ross. 1985. A theory of the term structure of interest rates. *Econometrica* 53(2):385–407.

Cramér, H. 1994a. *Collected Works* (ed. A. Martin-Löf), vol. I. Springer.

——. 1994b. *Collected Works* (ed. A. Martin-Löf), vol. II. Springer.

Credit Suisse Financial Products. 1997. CreditRisk$^+$: a credit risk management framework. Technical Document. (Available from www.csfb.com/creditrisk.)

Crépey, S. 2012a. Bilateral counterparty risk under funding constraints. I: Pricing. *Mathematical Finance*, doi: 10.1111/mafi.12004.

Crépey, S. 2012b. Bilateral counterparty risk under funding constraints. II: CVA. *Mathematical Finance*, doi: 10.1111/mafi.12005.

Crosbie, P. J. and J. R. Bohn. 2002. Modeling default risk. Technical Document, Moody's/KMV, New York.

Crouhy, M., D. Galai and R. Mark. 2000. A comparative analysis of current credit risk models. *Journal of Banking and Finance* 24:59–117.

——. 2001. *Risk Management*. New York: McGraw-Hill.

Crouhy, M., R. A. Jarrow and S. Turnbull. 2008. The subprime crisis of 2007. *Journal of Derivatives* 16:81–110.

Crowder, M. J. 1976. Maximum likelihood estimation for dependent observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 38:45–53.

Crowder, M. J., M. H. A. Davis and G. Giampieri. 2005. Analysis of default data unsing hidden Markov models. *Quantitative Finance* 5:27–34.

Cruz, M. G. 2002. *Modelling, Measuring and Hedging Operational Risk*. Wiley.

——. 2004. *Operational Risk Modelling and Analysis: Theory and Practice*. London: Risk Books.

Cruz, M. G., G. W. Peters and P. V. Shevchenko. 2015. *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk*. Wiley.

Czado, C. 2010. Pair copula constructions of multivariate copulas. In *Copula Theory and Its Applications* (ed. P. Jaworski, F. Durante, W. Härdle and T. Rychlik), pp. 93–109. Lecture Notes in Statistics. Springer.

Dacorogna, M. M., R. Gençay, U. Müller, R. B. Olsen and O. V. Pictet. 2001. *An Introduction to High-Frequency Finance*. San Diego, CA: Academic Press.

Daley, D. J. and D. Vere-Jones. 2003. *An Introduction to the Theory of Point Processes*, vol. I: *Elementary Theory and Methods*, 2nd edn. Springer.

Daníelsson, J. 2011. *Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in R and* MATLAB. Wiley.

Daníelsson, J. and C. G. de Vries. 1997a. Beyond the sample: extreme quantile and probability estimation. Discussion Paper, Tinbergen Institute, Rotterdam.

——. 1997b. Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance* 4:241–257.

——. 1997c. Value-at-risk and extreme returns. In *Extremes and Integrated Risk Management* (ed. P. Embrechts), pp. 85–106. London: Risk Books.

Daníelsson, J., L. de Haan, L. Peng and C. G. de Vries. 2001a. Using a bootstrap method to choose the sample fraction in the tail index estimation. *Journal of Multivariate Analysis* 76:226–248.

Daníelsson, J., P. Embrechts, C. Goodhart, C. Keating, F. Muennich and O. Renault. 2001b. An academic response to Basel II. Special Paper 130, Financial Markets Group, London School of Economics.

Das, B., P. Embrechts and V. Fasen. 2013. Four theorems and a financial crisis. *International Journal of Approximate Reasoning* 54:701–716.

Daul, S., E. De Giorgi, F. Lindskog and A. J. McNeil. 2003. The grouped t-copula with an application to credit risk. *Risk* 16(11):73–76.

Davis, M. H. A. 2014. Consistency of risk measure estimates. Preprint (arXiv:1410.4382v1).

Davis, M. H. A. and V. Lo. 2001. Infectious defaults. *Quantitative Finance* 1(4):382–387.

Davis, M. H. A. and S. I. Marcus. 1981. An introduction to nonlinear filtering. In *Stochastic Systems: The Mathematics of Filtering and Identification* (ed. M. Hazewinkel and J. C. Willems), pp. 53–75. Dordrecht: Reidel.

Davison, A. C. 1984. Modelling excesses over high thresholds, with an application. In *Statistical Extremes and Applications* (ed. J. T. de Oliveira), pp. 461–482. Dordrecht: Reidel.

——. 2003. *Statistical Models*. Cambridge University Press.

Davison, A. C. and R. L. Smith. 1990. Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 52:393–442.

de Fontnouvelle, P., V. DeJesus-Rueff, J. Jordan and E. Rosengren. 2003. Using loss data to quantify operational risk. Working Paper, Federal Reserve Bank of Boston.

de Jongh, R., T. de Wet, H. Raubenheimer and J. Venter. 2014. Combining scenario and historical data in the loss distribution approach: a new procedure that incorporates measures of agreement between scenarios and historical data. Preprint, North-West University, South Africa.

de la Vega, J. 1996. Confusión de confusiones. In *Extraordinary Popular Delusions and the Madness of Crowds & Confusión de Confusiones* (ed. M. S. Fridson), pp. 125–211. Wiley.

de Matos, J. A. 2001. *Theoretical Foundations of Corporate Finance*. Princeton University Press.

de Servigny, A. and O. Renault. 2002. Default correlation: empirical evidence. Standard & Poor's, Risk Solutions.

De Vylder, F. 1996. *Advanced Risk Theory. A Self-Contained Introduction*. Brussels: Editions de l'Université de Bruxelles.

Deheuvels, P. 1980. The decomposition of infinite order and extreme multivariate distributions. In *Asymptotic Theory of Statistical Tests and Estimation* (ed. I. M. Chakravarti), pp. 259–286. New York: Academic Press.

Dekkers, A. L. M. and L. de Haan. 1989. On the estimation of the extreme-value index and large quantile estimation. *Annals of Statistics* 17(4):1795–1832.

——. 1993. Optimal choice of sample fraction in extreme-value estimation. *Journal of Multivariate Analysis* 47:173–195.

Dekkers, A. L. M., J. H. J. Einmahl and L. de Haan. 1989. A moment estimator for the index of an extreme-value distribution. *Annals of Statistics* 17:1833–1855.

Delbaen, F. 2000. *Coherent Risk Measures*. Pisa: Cattedra Galiliana, Scuola Normale Superiore.

——. 2002. Coherent risk measures on general probability spaces. In *Advances in Finance and Stochastics* (ed. K. Sandmann and P. Schönbucher), pp. 1–37. Springer.

——. 2012. *Monetary Utility Functions*. Osaka University Press.

Delbaen, F. and W. Schachermayer. 2006. *The Mathematics of Arbitrage*. Springer.

Delbaen, F., P. Grandits, T. Rheinländer, D. Samperi, M. Schweizer and C. Stricker. 2002. Exponential hedging and entropic penalties. *Mathematical Finance* 12:99–123.

Demarta, S. and A. J. McNeil. 2005. The t copula and related copulas. *International Statistical Review* 73(1):111–129.

Dembo, A., J.-D. Deuschel and D. Duffie. 2004. Large portfolio losses. *Finance and Stochastics* 8:3–16.

Denault, M. 2001. Coherent allocation of risk capital. *Journal of Risk* 4(1):1–34.

Denneberg, D. 1990. Distorted probabilities and insurance premiums. *Methods of Operations Research* 63:3–5.

Denuit, M. and A. Charpentier. 2004. *Mathématiques de l'Assurance Non-Vie*, vol. 1: *Principes Fondamentaux de Théorie de Risque*. Paris: Economica.

Denuit, M., J. Dhaene, M. Goovaerts and R. Kaas. 2005. *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. Wiley.

Devlin, S. J., R. Gnanadesikan and J. R. Kettenring. 1975. Robust estimation and outlier detection with correlation coefficients. *Biometrika* 62(3):531–545.

——. 1981. Robust estimation of dispersion matrices and principal components. *Journal of the American Statistical Association* 76:354–362.

Dewatripont, M., J.-C. Rochet and J. Tirole. 2010. *Balancing the Banks: Global Lessons from the Financial Crisis*. Princeton University Press.

Di Graziano, G. and L. C. G. Rogers. 2009. A dynamic approach to the modelling of correlation credit derivatives using Markov chains. *International Journal of Theoretical and Applied Finance* 12:45–62.

Dias, A. and P. Embrechts. 2009. Testing for structural changes in exchange rates: dependence beyond linear correlation. *European Journal of Finance* 15:619–637.

Dickey, D. and W. Fuller. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74:427–431.

Dickson, D. C. M., M. R. Hardy and H. R. Waters. 2013. *Actuarial Mathematics for Life Contingent Risks*, 2nd edn. Cambridge University Press.

Diebold, F. X. and C. Li. 2006. Forecasting the structure of government bond yields. *Journal of Econometrics* 130(2):337–364.

Diebold, F. X., T. Schuermann, A. Hickmann and A. Inoue. 1998. Scale models. *Risk* 11:104–107.

Dillmann, T. 2002. *Modelle zur Bewertung von Optionen in Lebensversicherungsverträgen*. Ulm: Ifa-Schriftenreihe.

Dimakos, X. K. and K. Aas. 2004. Integrated risk modelling. *Statistical Modelling* 4(4):265–277.

Ding, Z. 1994. Time Series Analysis of Speculative Returns. PhD thesis, University of California, San Diego.

Ding, Z., C. W. Granger and R. F. Engle. 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1:83–106.

Doebeli, B., M. Leippold and P. Vanini. 2003. From operational risk to operational excellence. In *Advances in Operational Risk: Firm-wide Issues for Financial Institutions*, 2nd edn, pp. 239–252. London: Risk Books.

Does, R. J. M. M., K. C. B. Roes and A. Trip. 1999. *Statistical Process Control in Industry: Implementation and Assurance of SPC*. Dordrecht: Kluwer Academic.

Doherty, J. N. A. 2000. *Integrated Risk Management. Techniques and Strategies for Reducing Risk*. New York: McGraw-Hill.

Donnelly, C. and P. Embrechts. 2010. The devil is in the tails: actuarial mathematics and the subprime mortgage crisis. *ASTIN Bulletin* 40(1):1–33.

Dowd, K. 1998. *Beyond Value at Risk: The New Science of Risk Management*. Wiley.

Drehmann, M., S. Sorenson and M. Stringa. 2010. The integrated impact of credit and interest rate risk on banks: a dynamic framework and stress testing application. *Journal of Banking and Finance* 34:713–729.

Driessen, J. 2005. Is default event risk priced in corporate bonds? *Review of Financial Studies* 18(1):165–195.

Duffee, G. 1999. Estimating the price of default risk. *Review of Financial Studies* 12:197–226.

Duffie, D. 2001. *Dynamic Asset Pricing Theory*, 3rd edn. Princeton University Press.

——. 2010. SEC v. Goldman Sachs: analyzing the complaint regarding the ABACUS 2007 AC-1 Deal. Slides, Stanford Graduate School of Business. (Available from www.darrell duffie.com/policy.cfm.)

Duffie, D. 2011. *Measuring Corporate Default Risk*. Oxford University Press.

Duffie, D. and N. Gârleanu. 2001. Risk and valuation of collateralized debt obligations. *Financial Analysts Journal* 57(1):41–59.

Duffie, D. and R. Kan. 1996. A yield-factor model of interest rates. *Mathematical Finance* 6(4):921–950.

Duffie, D. and D. Lando. 2001. Term structure of credit risk with incomplete accounting observations. *Econometrica* 69:633–664.

Duffie, D. and J. Pan. 1997. An overview of value at risk. *Journal of Derivatives* 4:7–49.

