How mathematical models can be developed to forecast stock prices and optimise profits?

Introduction

With the rise of the internet and advancements in technology, more and more financial transactions are being conducted electronically. This includes stock trading, as well as other forms of digital trading, such as cryptocurrency trading and foreign exchange trading. Since I grew up witnessing my parents investing in the stock market, I naturally grew curious about the intricacies of stock trading. Initially, it was difficult for me to gauge a technical analysis, but as I developed my mathematical skills, It became easier to understand and analyse visual representations like graphs and charts. This is because the stock market uses mathematics extensively and is crucial for conceptualising market behaviour, examining financial data, and making educated choices about investments. Since the stock market is extremely data-driven, quantitative analysis is employed as an important tool to decipher patterns and trends. Studying topics like statistics, calculus, and probability provided me with the foundation to analyse stock market behaviour quantitatively.

Mathematics is also further extended to stock forecasting, it helps investors to make informed decisions. I would often observe discussions within my family while deciding on trading stocks. They would take into account a variety of factors, including market trends, company financials, and personal investment strategies. I noticed numerous mathematical indicators being considered while determining when and which stocks to buy/sell. This prompted me to look into mathematical models that aid in stock analysis. I was interested in portfolio optimization, which is the method used to select the optimal asset for a particular level of risk. Mathematical tools were used to identify the combination of assets that will maximize return while minimizing risk.

Hence the research question, “How mathematical models can be developed to forecast stock prices and optimise profits?”.

Background information

The stock market is the financial epicentre for the trading of shares of companies. It is where investors can buy and sell company stock. A stock exchange like the New York Stock Exchange (NYSE) is a platform where stocks can be bought and sold through a network of traders. In order to forecast stock prices and maximise profits, mathematical models are created using a combination of mathematical concepts and algorithms that let traders and investors make profitable investments. These models use historical data and machine learning methods to spot patterns and forecast future market movements. Many mathematical algorithms already exist that are extremely nuanced in predicting stock market performance. They offer a helpful framework for comprehending the trends and patterns in the stock market, despite their imperfections and inability to accurately predict the future.

By analyzing historical trends and patterns and identifying relationships between stock prices and various factors, analysts can make more accurate predictions about future prices and optimize profits for investors. The two variables investigated in portfolio optimisation include stock values and stock returns. Portfolio optimization requires stock value forecasting because it enables investors to choose which stocks to include in their portfolio and how much capital to devote to each stock. Returns are considered when optimising a portfolio because they are a major factor in the performance of investments. Investors can maximise their portfolio's returns with the least amount of risk by estimating the future performance of a stock's potential returns and risks. Returns are an indicator of how profitable an investment is because they show the gains or losses an investor experiences over a specific time frame.

Aim

In this exploration, multiple mathematical tools like time series analysis, regression analysis and statistics were used for stock forecasting and optimising returns. The theories applied to build the forecasting model was evaluated by comparing the produced values to the actual data. Through the course of this investigation, the model was developed to account for and correct the discrepancies encountered. The stock returns were analysed using probability, probability is a useful tool for assessing stock returns because it enables investors to calculate the probability of different possible outcomes.

