ASQ01: Lecture Summary

Overview

This document provides a summary of the ASQ-01 lecture on the basics of time series.
The lecture covers the following topics:

- What is a time series?
- Difference between technical analysis and time series analysis
- What is stationarity?
- ACF and PACF
- Transformation of prices to returns
- How to visualize the data?
- Realized volatility
- How to detect outliers?
- Features of asset returns observed empirically.

What is a timeseries?

Simply put, data for any variable collected over a period is called a time-series.
For example, a person's daily weight for the last 'n' days will be a time series of 'n' observations. The stock price fetched every minute for the previous 'm' minutes will be a timeseries of 'm' elements.

What is a time series analysis?

Time series analysis uses the time series data and looks for patterns and inconsistencies. The most important question we ask and seek the answer to is:
What will be the value of the variable in the following period?

How is time series analysis different from the technical analysis?

Technical analysis is an investment style that uses charts and averages to identify patterns in asset prices. In comparison, time series analysis is a much broader approach. It uses statistical theory to formulate and validate the hypothesis proposed to forecast.

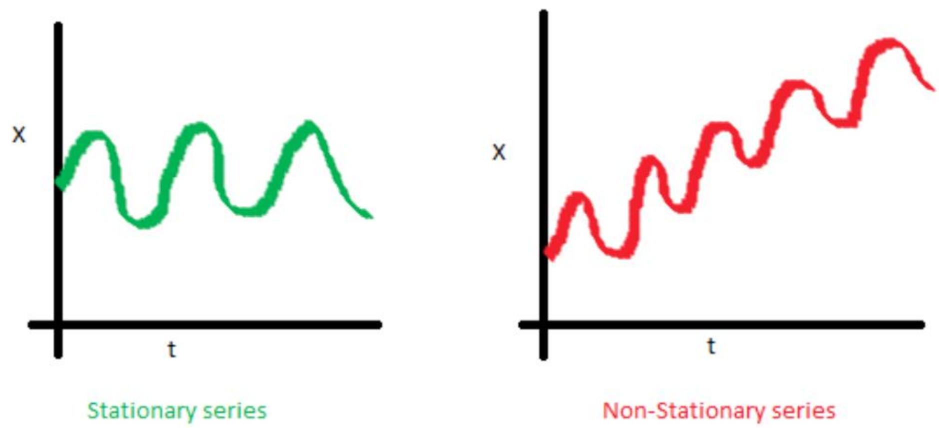
What is stationarity?

A time series is considered stationary if its properties (mean, variance) don't change over time.
Empirically it has been observed that most time series in any field are non-stationary. However, you can build synthetic stationary timeseries with two or more of the non-stationary time series data.

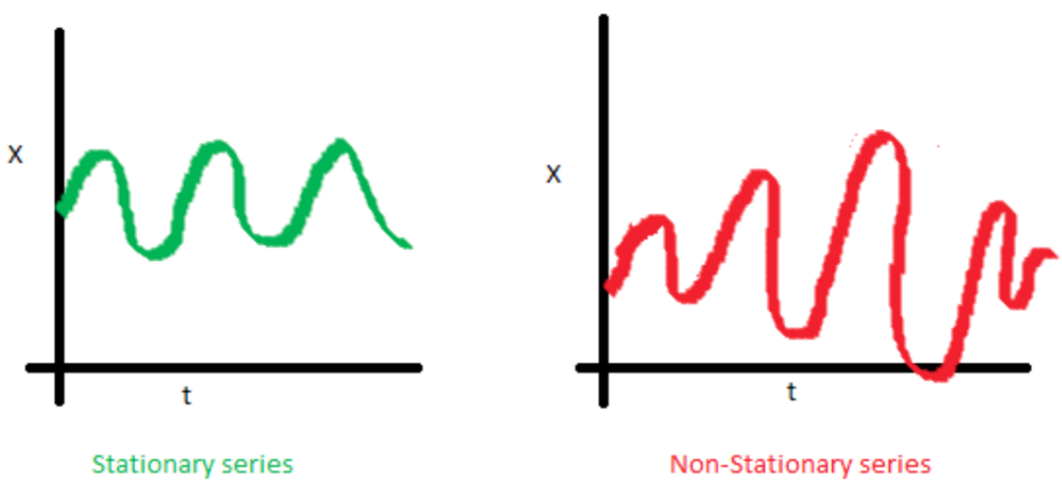
Analysis on a stationary series is easier to deal with because of its stable statistical properties. So, we always look for ways to transform our original series into a stationary one.

“The essence of mathematics is not to make simple things complicated, but to make complicated things simple.”Stan Gudder”

The following snapshot shows examples of a stationary and a non-stationary series.



Source



Source

ACF (Auto Correlation Function)

One of the properties of time series data is persistence in observations. This is also intuitive; we don’t expect a person’s weight to suddenly jump from one day to another or a stock price to move significantly (under normal conditions) within a few hours or days. Now, intuition helps us understand things, but we need to quantify them to make decisions. Therefore, the Auto-Correlation Function (ACF) is a metric for quantifying this persistence behaviour of time-series data.

Correlation between two variables helps us identify a linear relationship. In contrast, a correlation with a variable and its lagged version in a time series data is called serial correlation or auto-correlation. ACF is the plot of this correlation.