——. 2001. Analytical value-at-risk with jumps and credit risk. *Finance and Stochastics* 5:155–180.

Duffie, D. and K. Singleton. 1999. Modeling term structures of defaultable bonds. *Review of Financial Studies* 12:687–720.

——. 2003. *Credit Risk: Pricing, Measurement and Management*. Princeton University Press.

Duffie, D., D. Filipović and W. Schachermayer. 2003. Affine processes and applications in finance. *Annals of Applied Probability* 13:984–1053.

Duffie, D., J. Pan and K. Singleton. 2000. Transform analysis and asset pricing for affine jump diffusions. *Econometrica* 68:1343–1376.

Duffie, D., A. Eckner, G. Horel and L. Saita. 2009. Frailty correlated default. *Journal of Finance* 64:2089–2123.

Dunbar, N. 2000. *Inventing Money*. Wiley.

Dwyer, D. and S. Qu. 2007. EDF 8.0 model enhancements. Technical Document, Moody's Analytics.

Eberlein, E. and U. Keller. 1995. Hyperbolic distributions in finance. *Bernoulli* 1:281–299.

Eberlein, E., R. Frey and E. A. von Hammerstein. 2008. Advanced credit portfolio modelling and CDO pricing. In *Mathematics: Key Technology for the Future* (ed. W. Jäger and H. J. Krebs), pp. 256–279. Springer.

Eberlein, E., U. Keller and K. Prause. 1998. New insights into smile, mispricing, and value at risk: the hyperbolic model. *Journal of Business* 38:371–405.

Ebnöther, S., P. Vanini, A. J. McNeil and P. Antolinez-Fehr. 2003. Operational risk: a practitioner's view. *Journal of Risk* 5(3):1–15.

Eddy, W. F. 1984. Set-valued orderings for bivariate data. In *Stochastic Geometry, Geometric Statistics, Stereology* (ed. R. Ambartzumian and W. Weil). Teubner-Texte für Mathematik, vol. 56, pp. 79–90. Leipzig: Teubner.

Efron, B. F. and D. V. Hinkley. 1978. Assessing the accuracy of the maximum likleihood estimator: observed versus expected Fisher information. *Biometrika* 65:457–487.

Efron, B. F. and R. J. Tibshirani. 1994. *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Ehlers, P. and P. Schönbucher. 2009. Background filtrations and canonical loss processes for top-down models of portfolio credit risk. *Finance and Stochastics* 13:79–103.

Eisenberg, L. and T. H. Noe. 2001. Systemic risk in financial systems. *Management Science* 47(2):236–249.

El Karoui, N., M. Jeanblanc-Picqué and S. E. Shreve. 1998. Robustness of the Black and Scholes formula. *Mathematical Finance* 8:93–126.

El Karoui, N., D. Kurtz and Y. Jiao. 2008. Gauss and Poisson approximation: applications to CDO tranche pricing. *Journal of Computational Finance* 12:31–58.

Elliott, B. and J. Elliott. 2013. *Financial Accounting and Reporting*, 16th edn. Pearson.

Elliott, R. J., M. Jeanblanc and M. Yor. 2000. On models of default risk. *Mathematical Finance* 10:179–195.

Elsinger, H., A. Lehar and M. Summer. 2006. Risk assessment for banking systems. *Management Science* 52:1301–1314.

Embrechts, P. 2002. Insurance analytics. *British Actuarial Journal* 8(IV):639–641.

——. 2009. Copulas: a personal view. *Journal of Risk and Insurance* 76(3):639–650.

Embrechts, P. and C. M. Goldie. 1980. On closure and factorization properties of subexponential and related distributions. *Journal of the Australian Mathematical Society* A 29:43–256.

Embrechts, P. and G. Puccetti. 2006. Bounds for functions of dependent risks. *Finance and Stochastics* 10(3):341–352.

Embrechts, P., H. Furrer and R. Kaufmann. 2003. Quantifying regulatory capital for operational risk. *Derivatives Use, Trading and Regulation* 9(3):217–233.

Embrechts, P., R. Gruebel and S. Pitts. 1993. Some applications of the fast Fourier transform algorithm in insurance mathematics. *Statistica Neerlandica* 47:59–75.

Embrechts, P., A. Höing and G. Puccetti. 2005. Worst VaR scenarios. *Insurance: Mathematics and Economics* 37:115–134.

Embrechts, P., R. Kaufmann and P. Patie. 2005. Strategic long-term financial risks: single risk factors. *Computational Optimization and Applications* 32:61–90.

Embrechts, P., R. Kaufmann and G. Samorodnitsky. 2004. Ruin theory revisited: stochastic models for operational risk. In *Risk Management for Central Bank Foreign Reserves* (ed. C. Bernadell et al.), pp. 243–261. Frankfurt: European Central Bank.

Embrechts, P., C. Klüppelberg and T. Mikosch. 1997. *Modelling Extremal Events for Insurance and Finance*. Springer.

Embrechts, P., T. Liniger and L. Lin. 2011. Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability* 48A:367–378.

Embrechts, P., A. J. McNeil and D. Straumann. 1999. Correlation: pitfalls and alternatives. *Risk* 12(5):93–113.

——. 2002. Correlation and dependency in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond* (ed. M. Dempster), pp. 176–223. Cambridge University Press.

Embrechts, P., G. Puccetti and L. Rüschendorf. 2013. Model uncertainty and VaR aggregation. *Journal of Banking and Finance* 37(8):2750–2764.

Embrechts, P., B. Wang and R. Wang. 2014. Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, forthcoming.

Embrechts, P., J. L. Jensen, M. Maejima and J. L. Teugels. 1985. Approximations for compound Poisson and Pólya processes. *Advances in Applied Probability* 17:623–637.

Embrechts, P., G. Puccetti, L. Rüschendorf, R. Wang and A. Beleraj. 2014. An academic response to Basel 3.5. *Risks* 2(1):25–48.

Emmer, S., C. Klüppelberg and R. Korn. 2001. Optimal portfolios with bounded capital at risk. *Mathematical Finance* 11(4):365–384.

Engle, R. F. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50:987–1008.

——. 2002. Dynamic conditional correlation: a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* 20:339–350.

——. 2009. *Anticipating Correlations: A New Paradigm for Risk Management*. Princeton University Press.

Engle, R. F. and T. Bollerslev. 1986. Modeling the persistence of conditional variances. *Econometric Reviews* 5:1–50.

Engle, R. F. and K. F. Kroner. 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory* 11:122–150.

Engle, R. F. and K. Sheppard. 2001. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Technical Report 2001-15, University of California, San Diego.

Errais, E., K. Giesecke and L. Goldberg. 2010. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics* 1(1):642–665.

Fahrmeir, L. and G. Tutz. 1994. *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer.

Falk, M., J. Hüsler, and R. D. Reiss. 1994. *Laws of Small Numbers: Extremes and Rare Events*. Basel: Birkhäuser.

Fang, H. B. and K. T. Fang. 2002. The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis* 82:1–16.

Fang, K.-T., S. Kotz and K.-W. Ng. 1990. *Symmetric Multivariate and Related Distributions*. London: Chapman & Hall.

Federal Reserve Bank of Boston. 2005. Results of the 2004 loss data collection exercise for operational risk. Technical Report, Federal Reserve System. Office of the Comptroller of the Currency. Office of Thrift Supervision. Federal Deposit Insurance Corporation.

Feilmeier, M. and J. Bertram. 1987. *Anwendung numerischer Methoden in der Risikotheorie*. Karlsruhe: Verlag Versicherungswirtschaft.

Feller, W. 1971. *An Introduction to Probability Theory and Its Applications*, vol. II. Wiley.

Ferstl, R. and J. Hayden. 2011. Zero-coupon yield curve estimation with the package termstrc. *Journal of Statistical Software* 36(1):1–34.

Filipović, D. 1999. A note on the Nelson–Siegel family. *Mathematical Finance* 9:349–359.

——. 2009. *Term-Structure Models: A Graduate Course*. Springer.

Filipović, D., L. Overbeck and T. Schmidt. 2011. Dynamic CDO term structure modelling. *Mathematical Finance* 21:53–71.

FINMA. 2012. Summary report: FINMA investigation into the events surrounding trading losses of USD2.3 billion incurred by the investment banking division of UBS AG in London. Technical Report, Swiss Financial Market Supervisory Authority.

Fischer, T. 2003. Risk capital allocation by coherent risk measures based on one-sided moments. *Insurance: Mathematics and Economics* 32:135–146.

Fisher, R. A. and L. H. C. Tippett. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society* 24:180–190.

Fleming, T. and D. Harrington. 2005. *Counting Processes and Survival Analysis*, 2nd edn. Wiley.

Föllmer, H. 1994. Stock price fluctuation as a diffusion in a random environment. *Philosophical Transactions of the Royal Society of London* A 347:471–483.

Föllmer, H. and A. Schied. 2002. Convex measures of risk and trading constraints. *Finance and Stochastics* 6:429–447.

——. 2011. *Stochastic Finance: An Introduction in Discrete Time*, 4th edn. Berlin: Walter de Gruyter.

Föllmer, H. and M. Schweizer. 1991. Hedging of contingent claims under incomplete information. In *Applied Stochastic Analysis* (ed. M. H. A. Davis and R. J. Elliot), pp. 389–414. London: Gordon and Breach.

Föllmer, H. and D. Sondermann. 1986. Hedging of non-redundant contingent-claims. In *Contributions to Mathematical Economics* (ed. W. Hildenbrand and A. Mas-Colell), pp. 147–160. Amsterdam: North-Holland.

Fons, J. S. 1994. Using default rates to model the term structure of credit risk. *Financial Analysts Journal* 50:25–33.

Fornari, F. and A. Mele. 1997. Sign- and volatility-switching ARCH models: theory and applications to international stock markets. *Journal of Applied Econometrics* 12:49–65.

Fortin, I. and C. Kuzmics. 2002. Tail-dependence in stock–return pairs. *Intelligent Systems in Accounting, Finance and Management* 11:89–107.

Frachot, A. 2004. Operational risk: from theory to practice. Presentation, The Three Way Seminar, Versailles.

Frachot, A., P. Georges and T. Roncalli. 2001. Loss distribution approach for operational risk. Preprint. (Available from SSRN: 1032523.)

Frahm, G. 2004. Generalized Elliptical Distributions: Theory and Applications. PhD thesis, Faculty of Economics, Business Administration and Social Sciences, University of Cologne.

Frahm, G., M. Junker and A. Szimayer. 2003. Elliptical copulas: applicability and limitations. *Statistics and Probability Letters* 63:275–286.

Francq, C. and J.-M. Zakoïan. 2010. *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.

Frank, M. 1979. On the simultaneous additivity of $f(x, y)$ and $x + y - f(x, y)$. *Aequationes Mathematicae* 19:194–226.

Frank, M., R. Nelsen and B. Schweizer. 1987. Best-possible bounds on the distribution of a sum—a problem of Kolmogorov. *Probability Theory and Related Fields* 74:199–211.

Franke, G. and J. P. Kahnen. 2009. The future of securitization. In *Prudent Lending Restored: Securitization after the Mortgage Meltdown* (ed. Y. Fuchita and E. Litan). Washington, DC: Brookings Institution.

Fréchet, M. 1957. Les tableaux de corrélation dont les marges sont données. *Annales Université de Lyon Série A: Sciences Mathématiques et Astronomie* 4:13–31.

Frees, E. W. and E. A. Valdez. 1997. Understanding relationships using copulas. *North American Actuarial Journal* 2(1):1–25.

Frey, R. and J. Backhaus. 2008. Pricing and hedging of portfolio credit derivatives with interacting default intensities. *International Journal of Theoretical and Applied Finance* 11(6):611–634.