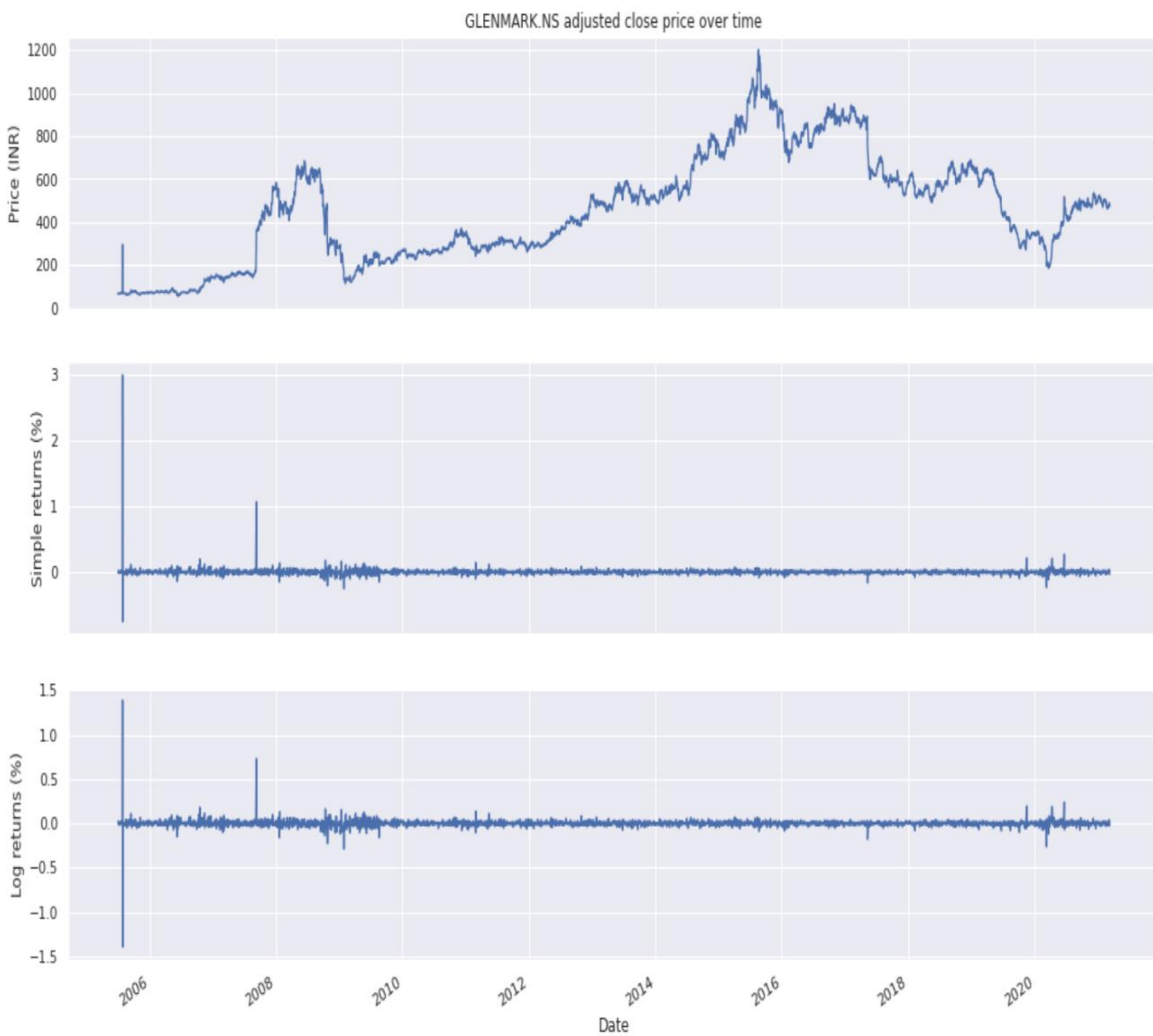
PACF (Partial Auto Correlation Function)

It only considers the direct relationship between the variable and its lagged version. It is a special case of the ACF.

A simple technique for making a series stationary

Stock prices are usually non-stationary. However, using returns instead of the raw prices is quite an effective technique in making the series stationary.

This snapshot shows the time series of price, simple returns and log returns for GLENMARK.



Realized volatility

It is a risk measure used to capture the change in price movement over time. One way to calculate monthly realized volatility is using the Barndorff-Nielsen and Shepherd Method.

Here are the steps to calculate it:

1. Calculate the daily log returns

$$R_t \quad = log(P_t) \; - \; log(P_{t-1})$$

2. Calculate the monthly realized variance by summing the squared returns for that month's `N` trading days.

$$\sum_{t=1}^{N} \quad {r^2}_t$$

3. The monthly realized volatility is the square root of the above expression,

4. We annualise the value by multiplying the monthly realized volatility by $\sqrt{12}$

The following snapshot shows the time series of price, log returns and monthly realized volatility for GLENMARK.

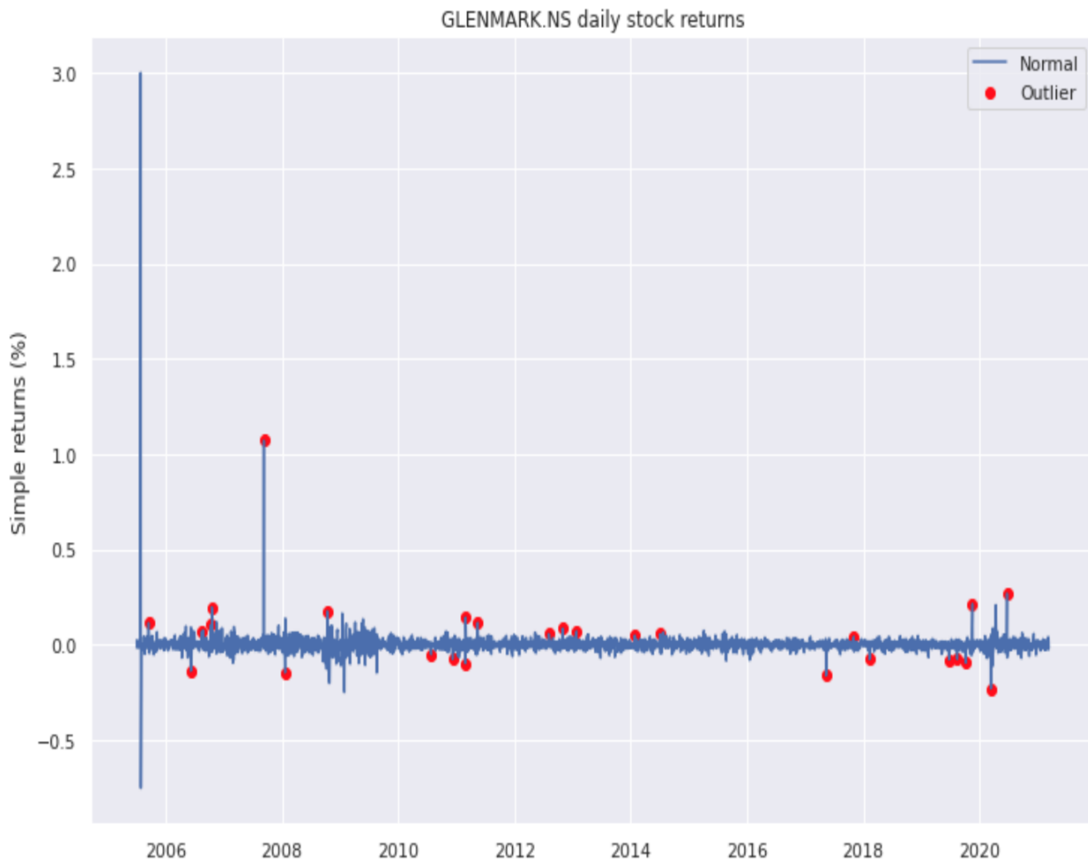
As you might have noticed, there are some unusually large values; they are called outliers.
In the next section, you'll see a method of detecting outliers.

How to detect outliers?

Outliers can be present for various reasons, such as a human error in filling the data, a black swan event, missing data filled with random values, etc. It is crucial to detect and remove/replace these outliers to analyse and make realistic predictions.
One criterion to detect outliers is using standard deviation(volatility).

For example - Any value 3 SD away from the mean can be considered an outlier.

The following snapshot shows the time series of simple returns of GLENMARK with outliers highlighted as red dots.



Features of asset returns

1. Non-normal distribution of returns

A standard assumption in finance models (like the CAPM, the Black-Scholes option pricing model) is normally distributed returns.

However, the following has been observed:

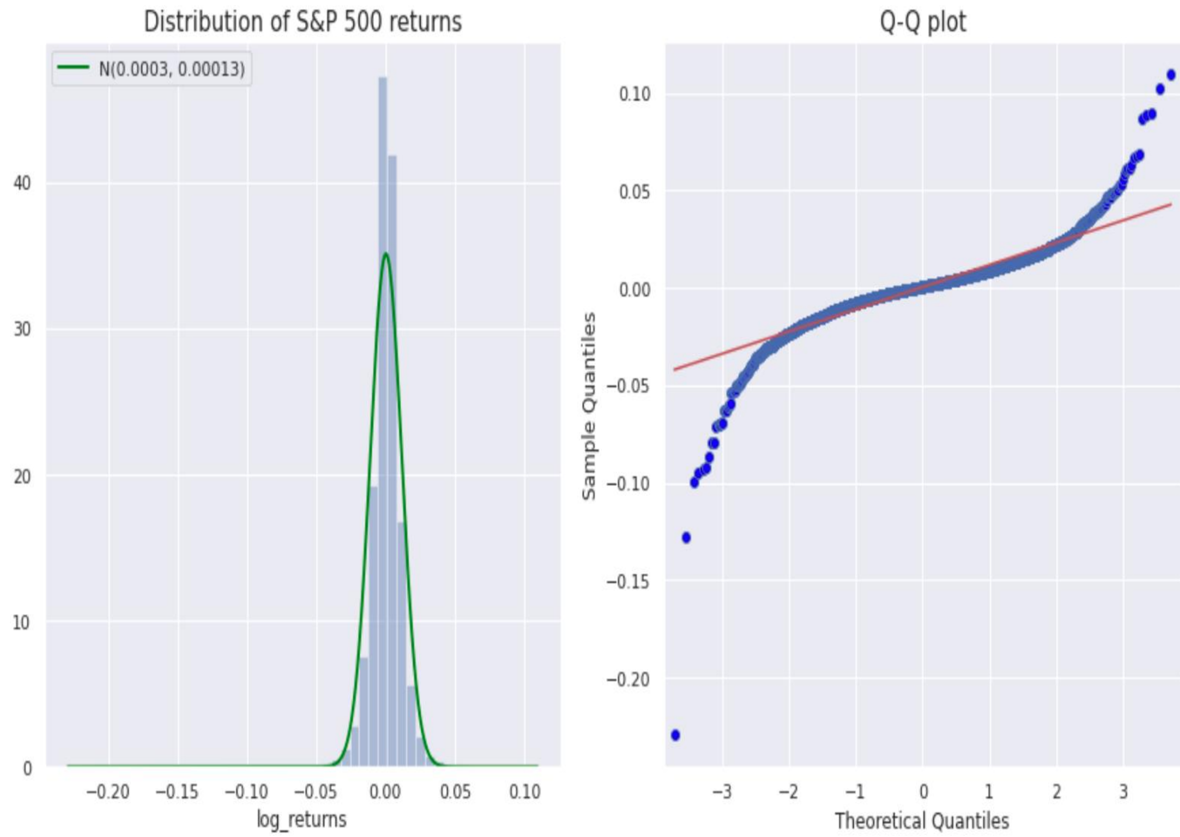
Left skewed: Gains and loss distributions are not symmetrical.