——. 2010. Dynamic hedging of synthetic CDO tranches with spread risk and default contagion. *Journal of Economic Dynamics and Control* 34:710–724.

Frey, R. and A. J. McNeil. 2001. Modelling dependent defaults. ETH E-Collection, ETH Zurich. (Available from www.e-collection.ethbib.ethz.ch/show?type=bericht&nr=273.)

——. 2002. VaR and expected shortfall in portfolios of dependent credit risks: conceptual and practical insights. *Journal of Banking and Finance* 26(7):1317–1344.

——. 2003. Dependent defaults in models of portfolio credit risk. *Journal of Risk* 6(1):59–92.

Frey, R. and L. Rösler. 2014. Contagion effects and collateralized credit value adjustments for credit default swaps. *International Journal of Theoretical and Applied Finance* 17(7):1450044.

Frey, R. and W. Runggaldier. 2010. Pricing credit derivatives under incomplete information: a nonlinear filtering approach. *Finance and Stochastics* 14:495–526.

Frey, R. and T. Schmidt. 2009. Pricing corporate securities under noisy asset information. *Mathematical Finance* 19:403–421.

——. 2011. Filtering and incomplete information in credit risk. In *Credit Risk Frontiers: Subprime Crisis, Pricing and Hedging, CVA, MBS, Ratings, and Liquidity* (ed. T. R. Bielecki, D. Brigo and F. Patras), pp. 185–218. Bloomberg Press.

——. 2012. Pricing and hedging of credit derivatives via the innovations approach to nonlinear filtering. *Finance and Stochastics* 12:105–133.

Frey, R. and R. Seydel. 2010. Optimal securitization of credit portfolios via impulse control. *Mathematics and Financial Economics* 4:1–28.

Frey, R., A. J. McNeil and M. Nyfeler. 2001. Copulas and credit models. *Risk* 14(10):111–114.

Frey, R., M. Popp and S. Weber. 2008. An approximation for credit portfolio losses. *Journal of Credit Risk* 4(1):3–20.

Frey, R., L. Rösler and D. Lu. 2014. Corporate security prices in structural credit risk models with incomplete information. Preprint, Institute of Statistics and Mathematics, WU Vienna.

Fristedt, B. and L. Gray. 1997. *A Modern Approach to Probability Theory*. Basel: Birkhäuser.

Frittelli, M. and Rosazza, E. 2002. Putting order in risk measures. *Journal of Banking and Finance* 26:1473–1486.

Froot, K. A. 2007. Risk management, capital budgeting and capital structure policy for insurers and reinsurers. *Journal of Risk and Insurance* 74(2):273–299.

Froot, K. A. and J. C. Stein. 1998. Risk management, capital budgeting and capital structure policy for financial institutions: an integrated approach. *Journal of Financial Economics* 47:55–82.

Froot, K. A., D. S. Scharfstein and J. C. Stein. 1993. Risk management: coordinating corporate investment and financing policies. *Journal of Finance* 48:1629–1658.

Frye, J. 2000. Depressing recoveries. Preprint, S&R 2000–8, Federal Reserve Bank of Chicago. (A short version appeared in *Risk* 13(11):108–111.)

Gagliardini, P. and C. Gouriéroux. 2005. Stochastic migration models with application to corporate risk. *Journal of Financial Econometrics* 3(2):188–226.

——. 2013. Granularity adjustment for risk measures: systematic vs unsystematic risks. *International Journal of Approximate Reasoning* 54(6):717–747.

Gai, P. and S. Kapadia. 2010. Contagion in financial networks. *Proceedings of the Royal Society of London* A 466:2401–2423.

Galambos, J. 1975. Order statistics of samples from multivariate distributions. *Journal of the American Statistical Association* 70:674–680.

——. 1987. *The Asymptotic Theory of Extreme Order Statistics*. Melbourne: Krieger.

Galbraith, J. K. 1993. *A Short History of Financial Euphoria*. New York: Whittle Books and Viking Penguin.

Gale, D. and L. S. Shapley. 1962. College admissions and stability of marriage. *American Mathematical Monthly* 69(1):9–15.

Gardner Jr, E. S. 1985. Exponential smoothing: the state of the art. *Journal of Forecasting* 2:1–21.

Geffroy, J. 1958. Contributions à la theorie des valeurs extrèmes. *Publications de l'Institut de Statistique de l'Université de Paris* 7–8:37–185.

Geman, H. 2005. *Commodities and Commodity Derivatives: Modelling and Pricing for Agriculturals, Metals and Energy*. Wiley.

——. 2009. *Risk Management in Commodity Markets: From Shipping to Agriculturals and Energy*. Wiley.

Gençay, R. and F. Selçuk. 2004. Extreme value theory and value-at-risk: relative performance in emerging markets. *International Journal of Forecasting* 20:287–303.

Gençay, R., F. Selçuk and A. Ulugülyağci. 2003. High volatility, thick tails and extreme value theory in value-at-risk estimation. *Insurance: Mathematics and Economics* 33:337–356.

Genest, C. and J. MacKay. 1986. The joy of copulas: bivariate distributions with uniform marginals. *American Statistician* 40:280–285.

Genest, C. and J. Nešlehová. 2007. A primer on copulas for count data. *ASTIN Bulletin* 37:475–515.

Genest, C. and L. Rivest. 1993. Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association* 88:1034–1043.

Genest, C., M. Gendron and M. Bourdeau-Brien. 2009. The advent of copulas in finance. *European Journal of Finance* 15(7–8):609–618.

Genest, C., K. Ghoudi and L. Rivest. 1995. A semi-parametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82:543–552.

——. 1998. Commentary on "Understanding relationships using copulas" by E. W. Frees and E. A. Valdez. *North American Actuarial Journal* 2:143–149.

Genest, C., B. Rémillard and D. Beaudoin. 2009. Goodness-of-fit tests for copulas: a review and a power study. *Insurance: Mathematics and Economics* 44:199–213.

Genton, M. G. (ed.). 2004. *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton, FL: Chapman & Hall.

Gerber, H. U. 1997. *Life Insurance Mathematics*, 3rd edn. Springer.

Gibson, M. S. 2004. Understanding the risk of synthetic CDOs. Working Paper, Federal Reserve Board.

Gibson, R. (ed.). 2000. *Model Risk: Concepts, Calibration and Pricing*. London: Risk Books.

Giese, G. 2003. Enhancing CreditRisk$^+$. *Risk* 16(4):73–77.

Giesecke, K. and L. R. Goldberg. 2004. Sequential defaults and incomplete information. *Journal of Risk* 7(1):1–26.

Giesecke, K. and S. Weber. 2004. Cyclical correlation, credit contagion and portfolio losses. *Journal of Banking and Finance* 28:3009–3036.

——. 2006. Credit contagion and aggregate losses. *Journal of Economic Dynamics and Control* 30:741–767.

Giesecke, K., L. R. Goldberg and X. Ding. 2011. A top-down approach to multi-name credit. *Operations Research* 59:283–300.

Giné, E. M. 1975. Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Annals of Statistics* 3(6):1243–1266.

Giri, N. C. 1996. *Multivariate Statistical Analysis*. New York: Marcel Dekker.

Glasserman, P. 2003. *Monte Carlo Methods in Financial Engineering*. Springer.

——. 2004. Tail approximations for portfolio credit risk. *Journal of Derivatives* 12:24–42.

Glasserman, P. and J. Li. 2003. Importance sampling for a mixed Poisson model of portfolio credit risk. In *Proceedings of the 2003 Winter Simulation Conference* (ed. S. Chick, P. J. Sánchez, D. Ferrin and D. J. Morrice).

——. 2005. Importance sampling for portfolio credit risk. *Management Science* 51:1643–1656.

Glasserman, P., P. Heidelberger and P. Shahabuddin. 1999. Importance sampling and stratification for value at risk. In *Proceedings of the 6th International Conference on Computational Finance* (ed. Y. S. Abu-Mostafa, B. LeBaron, A. W. Lo and A. S. Weigand). Cambridge, MA: MIT Press.

Glosten, L. R., R. Jagannathan and D. E. Runkle. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48(5):1779–1801.

Gnanadesikan, R. 1997. *Methods for Statistical Data Analysis of Multivariate Observations*, 2nd edn. Wiley.

Gnanadesikan, R. and J. R. Kettenring. 1972. Robust estimates, residuals and outlier detection with multiresponse data. *Biometrics* 28:81–124.

Gnedenko, B. V. 1941. Limit theorems for the maximal term of a variational series. *Comptes Rendus (Doklady) de L'Académie des Sciences de l'URSS* 32:7–9.

——. 1943. Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics* 44:423–453.

Gneiting, T. 2011. Making and evaluating point forecasts. *Journal of the American Statistical Association* 106(494):746–762.

Goovaerts, M. J., F. De Vylder and J. Haezendonck. 1984. *Insurance Premiums*. Amsterdam: North-Holland.

Goovaerts, M. J., J. Dhaene and R. Kaas. 2003. Capital allocation derived from risk measures. *North American Actuarial Journal* 7:44–49.

Goovaerts, M. J., E. van den Borre and R. J. A. Laeven. 2005. Managing economic and virtual economic capital within financial conglomerates. *North American Actuarial Journal* 9:77–89.

Goovaerts, M. J., R. Kaas, J. Dhaene and Q. Tang. 2003. A unified approach to generate risk measures. *ASTIN Bulletin* 33(2):173–191.

Gordy, M. B. 2000. A comparative anatomy of credit risk models. *Journal of Banking and Finance* 24:119–149.

——. 2002. Saddlepoint approximation of CreditRisk$^+$. *Journal of Banking and Finance* 26:1335–1353.

——. 2003. A risk-factor model foundation for ratings-based capital rules. *Journal of Financial Intermediation* 12(3):199–232.

——. 2004. Granularity adjustment in portfolio credit risk measurement. In *Risk Measures for the 21st Century* (ed. G. Szegö), pp. 109–121. Wiley.

Gordy, M. B. and E. Heitfield. 2002. Estimating default correlations from short panels of credit rating performance data. Technical Report, Federal Reserve Board.

Gordy, M. B. and E. Lütkebohmert. 2013. Granularity adjustment for regulatory capital assessment. *International Journal of Central Banking* 9(3):33–71.

Gordy, M. B. and J. Marrone. 2012. Granularity adjustment for mark-to-market credit risk models. *Journal of Banking and Finance* 36(7):1896–1910.

Gouriéroux, C. 1997. *ARCH Models and Financial Applications*. Springer.

Gouriéroux, C. and J. Jasiak. 2001. *Financial Econometrics: Problems, Models and Methods*. Princeton University Press.

Gouriéroux, C., J. P. Laurent and O. Scaillet. 2000. Sensitivity analysis of values at risk. *Journal of Empirical Finance* 7:225–245.

Gouriéroux, C., A. Montfort and A. Trognon. 1984. Pseudo maximum likelihood methods: theory. *Econometrica* 52:681–700.

Graham, J. R. and D. A. Rogers. 2002. Do firms hedge in response to tax incentives? *Journal of Finance* 57(2):815–839.

Grandell, J. 1997. *Mixed Poisson Processes*. London: Chapman & Hall.

Greenspan, A. 2002. Speech before the Council on Foreign Relations. In *International Financial Risk Management, Washington, DC*, 19 November.

Gregory, J. (ed.). 2003. *Credit Derivatives: The Definitive Guide*. London: Risk Books.

Gregory, J. 2012. *Counterparty Credit Risk and Credit Value Adjustment: A Continuing Challenge for Global Financial Markets*. Wiley.

Grothe, O., V. Korniichuk and H. Manner. 2014. Modeling multivariate extreme events using self-exciting point processes. *Journal of Econometrics* 182(2):269–289.