- We see longer left tails compared to the right tails
- The large negative returns are higher in magnitude compared to the large positive returns.

Excess kurtosis: The distribution is fat at the tails and higher than normal peaks

- Large (and small) returns occur more frequently than the distributional assumption of normality expects.

The following snapshot shows the normal distribution and Q-Q plot (right) as compared to the distribution of S&P 500 returns (left)



Observations

Distribution of S&P 500 returns

- There is a visible difference in the shape of the returns histogram and the Gaussian (normal) distribution curve.
- The peak is higher in the histogram than the normal curve.
- The left tail of the distribution is longer.

Test for normality

Method-1: Q-Q plot

- Q-Q plots are used to compare empirical data to theoretical distributions. They help find deviations at the tails.
- Here, we compare the distribution of the observed returns to a normal distribution.
- If we find that the dots are more or less on the red line, then the data (in this case returns) is normally distributed.
- There is drift at both ends of the tails. This means that we have fatter tails violating the normality assumption.

- The drift size is higher on the left side of the plot than on the right side. This means that we observe large drops in returns but not equally large growths in returns.

Method 2 - Jarque-Bera test

Jarque-Bera test is one of the widely used methods for testing normality.

The following snapshot shows the summary statistics of an S&P 500 log-returns timeseries with the Jarque-Bera statistic.

```
------------------- Summary Statistics --------------------
Range of dates: 1980-01-02 to 2020-12-30
Number of observations: 10339
Mean: 0.0003
Median: 0.0006
Min: -0.2290
Max: 0.1096
Standard Deviation: 0.0114
Skewness: -1.1531
Kurtosis: 25.8875
Jarque-Bera statistic: 290698.57 with p-value: 0.00
-----------------------------------------------------------
```

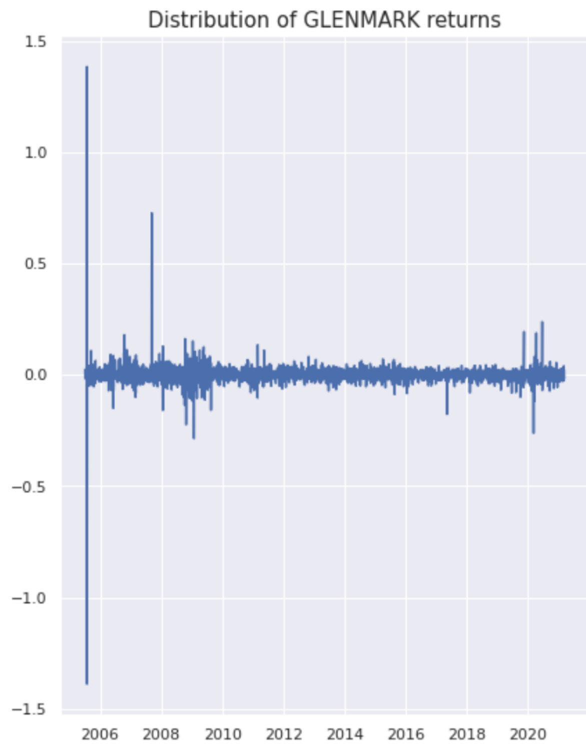
Observations

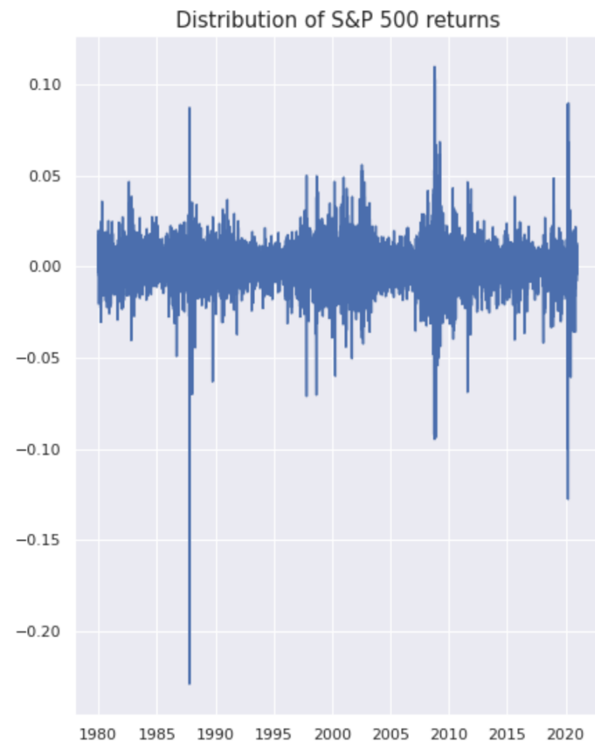
- The mean is less than the median that is seen in negatively skewed distributions.
- This is confirmed by the coefficient of skewness (which is negative).
- Excess kurtosis is seen (anything above 0 is deemed as excess kurtosis. Normal distribution has a kurtosis of 0 when measured using the `pandas kurtosis` method).
- The p-value of the Jarque-Bera test shows the non-normality of the data.