Grübel, R. and R. Hermesmeier. 1999. Computation of compound distributions. I. Aliasing errors and exponential tilting. *ASTIN Bulletin* 29(2):197–214.

——. 2000. Computation of compound distributions. II. Discretization errors and Richardson extrapolation. *ASTIN Bulletin* 30(2):309–331.

Guégan, D. and J. Houdain. 2005. Collateralized debt obligations pricing and factor models: a new methodology using normal inverse Gaussian distributions. Preprint, ENS, Cachan.

Gumbel, E. J. 1958. *Statistics of Extremes*. New York: Columbia University Press.

Haaf, H., O. Reiss and J. Schoenmakers. 2004. Numerically stable computation of CreditRisk$^+$. *Journal of Risk* 6(4):1–10.

Haan, L. de. 1985. Extremes in higher dimensions: the model and some statistics. In *Proceedings of the 45th Session International Statistical Institute*, paper 26.3. The Hague: International Statistical Institute.

Haan, L. de and A. Ferreira. 2000. *Extreme Value Theory: An Introduction*. Springer.

Haan, L. de and S. I. Resnick. 1977. Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie and verwandte Gebiete* 40:317–337.

Haan, L. de and H. Rootzén. 1993. On the estimation of high quantiles. *Journal of Statistical Planning and Inference* 35:1–13.

Haan, L. de, S. I. Resnick, H. Rootzén and C. G. de Vries. 1989. Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stochastic Processes and Their Applications* 32:213–224.

Hafner, C. M. and P. H. Franses. 2009. A generalized dynamic conditional correlation model: simulation and application to many assets. *Econometric Reviews* 28(6):612–631.

Hall, P. G. 1982. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 44:37–42.

——. 1990. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis* 32:177–203.

Hamilton, J. 1994. *Time Series Analysis*. Princeton University Press.

Hamilton, R., S. Varma, S. Ou and R. Cantor. 2005. Default and recovery rates of corporate cbond issuers. Technical Report, Moody's Investor Service.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel. 1986. *Robust Statistics: The Approach Based on Influence Functions*. Wiley.

Hancock, J., P. Huber and P. Koch. 2001. Value creation in the insurance industry. *Journal of Risk and Insurance* 4(2):1–9.

Hand, D. J. and W. E. Henley. 1997. Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 160(3):523–541.

Hardy, M. 2003. *Investment Guarantees*. Wiley.

Harrison, J. M. and D. M. Kreps. 1979. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* 20:381–408.

Harrison, J. M. and S. R. Pliska. 1981. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications* 11:215–260.

Haug, E. G. 1998. *Option Pricing Formulas*. New York: McGraw-Hill.

Hawke Jr, J. D. 2003. Remarks before the Institute of International Bankers, 3 March 2003, Washington, DC. (Available from www.occ.treas.gov/ftp/release/2003-15a.pdf/.)

Hawkes, A. G. 1971. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 33:438–443.

He, X. D. and X. Y. Zhou. 2011. Portfolio choice under cumulative prospect theory: an analytical treatment. *Management Science* 57:315–331.

Heath, D., R. A. Jarrow and A. Morton. 1992. Bond pricing and the term structure of interest rates: a new methodology for contingent claim valuation. *Econometrica* 60(1):77–105.

Heffernan, J. E. and J. A. Tawn. 2004. A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66(3):497–546.

Herbertsson, A. 2008. Pricing synthetic CDO tranches in a model with default contagion using the matrix-analytic approach. *Journal of Credit Risk* 4(4):3–35.

Hesselager, O. 1996. Recursions for certain bivariate counting distributions and their compound distributions. *ASTIN Bulletin* 26(1):35–52.

Hilberink, B. and L. C. G. Rogers. 2002. Optimal capital structure and endogenous default. *Finance and Stochastics* 6(2):237–263.

Hill, B. M. 1975. A simple general approach to inference about the tail of a distribution. *Annals of Statistics* 3:1163–1174.

Hofert, M. 2008. Sampling Archimedean copulas. *Computational Statistics and Data Analysis* 52:5163–5174.

——. 2011. Efficiently sampling nested Archimedean copulas. *Computational Statistics and Data Analysis* 55:57–70.

——. 2012. A stochastic representation and sampling algorithm for nested Archimedean copulas. *Journal of Statistical Computation and Simulation* 82(9):1239–1255.

Hofert, M. and M. Mächler. 2011. Nested Archimedean copulas meet R: the nacopula package. *Journal of Statistical Software* 39(9):1–30.

——. 2014. A graphical goodness-of-fit test for dependence models in higher dimensions. *Journal of Computational and Graphical Statistics* 23:700–716.

Hofert, M., M. Mächler and A. J. McNeil. 2012. Archimedean copulas in high dimensions: estimators and numerical challenges motivated by financial applications. *Journal de la Société Française de Statistique* 153(3):25–63.

Höffding, W. 1940. Massstabinvariante Korrelationstheorie. *Schriften des Mathematischen Seminars und des Instituts für Angewandte Mathematik der Universität Berlin* 5:181–233.

Hogg, R. V. and S. A. Klugman. 1984. *Loss Distributions*. Wiley.

Horst, U. 2007. Stochastic cascades, credit contagion and large portfolio losses. *Journal of Economic Behavior and Organization* 63:25–54.

Hosking, J. R. M. 1985. Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution. *Applied Statistics* 34:301–310.

Hosking, J. R. M. and J. R. Wallis. 1987. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* 29:339–349.

Hosking, J. R. M., J. R. Wallis and E. F. Wood. 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* 27:251–261.

Hsing, T. 1991. On tail index estimation using dependent data. *Annals of Statistics* 19:1547–1569.

Hu, Y.-T., R. Kiesel and W. Perraudin. 2002. The estimation of transition matrices for sovereign credit ratings. *Journal of Banking and Finance* 26:1383–1406.

Huang, J.-Z. and M. Huang. 2012. How much of the corporate-treasury yield spread is due to credit risk? *Review of Asset Pricing Studies* 2(2):153–202.

Huang, J.-Z., M. Lanfranconi, N. Patel and L. Pospisil. 2012. Modeling credit correlations: an overview of the Moody's Analytics GCorr Model. Technical Report, Moody's Analytics.

Huber, P. J. and E. M. Ronchetti. 2009. *Robust Statistics*, 2nd edn. Wiley.

Hull, J. C. 1997. *Options, Futures and Other Derivatives*, 3rd edn. Englewood Cliffs, NJ: Prentice Hall.

——. 2009. The credit crunch of 2007: what went wrong? Why? What lessons can be learned? *Journal of Credit Risk* 5(2):3–18.

——. 2014. *Options, Futures and Other Derivatives*, 9th edn. Englewood Cliffs, NJ: Prentice Hall.

Hull, J. C. and A. White. 1998. Incorporating volatility updating into the historical simulation method for value-at-risk. *Journal of Risk* 1(1):5–19.

——. 2001. Valuing credit default swaps. II. Modelling default correlations. *Journal of Derivatives* 8(3):12–22.

——. 2004. Valuation of a CDO and an *n*th to default CDS without Monte Carlo simulation. *Journal of Derivatives* 12:8–23.

——. 2006. The implied copula model. *Journal of Derivatives* 14:8–28.

——. 2010. The risk of tranches created from mortgages. *Financial Analysts Journal* 66:54–67.

——. 2012. CVA and wrong-way risk. *Financial Analysts Journal* 68:58–69.

——. 2013. Collateral and credit issues in derivative pricing. Preprint, Rotman School of Management.

Hult, H. and F. Lindskog. 2002. Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability* 34:587–608.

Hult, H., F. Lindskog, O. Hammarlind and C. J. Rehn. 2012. *Risk and Portfolio Analysis: Principles and Methods*. Springer.

Hürlimann, W. 2008a. Extremal moment methods and stochastic orders. Application in actuarial science: chapters I, II and III. *Boletín de la Asociación Matemática Venezolana* 15(1):5–110.

Hürlimann, W. 2008b. Extremal moment methods and stochastic orders. Application in actuarial science: chapters IV, V and VI. *Boletín de la Asociación Matemática Venezolana* 15(2):153–301.

Hutchinson, T. P. and C. D. Lai. 1990. *Continuous Bivariate Distributions, Emphasizing Applications*. Adelaide: Rumsby Scientific.

Hyndman, R. J. and Y. Fan. 1996. Sample quantiles in statistical packages. *American Statistician* 50:361–365.

IFRI Foundation and CRO Forum. 2007. Insights from the joint IFRI/CRO Forum survey on economic capital practice and applications. KPMG Business Advisory Services.

Iman, R. L. and W. J. Conover. 1982. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics: Simulation and Computation* 11:311–334.

Jacque, L. L. 2010. *Global Derivatives Debacles: From Theory to Malpractice*. World Scientific.

Jaeger, L. 2005. *Through the Alpha Smoke Screens: A Guide to Hedge Fund Return Sources*. New York: Euromoney Institutional Investor.

Janson, S., S. M'Baye and P. Protter. 2011. Absolutely continuous compensators. *International Journal of Theoretical and Applied Finance* 14:335–351.

Jarque, C. M. and A. K. Bera. 1987. A test for normality of observations and regression residuals. *International Statistical Review* 55(5):163–172.

Jarrow, R. A. and P. Protter. 2004. Structural versus reduced-form models: a new information based perspective. *Journal of Investment Management* 2:1–10.

Jarrow, R. A. and S. M. Turnbull. 1995. Pricing derivatives on financial securities subject to credit risk. *Journal of Finance* 50(1):53–85.

——. 1999. *Derivative Securities*, 2nd edn. Cincinnati, OH: South Western Publishing.

Jarrow, R. A. and F. Yu. 2001. Counterparty risk and the pricing of defaultable securities. *Journal of Finance* 53:2225–2243.

Jarrow, R. A., D. Lando and S. M. Turnbull. 1997. A Markov model for the term structure of credit risk spreads. *Review of Financial Studies* 10:481–523.

Jarrow, R. A., D. Lando and F. Yu. 2005. Default risk and diversification: theory and empirical implications. *Mathematical Finance* 15:1–26.

Jeanblanc, M., M. Yor and M. Chesney. 2009. *Mathematical Methods for Financial Markets*. Springer.

Jeantheau, T. 1998. Strong consistency of estimators of multivariate ARCH models. *Econometric Theory* 14:70–86.

Jenkins, S. and K. Roberts (eds). 2004. *Advances in Operational Risk: Firm-Wide Issues for Financial Institutions*, 2nd edn. London: Risk Books.

Jensen, J. L. 1995. *Saddlepoint Approximations*. Oxford: Clarendon Press.

Jewell, W. S. 1990. Credibility prediction of first and second moments in the hierarchical model. In *Festgabe Bühlmann*, pp. 81–110. Bern: Stämpfli.

Joag-Dev, K. 1984. Measures of dependence. In *Handbook of Statistics* (ed. P. R. Krishnaiah), vol. 4, pp. 79–88. New York: North-Holland/Elsevier.

Joe, H. 1993. Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis* 46(2):262–282.

——. 1997. *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Joe, H., R. L. Smith and I. Weissman. 1992. Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 54(1):171–183.

Joenssen, D. W. and J. Vogel. 2014. A power study of goodness-of-fit tests for multivariate normality implemented in R. *Journal of Statistical Computation and Simulation* 84:1055–1078.

Johnson, N. L. and S. Kotz. 1969. *Distributions in Statistics: Discrete Distributions*. Boston, MA: Houghton Mifflin.

Johnson, R. A. and D. W. Wichern. 2002. *Applied Multivariate Statistical Analysis*, 5th edn. Upper Saddle River, NJ: Prentice Hall.

Jones, M. C. 1994. Expectiles and M-quantiles are quantiles. *Statistics and Probability Letters* 20:149–153.