2. Volatility clustering

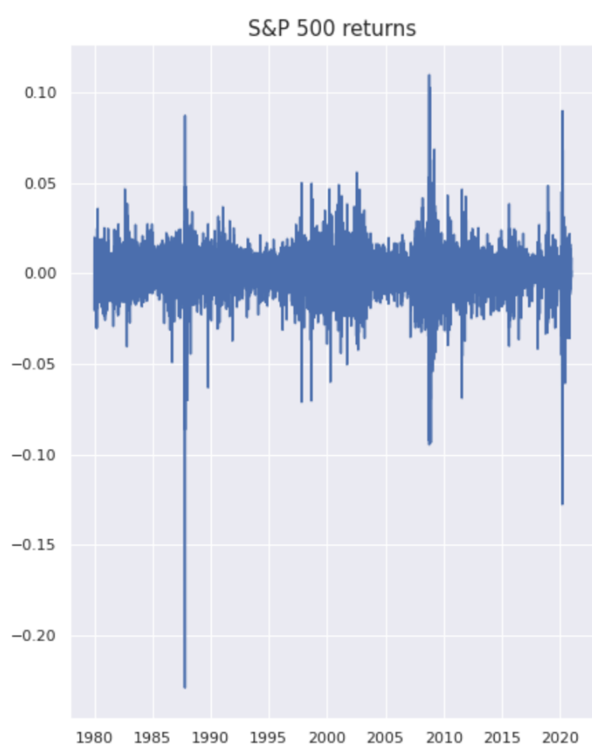
Simply put, volatility clustering means that large changes tend to follow large changes, and hence the volatility is "clustered".
Here's an example:

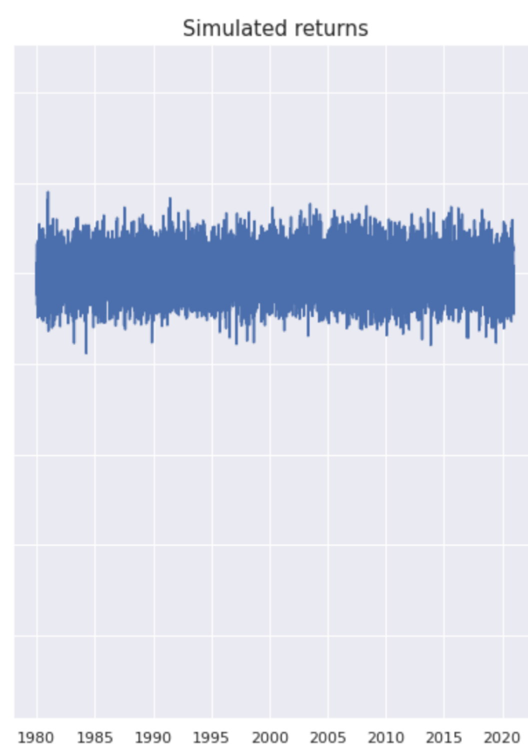
The following snapshot shows the distribution of returns for GLENMARK (left) and S&P500 (right)





The following snapshot compares the S&P500 returns (left) to the simulated returns of a normal distribution. (right)





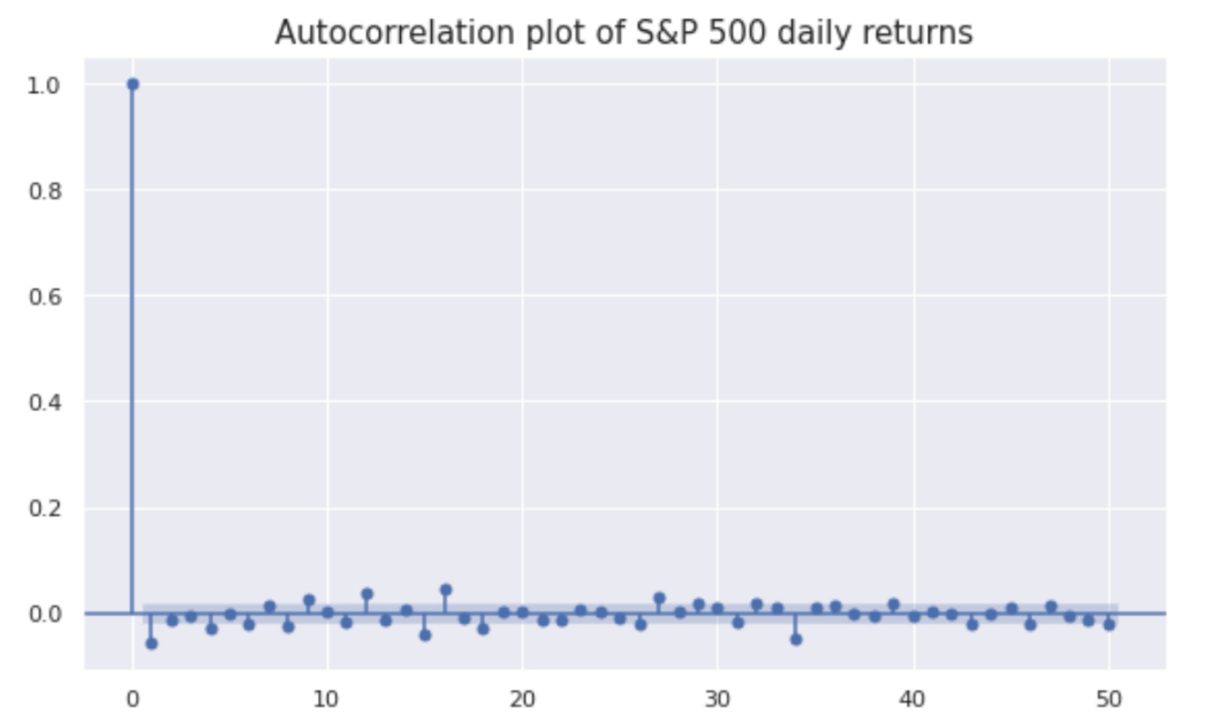
Observations

- For Glenmark, between 2008 and 2010, there's a higher swing of positive and negative returns. Notice how it's quite different between 2016 and 2018.
- Similar waves of high and low volatility periods seen in S&P returns

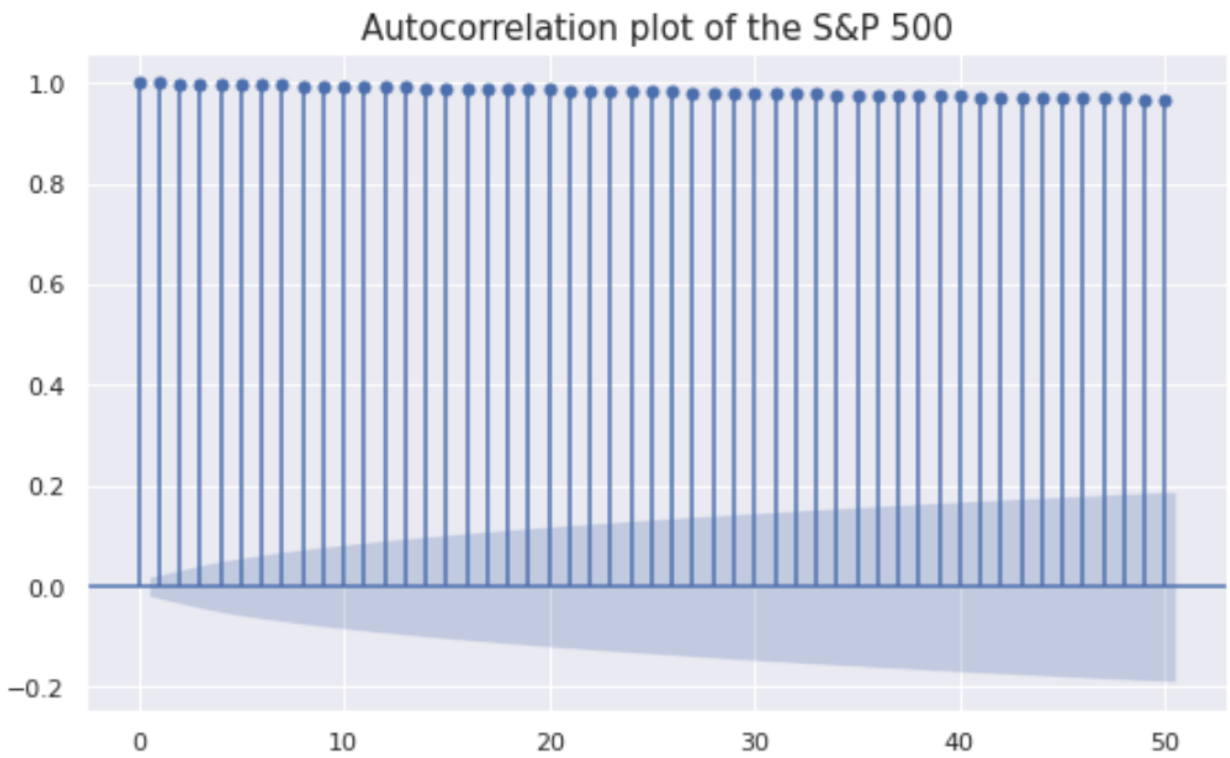
3. No auto-correlation

Auto-correlation is the similarity in values of a time series with a delayed version of itself.

The following snapshot shows the autocorrelation of S&P500 daily returns



The following snapshot shows the autocorrelation of S&P500 prices.