Jorion, P. 2000. Risk management lessons from Long-Term Capital Management. *European Financial Management* 6:277–300.

——. 2002a. Fallacies about the effects of market risk management systems. *Journal of Risk* 5(1):75–96.

——. 2002b. *Financial Risk Manager Handbook 2001–2002 (GARP)*. Wiley.

——. 2007. *Value at Risk: The New Benchmark for Measuring Financial Risk*, 3rd edn. New York: McGraw-Hill.

JPMorgan. 1996. *RiskMetrics Technical Document*, 4th edn. New York: JPMorgan.

Juri, A. and M. V. Wüthrich. 2002. Copula convergence theorems for tail events. *Insurance: Mathematics and Economics* 30:411–427.

——. 2003. Tail dependence from a distributional point of view. *Extremes* 6:213–246.

Kaas, R., M. J. Goovaerts, J. Dhaene and M. Denuit. 2001. *Modern Actuarial Risk Theory*. Boston, MA: Kluwer Academic.

Kalemanova, A., B. Schmid and R. Werner. 2005. The normal inverse Gaussian distribution for synthetic CDO pricing. *Journal of Derivatives* 14:80–94.

Kalkbrener, M. 2005. An axiomatic approach to capital allocation. *Mathematical Finance* 15:425–437.

Kalkbrener, M., H. Lotter and L. Overbeck. 2004. Sensible and efficient capital allocation for credit portfolios. *Risk* 17(1):S19–S24.

Kallenberg, O. 1983. *Random Measures*, 3rd edn. London: Academic Press.

Karatzas, I. and S. E. Shreve. 1988. *Brownian Motion and Stochastic Calculus*. Springer.

Kariya, T. and M. L. Eaton. 1977. Robust tests for spherical symmetry. *Annals of Statistics* 5(1):206–215.

Karr, A. F. 1991. *Point Processes and Their Statistical Inference*, 2nd edn. New York: Marcel Dekker.

Kaufmann, R. 2004. Long-Term Risk Management. PhD thesis, ETH Zurich.

Kaufmann, R., R. Gadmer and R. Klett. 2001. Introduction to dynamical financial analysis. *ASTIN Bulletin* 31(1):213–249.

Kelker, D. 1970. Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā* A 32:419–430.

Kerkhof, J. and B. Melenberg. 2004. Backtesting for risk-based regulatory capital. *Journal of Banking and Finance* 28(8):1845–1865.

Kesten, H. 1973. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica* 131:207–248.

Keynes, J. M. 1920. *A Treatise on Probability*. London: Macmillan.

Khoudraji, A. 1995. Contribution à l'Étude des Copules et à la Modélisation des Copules de Valeurs Extrèmes Bivariées. PhD thesis, Université Laval, Québec.

Kiesel, R., M. Scherer and R. Zagst (eds). 2010. *Alternative Investments and Strategies*. World Scientific.

Kimberling, C. H. 1974. A probabilistic interpretation of complete monotonicity. *Aequationes Mathematicae* 10:152–164.

Kindleberger, C. P. 2000. *Manias, Panics and Crashes: A History of Financial Crises*, 4th edn. Wiley.

King, J. L. 2001. *Measurement and Modelling Operational Risk*. Wiley.

Klaassen, P. and I. van Eeghen. 2009. *Economic Capital: How It Works and What Every Manager Needs to Know*. Amsterdam: Elsevier.

Kloman, H. F. 1990. Risk management agonists. *Risk Analysis* 10:201–205.

Klugman, S. A. and R. Parsa. 1999. Fitting bivariate loss distributions with copulas. *Insurance: Mathematics and Economics* 24:139–148.

Klugman, S. A., H. H. Panjer and G. E. Willmot. 2012. *Loss Models: From Data to Decisions*, 4th edn. Wiley.

——. 2013. *Loss Models: Further Topics*. Wiley.

Knight, F. H. 1921. *Risk, Uncertainty and Profit*. Boston, MA: Houghton Mifflin.

Koedijk, K. G., M. M. A. Schafgans and C. G. de Vries. 1990. The tail index of exchange rate returns. *Journal of International Economics* 29:93–108.

Koller, M. 2011. *Life Insurance Risk Management Essentials*. Springer.

Kolmogorov, A. N. 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer.

Koopman, S. J., A. Lucas and P. Klaassen. 2005. Empirical credit cycles and capital buffer formation. *Journal of Banking and Finance* 29:3159–3179.

Koopman, S. J., A. Lucas and A. Monteiro. 2008. The multi-state latent factor intensity model for credit rating transitions. *Journal of Econometrics* 142:399–424.

Koopman, S. J., A. Lucas and B. Schwaab. 2012. Dynamic factor models with macroeconomic, frailty and industry effects for U.S. default counts: the credit crisis of 2008. *Journal of Business and Economic Statistics* 30(4):521–532.

Koryciorz, S. 2004. *Sicherheitskapitalbestimmung und -allokation in der Schadenversicherung*. Karlsruhe: Verlag Versicherungswirtschaft.

Kotz, S. and S. Nadarajah. 2000. *Extreme Value Distributions, Theory and Applications*. London: Imperial College Press.

——. 2004. *Multivariate t Distributions and Their Applications*. Cambridge University Press.

Kotz, S., N. L. Johnson and C. B. Read (eds). 1985. *Encyclopedia of Statistical Sciences*. Wiley.

Kotz, S., T. J. Kozubowski and K. Podgórski. 2001. *The Laplace Distribution and Generalizations*. Boston, MA: Birkhäuser.

Koyluoglu, U. and A. Hickman. 1998. Reconciling the differences. *Risk* 11(10):56–62.

Kretzschmar, G., A. J. McNeil and A. Kirchner. 2010. Integrated models of capital adequacy: why banks are undercapitalised. *Journal of Banking and Finance* 34:2838–2850.

Krokhmal, P., J. Palmquist and S. Uryasev (2001). Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk* 4(2):43–68.

Krupskii, P. and H. Joe. 2013. Factor copula models for multivariate data. *Journal of Multivariate Analysis* 120:85–101.

Kruskal, W. H. 1958. Ordinal measures of association. *Journal of the American Statistical Association* 53:814–861.

Kupiec, P. H. 1995. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* 3(2):73–84.

——. 1998. Stress testing in a value at risk framework. *Journal of Derivatives* 6:7–24.

Kurowicka, D. and R. Cooke. 2006. *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley.

Kurth, A. and D. Tasche. 2003. Contributions to credit risk. *Risk* 16(3):84–88.

Kusuoka, S. 1999. A remark on default risk models. *Advances in Mathematical Economics* 1:69–81.

——. 2001. On law invariant coherent risk measures. *Advances in Mathematical Economics* 3:83–95.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* 54:159–178.

Lamberton, D. and B. Lapeyre. 1996. *Introduction to Stochastic Calculus Applied to Finance*. London: Chapman & Hall.

Lando, D. 1998. Cox processes and credit risky securities. *Review of Derivatives Research* 2:99–120.

——. 2004. *Credit Risk Modeling: Theory and Applications*. Princeton University Press.

Lando, D. and M. S. Nielsen. 2010. Correlation in corporate defaults: contagion or conditional independence? *Journal of Financial Intermediation* 19:355–372.

Lando, D. and T. Skodeberg. 2002. Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking and Finance* 26:423–444.

Landsman, Z. M. and E. A. Valdez. 2003. Tail conditional expectations for elliptical distributions. *North American Actuarial Journal* 7(4):55–71.

Lang, L. and R. Stulz. 1992. Contagion and competitive intra-industry effects of bankruptcy announcements. *Journal of Financial Economics* 32:45–60.

Laurent, J.-P. and J. Gregory. 2005. Basket default swaps, CDOs and factor copulas. *Journal of Risk* 7:103–122.

Laurent, J.-P., A. Cousin and J.-D. Fermanian. 2011. Hedging default risks of CDOs in Markovian contagion models. *Quantitative Finance* 11:1773–1791.

Laux, C. and C. Leuz. 2010. Did fair-value accounting contribute to the financial crisis? *Journal of Economic Perspectives* 24:93–118.

Leadbetter, M. R. 1983. Extremes and local dependence of stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie and verwandte Gebiete* 65:291–306.

——. 1991. On a basis for peaks over threshold modeling. *Statistics and Probability Letters* 12:357–362.

Leadbetter, M. R., G. Lindgren and H. Rootzén. 1983. *Extremes and Related Properties of Random Sequences and Processes*. Springer.

Ledford, A. W. and J. A. Tawn. 1996. Statistics for near independence in multivariate extreme values. *Biometrika* 83(1):169–187.

Lee, S. and B. Hansen. 1994. Asymptotic properties of the maximum likelihood estimator and test of the stability of parameters of GARCH and IGARCH models. *Econometric Theory* 10:29–52.

Lehmann, E. L. 1983. *Theory of Point Estimation*. Wiley.

——. 1986. *Testing Statistical Hypotheses*, 2nd edn. Wiley.

Leland, H. E. 1994. Corporate debt value, bond covenants and optimal capital structure. *Journal of Finance* 49(4):157–196.

Leland, H. E. and K. Toft. 1996. Optimal capital structure, endogenous bankruptcy and the term structure of credit spreads. *Journal of Finance* 51:987–1019.

Lemaire, J. 1984. An application of game theory: cost allocation. *ASTIN Bulletin* 14(1):61–82.

Leoni, P. 2014. *The Greeks and Hedging Explained*. Basingstoke: Palgrave Macmillan.

Letac, G. 2004. From Archimedes to statistics: the area of the sphere. Preprint, Université Paul Sabatier, Toulouse.

Li, D. 2000. On default correlation: a copula function approach. *Journal of Fixed Income* 9:43–54.

Li, R.-Z., K.-T. Fang and L.-X. Zhu. 1997. Some Q–Q probability plots to test spherical and elliptical symmetry. *Journal of Computational and Graphical Statistics* 6(4):435–450.

Lindsey, J. K. 1996. *Parametric Statistical Inference*. Oxford University Press.

Lindskog, F. 2000. Linear correlation estimation. RiskLab Report, ETH Zurich.

Lindskog, F. and A. J. McNeil. 2003. Common Poisson shock models: applications to insurance and credit risk modelling. *ASTIN Bulletin* 33(2):209–238.

Lindskog, F., A. J. McNeil and U. Schmock. 2003. Kendall's tau for elliptical distributions. In *Credit Risk: Measurement, Evaluation and Management* (ed. G. Bol, G. Nakhaeizadeh, S. T. Rachev, T. Ridder and K.-H. Vollmer), pp. 149–156. Heidelberg: Physica.

Lipton, A. and A. Rennie (eds). 2008. *Credit Correlation: Life After Copulas*. World Scientific.

Lipton, A. and A. Sepp. 2009. Credit value adjustments for credit default swaps via the structural default model. *Journal of Credit Risk* 5(2):127–150.

Liu, C. 1997. ML estimation of the multivariate t distribution and the EM algorithm. *Journal of Mathematical Economics* 63:296–312.

Liu, C. and D. B. Rubin. 1994. The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika* 81:633–648.

——. 1995. ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Statistica Sinica* 5:19–39.

Ljung, G. M. and G. E. P. Box. 1978. On a measure of lack of fit in time series models. *Biometrika* 65:297–303.

Lo, A. W. and A. C. MacKinlay. 1999. *A Non-Random Walk Down Wall Street*. Princeton University Press.

Longin, F. M. 1996. The asymptotic distribution of extreme stock market returns. *Journal of Business* 69:383–408.

Longstaff, F. A. and E. S. Schwartz. 1995. Valuing risky debt: a new approach. *Journal of Finance* 50:789–821.