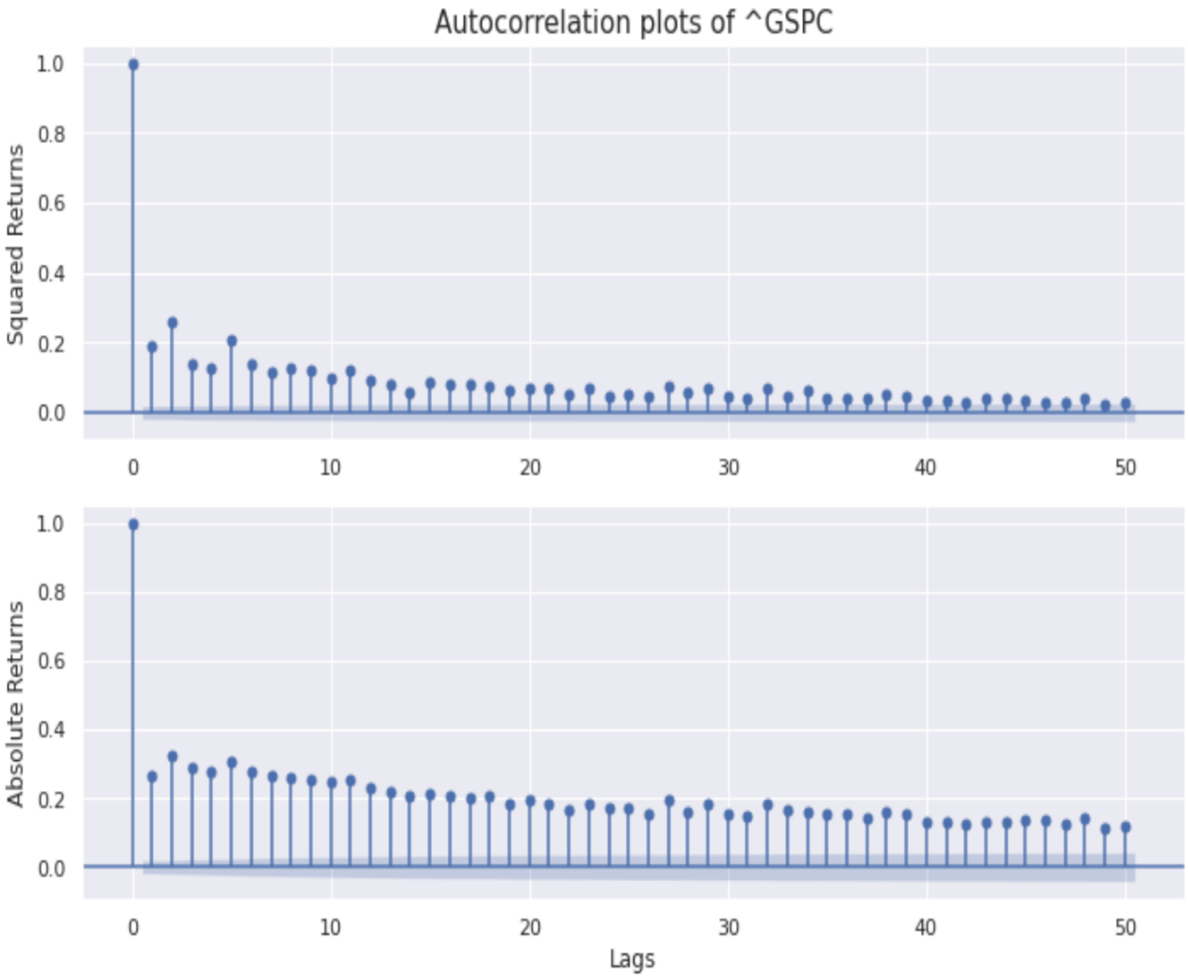


Observations

- We do not look at autocorrelation at lag 0.
- A few values that lie outside the blue confidence interval are statistically significant.
- Visually, it appears that there is very little autocorrelation in the returns.

4. Slow decay of autocorrelation in absolute terms

The following snapshot shows the ACF plot for absolute and squared returns of S&P 500.



Observations

- There is a slow and uneven decay of the ACF plot.
- The ACF of the squared returns decreases faster than that of the absolute returns.
- There is significant autocorrelation, as seen in the literature.

5. Leverage effect

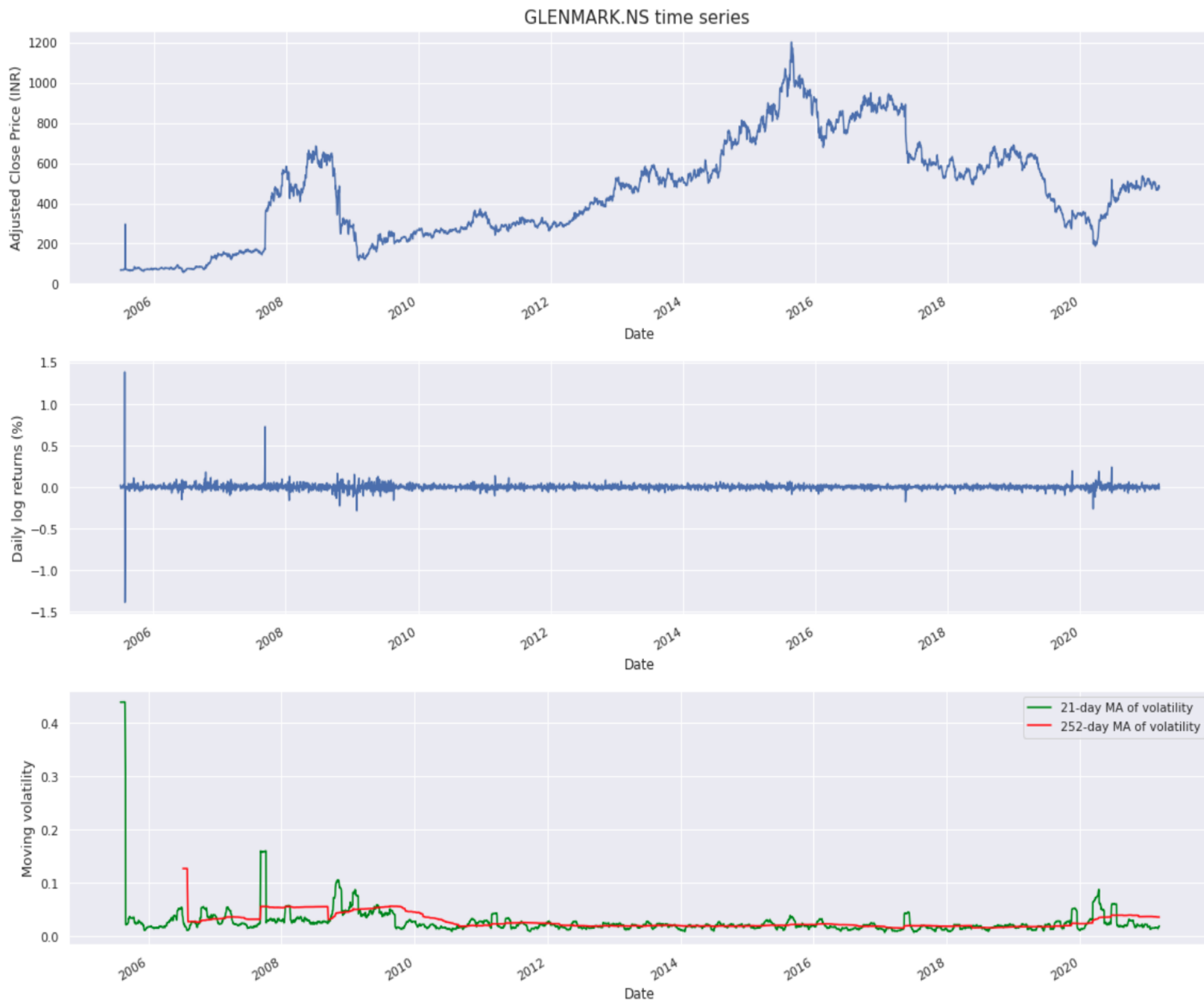
There is a negative correlation between the volatility (most measures) of an asset and its returns. i.e. When prices go up, there is likely to be less volatility in the asset returns and vice-versa.

We verify it using two different methods:

- In the first one, we measure volatility as the standard deviation of the asset returns.
- In the second one, we will use the VIX index (often called the *fear index* of Wall Street), a popular market metric that tracks volatility expectations.

Method-1: Measuring volatility

The following snapshot compares the stock returns with the monthly volatility of GLENMARK.

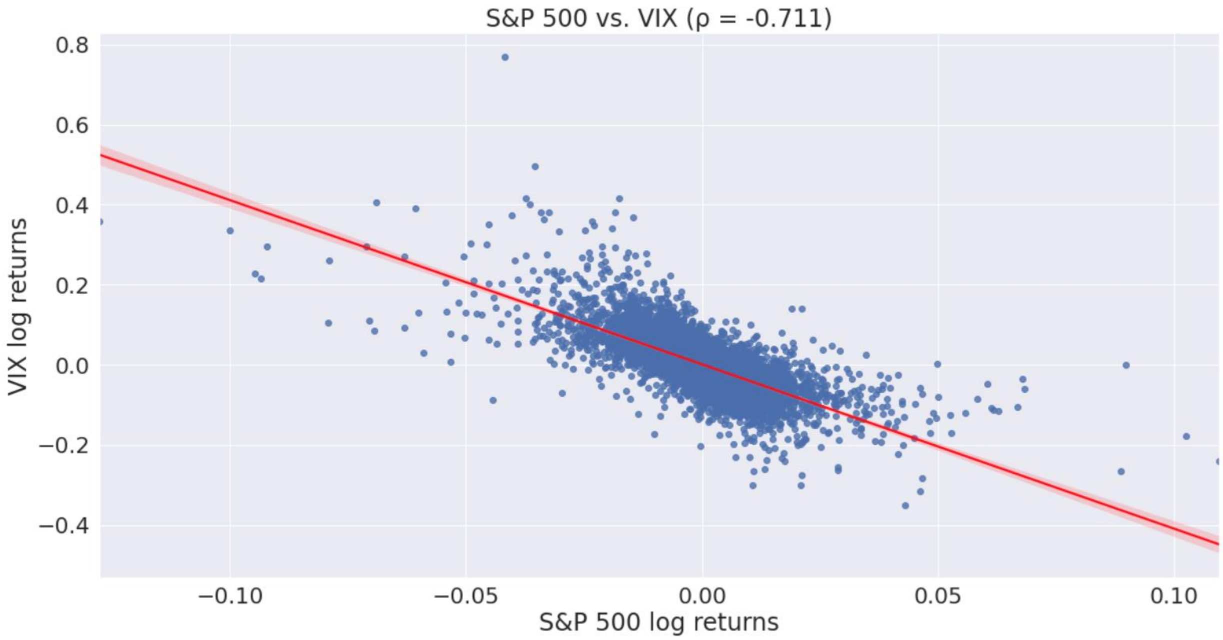


Observations

- The leverage effect is visible. There seems to be a pattern of prices going up and volatility being low in those phases and vice-versa.

Method 2 - VIX

The following snapshot shows the regression of S&P500 log-returns with VIX log returns.



Observations

- The correlation coefficient has a high negative value.
- The slope of the regression line is also negative.

<u>Additional resources</u>

- https://blog.quantinsti.com/time-series-analysis/
- https://blog.quantinsti.com/volatility-and-measures-of-risk-adjusted-return-based-on-volatility/
- https://pyflux.readthedocs.io/en/latest/getting_started.html
- https://tomaugspurger.github.io/modern-7-timeseries