Lord Penrose. 2004. *Report of the Equitable Life Inquiry*. London: The Stationery Office.

Lord Turner. 2009. *The Turner Review: A Regulatory Response to the Global Banking Crisis*. London: Financial Services Authority.

Loretan, M. and W. B. English. 2000. Evaluating correlation breakdowns during periods of market volatility. Technical Report 658, FRB International Finance Discussion Paper, Federal Reserve Board.

Lowenstein, R. 2000. *When Genius Failed: The Rise and Fall of Long-Term Capital Management*. London: Random House.

Lucas, D. J. 1995. Default correlation and credit analysis. *Journal of Fixed Income* 4(4):76–87.

Lumsdaine, R. 1996. Asymptotic properties of the quasi maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models. *Econometrica* 64:575–596.

Luo, X. and P. V. Shevchenko. 2010. The t copula with multiple parameters of degrees of freedom: bivariate characteristics and application to risk management. *Quantitative Finance* 10(9):1039–1054.

——. 2012. Bayesian model choice of grouped t-copula. *Methodology and Computing in Applied Probability* 14(4):1097–1119.

Lütkebohmert, E. 2009. *Concentration Risk in Credit Portfolios*. EAA Lecture Notes. Springer.

Lütkepohl, H. 1993. *Introduction to Multiple Time Series Analysis*, 2nd edn. Springer.

Lux, T. 1996. The stable Paretian hypothesis and the frequency of large returns: an analysis of major German stocks. *Applied Financial Economics* 6:463–475.

McCullagh, P. and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd edn. London: Chapman & Hall.

McCullough, B. D. and C. G. Renfro. 1999. Benchmarks and software standards: a case study of GARCH procedures. *Journal of Economic and Social Measurement* 25:59–71.

McLeish, D. L. and C. G. Small. 1988. *The Theory and Applications of Statistical Inference Functions*. Springer.

McNeil, A. J. 1997. Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin* 27:117–137.

——. 1998. History repeating. *Risk* 11(1):99.

——. 2008. Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation* 78(6):567–581.

McNeil, A. J. and R. Frey. 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* 7:271–300.

McNeil, A. J. and J. Nešlehová. 2009. Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. *Annals of Statistics* 37(5b):3059–3097.

——. 2010. From Archimedean to Liouville distributions. *Journal of Multivariate Analysis* 101(8):1772–1790.

McNeil, A. J. and T. Saladin. 2000. Developing scenarios for future extreme losses using the POT method. In *Extremes and Integrated Risk Management* (ed. P. Embrechts), pp. 253–267. London: Risk Books.

McNeil, A. J. and A. D. Smith. 2012. Multivariate stress scenarios and solvency. *Insurance: Mathematics and Economics* 50(3):299–308.

McNeil, A. J. and J. Wendin. 2006. Dependent credit migration. *Journal of Credit Risk* 2(3):87–114.

——. 2007. Bayesian inference for generalized linear mixed models of portfolio credit risk. *Journal of Empirical Finance* 14(2):131–149.

Madan, D. B. and E. Seneta. 1990. The variance gamma (VG) model for share market returns. *Journal of Business* 63:511–524.

Madan, D. B., P. P. Carr and E. C. Chang. 1998. The variance gamma process and option pricing. *European Finance Review* 2:79–105.

Mai, J.-F. and M. Scherer. 2012. *Simulating Copulas: Stochastic Models, Sampling Algorithms and Applications*. World Scientific.

Makarov, G. 1981. Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory of Probability and Its Applications* 26:803–806.

Marazzi, A. 1993. *Algorithms, Routines and S-Functions for Robust Statistics*. New York: Chapman & Hall.

Mardia, K. V. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57:519–530.

——. 1972. *Statistics of Directional Data*. New York: Academic Press.

——. 1974. Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā* 36:115–128.

——. 1975. Assessment of multinormality and the robustness of Hotelling's T^2 test. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 24:163–171.

Mardia, K. V., J. T. Kent and J. M. Bibby. 1979. *Multivariate Analysis*. London: Academic Press.

Markowitz, H. M. 1952. Portfolio selection. *Journal of Finance* 7:77–91.

——. 1959. *Portfolio Selection: Efficient Diversification of Investments*. Wiley.

Maronna, R. A. 1976. Robust M-estimators of multivariate location and scatter. *Annals of Statistics* 4:51–67.

Marshall, A. W. 1996. Copulas, marginals and joint distributions. In *Distributions with Fixed Marginals and Related Topics* (ed. L. Rüschendorff, B. Schweizer and M. D. Taylor), pp. 213–222. Hayward, CA: Institute of Mathematical Statistics.

Marshall, A. W. and I. Olkin. 1967a. A generalized bivariate exponential distribution. *Journal of Applied Probability* 4:291–302.

——. 1967b. A multivariate exponential distribution. *Journal of the American Statistical Association* 62:30–44.

——. 1988. Families of multivariate distributions. *Journal of the American Statistical Association* 83:834–841.

——. 2007. *Life Distributions: Structure of Nonparametric, Semiparametric, and Parametric Families*. Springer.

Martin, R. and T. Wilde. 2002. Unsystematic credit risk. *Risk* 15(11):123–128.

Martin, R., K. Thompson and C. Browne. 2001. Taking to the saddle. *Risk* 14(6):91–94.

Mas-Colell, A., M. D. Whinston and J. R. Green. 1995. *Microeconomic Theory*. Oxford University Press.

Mashal, R. and A. Zeevi. 2002. Beyond correlation: extreme co-movements between financial assets. Preprint, Columbia Business School.

Mashal, R., M. Naldi and A. Zeevi. 2003. On the dependence of equity and asset returns. *Risk* 16(10):83–87.

Massé, J. C. and R. Theodorescu. 1994. Halfplane trimming for bivariate distributions. *Journal of Multivariate Analysis* 48:188–202.

Matten, C. 2000. *Managing Bank Capital: Capital Allocation and Performance Measurement*, 2nd edn. Wiley.

Matthys, G. and J. Beirlant. 2000. Adaptive selection in tail index estimation. In *Extremes and Integrated Risk Management* (ed. P. Embrechts), pp. 37–49. London: Risk Books.

Medova, E. A. 2000. Measuring risk by extreme values. *Risk* 13(11):S20–S26.

Medova, E. A. and M. Kyriacou. 2000. Extreme values and the measurement of operational risk. I. *Operational Risk* 1(7):13–17.

Meng, X. L. and D. B. Rubin. 1993. Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* 80:267–278.

Meng, X. L. and D. van Dyk. 1997. The EM algorithm—an old folk song sung to a fast new tune (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 59:511–567.

Merino, S. and M. A. Nyfeler. 2004. Applying importance sampling for estimating coherent credit risk contributions. *Quantitative Finance* 4:199–207.

Merton, R. C. 1974. On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance* 29:449–470.

Mikosch, T. 2003. Modeling dependence and tails of financial time series. In *Extreme Values in Finance, Telecommunications, and the Environment* (ed. B. Finkenstaedt and H. Rootzén), pp. 185–286. London: Chapman & Hall.

——. 2004. *Non-Life Insurance Mathematics: An Introduction with Stochastic Processes*. Springer.

——. 2013. Modeling heavy-tailed time series. Lecture Notes, ETH Zurich.

Mikosch, T. and C. Stărică. 2000. Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Annals of Statistics* 28:1427–1451.

Mina, J. and J. Y. Xiao. 2001. Return to RiskMetrics: the evolution of a standard. Technical Report, RiskMetrics Group, New York.

Modigliani, F. and M. H. Miller. 1958. The cost of capital, corporation finance, and the theory of investment. *American Economic Review* 48:261–97.

Møller, T. and M. Steffensen. 2007. *Market Valuation Methods in Life and Pensions Insurance*. Cambridge University Press.

Moscadelli, M. 2004. The modelling of operational risk: experience with the analysis of the data, collected by the Basel Committee. Banca d'Italia, Temi di discussione del Servizio Studi 517 (July).

Mowbray, A. H. 1914. How extensive a payroll exposure is necessary to give a dependable pure premium. *Proceedings of the Casualty Actuarial Society* 1:24–30.

Müller, A. and D. Stoyan. 2002. *Comparison Methods for Stochastic Models and Risks*. Wiley.

Musiela, M. and M. Rutkowski. 1997. *Martingale Methods in Financial Modelling*. Springer.

Nagpal, K. and R. Bahar. 2001. Measuring default correlation. *Risk* 14(3):129–132.

Neftci, S. 2008. *Principles of Financial Engineering*, 2nd edn. Academic Press.

Nelsen, R. B. 2006. *An Introduction to Copulas*, 2nd edn. Springer.

Nelson, C. R. and A. F. Siegel. 1987. Parsimonious modeling of yield curves. *Journal of Business* 60:473–489.

Nelson, D. B. 1990. Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6:318–334.

——. 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59:347–370.

Nešlehová, J., P. Embrechts and V. Chavez-Demoulin. 2006. Infinite-mean models and the LDA for operational risk. *Journal of Operational Risk* 1(1):3–25.

Newey, W. K. and J. L. Powell. 1987. Asymmetric least squares estimation and testing. *Econometrica* 55:819–847.

Nickell, P., W. Perraudin and S. Varotto. 2000. Stability of ratings transitions. *Journal of Banking and Finance* 24:203–227.

Nolan, D. 1992. Asymptotics for multivariate trimming. *Stochastic Processes and Their Applications* 42:157–169.

Nolan, J. P. 2003. *Stable Distributions: Models for Heavy-Tailed Data*. Birkhauser.

Norberg, R. 1979. The credibility approach to experience rating. *Scandinavian Actuarial Journal* 4:181–221.

Norris, J. R. 1997. *Markov Chains*. Cambridge University Press.

Ogata, Y. 1988. Statistical models for earthquake occurrences and residuals analysis for point processes. *Journal of the American Statistical Association* 83:9–27.

Ong, M. K. 2003. *The Basel Handbook: A Guide for Financial Practitioners*. London: Risk Books.

Osband, K. H. 1985. Providing Incentives for Better Cost Forecasting. PhD thesis, University of California, Berkeley.

Ou, S. 2013. Annual default study: corporate default and recovery rates, 1920–2012. Technical Document, Moody's Investor Service.

Overbeck, L. and W. Schmidt. 2005. Modeling default dependence with threshold models. *Journal of Derivatives* 12:10–19.

Owen, R. and R. Rabinovitch. 1983. On the class of elliptical distributions and their application to the theory of portfolio choice. *Journal of Finance* 38:745–752.

Palisade. 1997. *Manual for @RISK*. Newfield, NY: Palisade Corporation.

Panjer, H. H. 1981. Recursive evaluation of a family of compound distributions. *ASTIN Bulletin* 12(1):22–26.

——. 2006. *Operational Risk: Modeling Analytics*. Wiley.

Panjer, H. H. and P. P. Boyle. 1998. *Financial Economics with Applications to Investment, Insurance and Pensions*. Schaumburg, IL: Actuarial Foundation.

Partrat, C. and J.-L. Besson. 2004. *Assurance Non-Vie: Modélisation, Simulation*. Paris: Economica.

Patrik, G., S. Bernegger and M. B. Rüegg. 1999. The use of risk adjusted capital to support business decision-making. *CAS Forum, Spring 1999*, Reinsurance Call Papers. Arlington, VA: Casualty Actuarial Society.

Patton, A. J. 2004. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Economics* 2(1):130–168.

——. 2006. Modelling asymmetric exchange rate dependence. *International Economic Review* 47:527–556.

Peeters, R. T. 2004. Financial Time and Volatility. PhD thesis, University of Amsterdam.

Pérignon, C. and D. R. Smith. 2010. The level and quality of value-at-risk disclosure by commercial banks. *Journal of Banking and Finance* 34(2):362–377.

Perraudin, W. (ed.). 2004. *Structured Credit Products: Pricing, Rating, Risk Management and Basel II*. London: Risk Books.

Petrov, V. V. 1995. *Limit Theorems of Probability Theory*. Oxford University Press.

Pézier, J. 2003a. A constructive review of the Basel proposals on operational risk. In *Operational Risk: Regulation, Analysis and Management* (ed. C. Alexander), chap. 4, pp. 49–73. Dorchester: FT–Prentice Hall.

——. 2003b. Operational risk management. In *Operational Risk: Regulation, Analysis and Management* (ed. C. Alexander), chap. 15, pp. 296–324. Dorchester: FT–Prentice Hall.

Pfeifer, D. 2004. VaR oder Expected Shortfall: welche Risikomaße sind für Solvency II geeignet? Preprint, University of Oldenburg.

Pfeifer, D. and J. Nešlehová. 2004. Modeling and generating dependent risk processes for IRM and DFA. *ASTIN Bulletin* 34(2):333–360.

Phillips, P. C. B. and P. Perron. 1988. Testing for unit roots in time series regression. *Biometrika* 75:335–346.

Pickands, J. 1971. The two-dimensional Poisson process and extremal processes. *Journal of Applied Probability* 8:745–756.

——. 1975. Statistical inference using extreme order statistics. *Annals of Statistics* 3:119–131.

——. 1981. Multivariate extreme value distribution. In *Proceedings of the 43rd Session of the International Statistics Institute, Buenos Aires*, vol. 2, pp. 859–878. The Hague: International Statistical Institute.

Prates, M. M. 2013. Why prudential regulation will fail to prevent financial crises. A legal approach. Working Paper 335 (Technical Report), Banco Central do Brasil.

Prause, K. 1999. The Generalized Hyperbolic Model: Estimation, Financial Derivatives and Risk Measures. PhD thesis, Institut für Mathematische Statistik, Albert-Ludwigs-Universität Freiburg.

Prentice, M. J. 1978. On invariant tests of uniformity for directions and orientations. *Annals of Statistics* 6(1):169–176.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery. 1992. *Numerical Recipes in C*. Cambridge University Press.

Priestley, M. B. 1981. *Spectral Analysis and Time Series*. New York: Academic Press.

Protassov, R. S. 2004. EM-based maximum likelihood parameter estimation of multivariate generalized hyperbolic distributions with fixed λ. *Statistics and Computing* 14:67–77.

Protter, P. 2005. *Stochastic Integration and Differential Equations: A New Approach*, 2nd edn. Springer.

Puccetti, G. 2013. Sharp bounds on the expected shortfall for a sum of dependent random variables. *Statistics and Probability Letters* 83(4):1227–1232.

Puccetti, G. and L. Rüschendorf. 2012. Computation of sharp bounds on the distribution of a function of dependent risks. *Journal of Computational and Applied Mathematics* 236(7):1833–1840.

Puccetti, G. and R. Wang. 2014. General extremal dependence concepts. Preprint. (Available from SSRN: 2436392.)

——. 2015. Detecting complete and joint mixability. *Journal of Computational and Applied Mathematics* 280:174–187.

Rebonato, R. 2007. *Plight of the Fortune Tellers: Why We Need to Manage Financial Risk Differently*. Princeton University Press.

——. 2010. *Coherent Stress Testing: A Bayesian Approach to the Analysis of Financial Stress*. Wiley.

Reinhart, C. M. and K. S. Rogoff. 2009. *This Time Is Different: Eight Centuries of Financial Folly*. Princeton University Press.

Reiss, R.-D. and M. Thomas. 1997. *Statistical Analysis of Extreme Values*. Basel: Birkhäuser.

Remillard, B. 2013. *Statistical Methods for Financial Engineering*. Chapman & Hall.

Resnick, S. I. 1992. *Adventures in Stochastic Processes*. Boston, MA: Birkhäuser.

Resnick, S. I. 1997. Discussion of the Danish data on large fire insurance losses. *ASTIN Bulletin* 27(1):139–151.

——. 2007. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer.

——. 2008. *Extreme Values, Regular Variation and Point Processes*. Springer.

Resnick, S. I. and C. Stărică. 1995. Consistency of Hill's estimator for dependent data. *Journal of Applied Probability* 32:239–167.

——. 1996. Tail index estimation for dependent data. Technical Report, School of ORIE, Cornell University.

——. 1997. Smoothing the Hill estimator. *Advances in Applied Probability* 29:291–293.

Rice, J. A. 1995. *Mathematical Statistics and Data Analysis*, 2nd edn. Belmont, CA: Duxbury Press.

Riedel, F. 2004. Dynamic coherent risk measures. *Stochastic Processes and Their Applications* 112:185–200.

Ripley, B. D. 1987. *Stochastic Simulation*. Wiley.

RiskMetrics Group. 1997. CreditMetrics technical document. (Available from www.msci.com/resources/technical_documentation/CMTD1.pdf.)

Robbins, H. 1955. An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 157–163. University of California Press.

Robbins, H. 1964. The empirical Bayes approach to statistical problems. *Annals of Mathematical Statistics* 35:1–20.

Robert, C. P. and G. Casella. 1999. *Monte Carlo Statistical Methods*. Springer.

Roberts, R. 2012. Did anyone learn anything from the Equitable Life? Lessons and learning from financial crises. Institute of Contemporary British History (ICBH), King's College London.

Rochet, J.-C. 2008. *Why Are There So Many Banking Crises? The Politics and Policy of Bank Regulation*. Princeton University Press.

Rockafellar, R. T. 1970. *Convex Analysis*. Princeton University Press.

Rockafellar, R. T. and S. Uryasev. 2000. Optimization of conditional value-at-risk. *Journal of Risk* 2:21–42.

——. 2002. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance* 26(7):1443–1471.

Rogers, L. C. G. and D. Williams. 1994. *Diffusions, Markov Processes, and Martingales. Volume One: Foundations*, 2nd edn. Wiley.

Rogge, E. and P. J. Schönbucher. 2003. Modeling dynamic portfolio credit risk. Preprint, Imperial College and ETH Zürich.

Rolski, T., H. Schmidli, V. Schmidt and J. L. Teugels. 1999. *Stochastic Processes for Insurance and Finance*. Wiley.

Rootzén, H. and N. Tajvidi. 1997. Extreme value statistics and wind storm losses: a case study. *Scandinavian Actuarial Journal* 1997(1):70–94.

Rosen, D. and D. Saunders. 2009. Valuing CDOs of bespoke portfolios with implied multi-factor models. *Journal of Credit Risk* 5(3):3–36.

——. 2010. Economic capital. In *Encyclopedia of Quantitative Finance* (ed. R. Cont). Wiley.

Rosenberg, J. V. and T. Schuermann. 2006. A general approach to integrated risk management with skewed, fat-tailed risks. *Journal of Financial Economics* 79(3):569–614.

Rousseeuw, P. J. and G. Molenberghs. 1993. Transformation of non positive semidefinite correlation matrices. *Communications in Statistics: Theory and Methods* 22(4):965–984.

Rousseeuw, P. J. and I. Ruts. 1999. The depth function of a population distribution. *Metrika* 49:213–244.

Rouvinez, C. 1997. Going Greek with VaR. *Risk* 10(2):57–65.

Rüschendorf, L. 1982. Random variables with maximum sums. *Advances in Applied Probability* 14:623–632.

——. 2009. On the distributional transform, Sklar's Theorem, and the empirical copula process. *Journal of Statistical Planning and Inference* 139:3921–3927.

——. 2013. *Mathematical Risk Analysis*. Springer.

Rüschendorf, L. and L. Uckelmann. 2002. Variance minimization and random variables with constant sum. In *Distributions with Given Marginals and Statistical Modelling* (ed. C. M. Cuadras, J. Fortiana and J. A. Rodríguez-Lallena), pp. 211–222. Springer.

Ryan, S. G. 2008. Fair-value accounting: understanding the issues raised by the credit crunch. White Paper (Technical Report), Council of Institutional Investors.

Sandström, A. 2006. *Solvency: Models, Assessment and Regulation*. Boca Raton, FL: Chapman & Hall.

——. 2011. *Handbook of Solvency for Actuaries and Risk Managers: Theory and Practice*. Boca Raton, FL: Chapman & Hall.

Schervish, M. J. 1995. *Theory of Statistics*. Springer.

Schlather, M. and J. A. Tawn. 2002. Inequalities for the extremal coefficients of multivariate extreme value distributions. *Extremes* 5:87–102.

Schmeidler, D. 1986. Integral representation without additivity. *Proceedings of the American Mathematical Society* 97:255–261.

Schmidt, R. 2002. Tail dependence for elliptically contoured distributions. *Mathematical Methods of Operations Research* 55:301–327.

Schmock, U. 1999. Estimating the value of the WINCAT coupons of the Winterthur insurance convertible bond: a study of the model risk. *ASTIN Bulletin* 29(1):101–163.

Schoenberg, I. J. 1938. Metric spaces and completely monotone functions. *Annals of Mathematics* 39:811–841.

Schönbucher, P. J. 2003. *Credit Derivatives Pricing Models*. Wiley.

——. 2004. Information-driven default contagion. Preprint, Department of Mathematics, ETH Zurich.

——. 2005. Taken to the limit: simple and not-so-simple loan loss distributions. In *The Best of Wilmott 1: Incorporating the Quantitative Finance Review* (ed. P. Wilmott), vol. 1, chap. 11, pp. 143–160. Wiley.

Schönbucher, P. J. and D. Schubert. 2001. Copula-dependent default risk in intensity models. Preprint, ETH Zurich and Universität Bonn.

Schoutens, W. 2003. *Lévy Processes in Finance: Pricing Financial Derivatives*. Wiley.

Schweizer, B. and A. Sklar. 1983. *Probabilistic Metric Spaces*. New York: North-Holland/Elsevier.

Schweizer, B. and E. F. Wolff. 1981. On nonparametric measures of dependence for random variables. *Annals of Statistics* 9:879–885.

Schweizer, M. 2001. A guided tour through quadratic hedging approaches. In *Option Pricing, Interest Rates and Risk Management* (ed. E. Jouini, J. Cvitanic and M. Musiela), pp. 538–574. Cambridge University Press.

Seal, H. L. 1969. *Stochastic Theory of a Risk Business*. Wiley.

Seber, G. A. F. 1984. *Multivariate Observations*. Wiley.

Seneta, E. 1976. *Regularly Varying Functions*. Lecture Notes in Mathematics. Springer.

Serfling, R. J. 1980. *Approximation Theorems of Mathematical Statistics*. Wiley.

Shaffer, S. 2011. Evaulating the impact of fair value accounting on financial institutions: implications for accounting standards setting and bank supervision. Working Paper QAU 12-01 (Technical Report), Federal Reserve Bank of Boston.

Sharpe, W. F., G. J. Alexander and J. V. Bailey. 1999. *Investments*, 6th edn. Englewood Cliffs, NJ: Prentice Hall.

Shea, G. A. 1983. Hoeffding's lemma. In *Encyclopaedia of Statistical Science* (ed. S. Kotz, N. L. Johnson and C. B. Read), vol. 3, pp. 648–649. Wiley.

Shephard, N. 1996. Statistical aspects of ARCH and stochastic volatility. In *Time Series Models in Econometrics, Finance and Other Fields* (ed. D. R. Cox, D. V. Hinkley and O. E. Barndorff-Nielsen), pp. 1–67. London: Chapman & Hall.

Shevchenko, P. V. 2011. *Modelling Operational Risk Using Bayesian Inference*. Springer.

Shiller, R. J. 2000. *Irrational Exuberance*. Princeton University Press.

——. 2003. *The New Financial Order: Risk in the Twenty-First Century*. Princeton University Press.

——. 2008. *The Subprime Solution: How Today's Global Crisis Happened and What to Do About It*. Princeton University Press.

——. 2012. *Finance and the Good Society*. Princeton University Press.

Shimpi, P. K. 2001. *Integrating Corporate Risk Management*. New York: Texere.

Shin, H. S. 2010. *Risk and Liquidity*. Oxford University Press.

Shreve, S. E. 2004a. *Stochastic Calculus for Finance. I. The Binomial Asset Pricing Model*. Springer.

——. 2004b. *Stochastic Calculus for Finance. II. Continuous-Time Models*. Springer.

——. 2008. Don't blame the quants. (Available from www.forbes.com/2008/10/07/securities-quants-models-oped-cx_ss_1008shreve.html.)

——. 2009. Did faulty mathematical models cause the financial fiasco? Blame it on Gaussian copula. *Analytics*, Spring. (Available at www.analytics-magazine.com.)

Sibuya, M. 1960. Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics* 11:195–210.

Sidenius, J., V. Piterbarg and L. Andersen. 2008. A new framework for dynamic credit portfolio loss modelling. *International Journal of Theoretical and Applied Finance* 11:163–197.

Siegel, A. F. and C. R. Nelson. 1988. Long-term behaviour of yield curves. *Journal of Financial and Quantitative Analysis* 23:105–110.

Sklar, A. 1959. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* 8:229–231.

Small, N. J. H. 1978. Plotting squared radii. *Biometrika* 65:657–658.

Smith, P. J. 1977. A nonparametric test for bivariate circular symmetry based on the empirical CDF. *Communications in Statistics: Theory and Methods* A 6(3):209–220.

Smith, R. L. 1985. Maximum likelihood estimation in a class of nonregular cases. *Biometrika* 72:67–92.

——. 1987. Estimating tails of probability distributions. *Annals of Statistics* 15:1174–1207.

——. 1989. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science* 4:367–393.

——. 1994. Multivariate threshold methods. In *Extreme Value Theory and Applications* (ed. J. Galambos), pp. 225–248. Boston, MA: Kluwer Academic.

Smith, R. L. and D. Goodman. 2000. Bayesian risk analysis. In *Extremes and Integrated Risk Management* (ed. P. Embrechts), pp. 235–267. London: Risk Books.

Smith, R. L. and T. S. Shively. 1995. Point process approach to modeling trends in tropospheric ozone based on exceedances of a high threshold. *Atmospheric Environment* 29:3489–3499.

Smith, R. L. and I. Weissman. 1994. Estimating the extremal index. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 56:515–528.

Smith, R. L., J. A. Tawn and S. G. Coles. 1997. Markov chain models for threshold exceedances. *Biometrika* 84:249–268.

Soros, G. 2009. One way to stop bear raids: credit default swaps need much stricter regulation. *Wall Street Journal*, 24 March.

Steinherr, A. 1998. *Derivatives: The Wild Beast of Finance*. Wiley.

Straumann, D. 2005. *Estimation in Conditionally Heteroscedastic Time Series Models*. Springer.

Straumann, D. and T. Mikosch. 2006. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *Annals of Statistics* 34(5):2449–2495.

Stuart, A., J. K. Ord and S. Arnold. 1999. Classical inference and the linear model. In *Kendall's Advanced Theory of Statistics*, 6th edn, vol. 2A. Oxford University Press.

Studer, G. 1997. Maximum Loss for Measurement of Market Risk. PhD thesis, ETH Zurich.

——. 1999. Market risk computation for nonlinear portfolios. *Journal of Risk* 1(4):33–53.

——. 2001. Stochastic Taylor Expansions and Saddlepoint Approximations for Risk Management. PhD thesis, ETH Zurich.

Stulz, R. M. 1996. Rethinking risk management. *Journal of Applied Corporate Finance* 9(3):8–24.

——. 2002. *Risk Management and Derivatives*. Mason, OH: South-Western.

——. 2010. Credit default swaps and the credit crisis. *Journal of Economic Perspectives* 24(1):73–92.

Sun, Z., D. Munves and D. Hamilton. 2012. Public firm expected default frequency (EDF) credit measures: methodology, performance and model extensions. Technical Document, Moody's Analytics.

Sundt, B. 1999. On multivariate Panjer recursions. *ASTIN Bulletin* 29(1):29–45.

——. 2000. On multivariate Vernic recursions. *ASTIN Bulletin* 30(1):111–122.

Sundt, B. and W. S. Jewell. 1982. Further results on recursive evaluation of compound distributions. *ASTIN Bulletin* 12(1):27–39.

Sundt, B. and R. Vernic. 2009. *Recursions for Convolutions and Compound Distributions with Insurance Applications*. Springer.

Svensson, L. E. O. 1994. Estimating and interpreting forward interest rates: Sweden 1992–1994. IMF Working Paper, Technical Report 94/114.

Takahashi, R. 1994. Domains of attraction of multivariate extreme value distributions. *Journal of Research of NIST* 99:551–554.

Taleb, N. N. 2001. *Fooled by Randomness: The Hidden Role of Chance in the Markets and Life*. New York: Texere.

——. 2007. *The Black Swan: The Impact of the Highly Improbable*. Random House.

Tarullo, D. K. 2008. Banking on Basel: the future of international financial regulation. Peterson Institute for International Economics, Washington, DC.

Tasche, D. 1999. Risk contributions and performance measurement. Preprint, TU-München.

——. 2000. Conditional expectation as a quantile derivative. Preprint, TU-München. (Available from arXiv math/0104190.)

——. 2002. Expected shortfall and beyond. *Journal of Banking and Finance* 26:1519–1533.

——. 2008. Capital allocation to business units and sub-portfolios: the Euler principle. In *Pillar II in the New Basel Accord: The Challenge of Economic Capital* (ed. A. Resti), pp. 423–453. London: Risk Books.

Tawn, J. A. 1988. Bivariate extreme value theory: models and estimation. *Biometrika* 75(3):397–415.

——. 1990. Modelling multivariate extreme value distributions. *Biometrika* 77(2):245–253.

Taylor, S. J. 2008. *Modelling Financial Time Series*, 2nd edn. World Scientific.

Teugels, J. L. and B. Sundt (eds). 2004. *Encyclopedia of Actuarial Science*. Wiley.

Thomas, L. C. 2009. *Consumer Credit Models: Pricing, Profit and Portfolios*. Oxford University Press.

Thoyts, R. 2010. *Insurance Theory and Practice*. Routledge.

Tiago de Oliveira, J. 1958. Extremal distributions. *Revista da Faculdade de Ciências de Lisboa* A 7:215–227.

Tiit, E. 1996. Mixtures of multivariate quasi-extremal distributions having given marginals. In *Distributions with Fixed Marginals and Related Topics* (ed. L. Rüschendorff, B. Schweizer and M. D. Taylor), pp. 337–357. Hayward, CA: Institute of Mathematical Statistics.

Tsay, R. S. 2002. *Analysis of Financial Time Series*. Wiley.

Tsay, R. S. and G. C. Tiao. 1984. Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of the American Statistical Association* 79:84–96.

Tse, Y. and A. Tsui. 2002. A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics* 20:351–362.

Tsukahara, H. 2009. One-parameter families of distortion risk measures. *Mathematical Finance* 19(4):691–705.

Tuckman, B. and A. Serrat. 2011. *Fixed Income Securities: Tools for Today's Markets*, 3rd edn. Wiley.

Tukey, J. W. 1975. Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians, Vancouver*, vol. 2, pp. 523–531. Montreal: Canadian Mathematical Society.

Tyler, D. E. 1983. Robustness and efficiency properties of scatter matrices. *Biometrika* 70(2):411–420.

——. 1987. A distribution-free M-estimator of multivariate scatter. *Annals of Statistics* 15(1):234–251.

Upper, C. 2011. Simulation methods to assess the danger of contagion in financial networks. *Journal of Financial Stability* 7:111–125.

Varnell, E. M. 2011. Economic scenario generators and Solvency II. *British Actuarial Journal* 16(1):121–159.

Vasicek, O. 1997. The loan loss distribution. Preprint, KMV Corporation.

Venter, J. H. and P. J. de Jongh. 2002. Risk estimation using the normal inverse Gaussian distribution. *Journal of Risk* 4(2):1–24.

Vrins, F. D. 2009. Double-t copula pricing of structured credit products: practical aspects of a trustworthy implementation. *Journal of Credit Risk* 5(3):91–109.

Vyncke, D. 2004. Comonotonicity. In *Encyclopedia of Actuarial Science*. Wiley.

Wang, B. and R. Wang. 2011. The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis* 102(10):1344–1360.

——. 2014. Extreme negative dependence and risk aggregation. Preprint, University of Waterloo.

Wang, R., L. Peng and J. Yang. 2013. Bounds for the sum of dependent risks and worst value-at-risk with monotone marginal densities. *Finance and Stochastics* 17(2):395–417.

Wang, S. 1996. Premium calculation by transforming the layer premium density. *ASTIN Bulletin* 26:71–92.

Wang, S. and J. Dhaene. 1998. Comonoticity, correlation order and premium principles. *Insurance: Mathematics and Economics* 22:235–242.

Weber, S. 2006. Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance* 16(2):419–441.

Weitzman, M. L. 2011. Fat-tailed uncertainty in the economics of catastrophic climate change. *Review of Environmental Economics and Policy* 5(2):275–292.

White, H. 1981. Maximum likelihood estimation of misspecified models. *Econometrica* 50:1–25.

Whitney, A. W. 1918. The theory of experience rating. *Proceedings of the Casualty Actuarial Society* 4:274–292.

Wilcox, R. 1997. *Introduction to Robust Estimation and Hypothesis Testing*. New York: Academic Press.

Williams, D. 1991. *Probability with Martingales*. Cambridge University Press.

Williamson, R. C. and T. Downs. 1990. Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning* 4:89–158.

Wilmot, P. 2000. *Paul Wilmott on Quantitative Finance*. Wiley.

Wilson, T. 1997a. Portfolio credit risk. I. *Risk* 10(9):111–117.

——. 1997b. Portfolio credit risk. II. *Risk* 10(10):56–61.

Wüthrich, M. V. 2011. An academic view on the illiquidity premium and market-consistent valuation. *European Actuarial Journal* 1(1):93–105.

——. 2013. Non-life insurance: mathematics and statistics. Preprint. (Available from SSRN: 2319328.)

Wüthrich, M. V. and M. Merz. 2013. *Financial Modeling, Actuarial Valuation and Solvency in Insurance*. Springer.

Wüthrich, M. V., H. Bühlmann and H. J. Furrer. 2010. *Market-Consistent Actuarial Valuation*, 2nd edn. Springer.

Yaari, M. E. 1987. The dual theory of choice under risk. *Econometrica* 55:95–115.

Yang, L., W. Härdle and J. P. Nielsen. 1999. Nonparametric autoregression with multiplicative volatility and additive mean. *Journal of Time Series Analysis* 20:579–604.

Yu, F. 2007. Correlated defaults in intensity-based models. *Mathematical Finance* 17:155–173.

Zhou, C. 2001. The term structure of credit spreads with jump risk. *Journal of Banking and Finance* 25(11):2015–2040.

Zhou, X. Y. 2010. Mathematicalising behavioural finance. In *Proceedings of the International Congress of Mathematicians, Hyderabad, India*, vol. 4, pp. 3185–3209. World Scientific.

Ziegel, J. F. 2015. Coherence and elicitability. *Mathematical Finance*, doi:10.1111/mafi.12080.

Zivot, E. and J. Wang. 2003. *Modeling Financial Time Series with S-PLUS*. Springer.

Index