Mathematical analysis and evaluation

Data collection

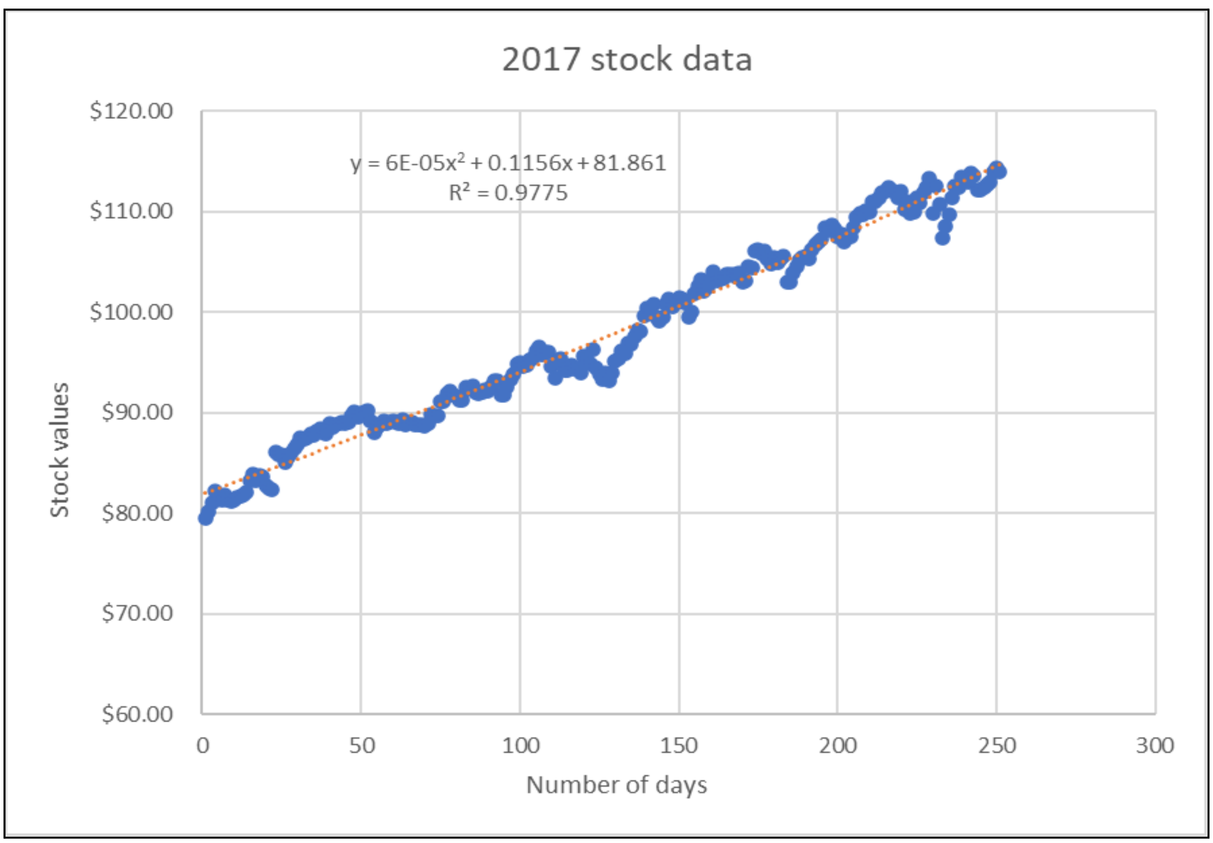
The stock data of Visa was used for this research, this is because Visa is an already mature and established company with a rather steady stock value. However, since it's been a public company for such a long period of time, the stock values were evaluated for only a fragment of this period. To avoid unnecessary discrepancies, extraneous variables like economic factors that could possibly influence the trends in the data set were avoided and so, data post 2020 wasn't included considering the repercussions of the Covid-19 pandemic in the financial sector. Data from the financial year of 2017 and 2018 was used in particular since the economy was relatively stable during this timeline.

Excel was used to gather the secondary data of Visa stock values through the 'stock history' function. The data was derived from the New York stock exchange (NYSE). The stock values and returns were presented in USD ($).

Mathematical models

<u>1.1 Non-linear regression analysis</u>

Regression analysis is a statistical model that looks at the correlation between two variables. We graphed the stock values against the number of days to illustrate the trends of the Visa stock for the financial year of 2017. The x-axis contained the independent variable 'number of days'. A total of 251 data points were obtained and plotted since the stock market is open only 5 days a week and remains closed on public holidays. The dates were assigned numerical values starting from 3 January 2017, which was assigned the value of 1, meaning that it was the first trading day. The graph illustrating the trends for the FY 2017 is as shown below through graph 1.



Graph 1 - Stock Visa non-linear regression analysis for 2017 (1 to 251 trading days)

The opening stock value for the first and last trading day for the financial year of 2017 are represented in table 1.

Trading day	Stock value ($)
1st day	$79.50
251st day	$114.02

Table 1: opening value for FY 2017

To analyse the graph, the best-fit line was used to make predictions about the dependent variable based on the independent variables. A study published in the Journal of Banking and Finance in 2009 provided evidence that non-linear models may be more effective than linear models for predicting stock returns, suggesting the importance of considering non-linear patterns in stock data.

Using technology, we produced an equation for the polynomial model.

The quadratic equation was as follows:

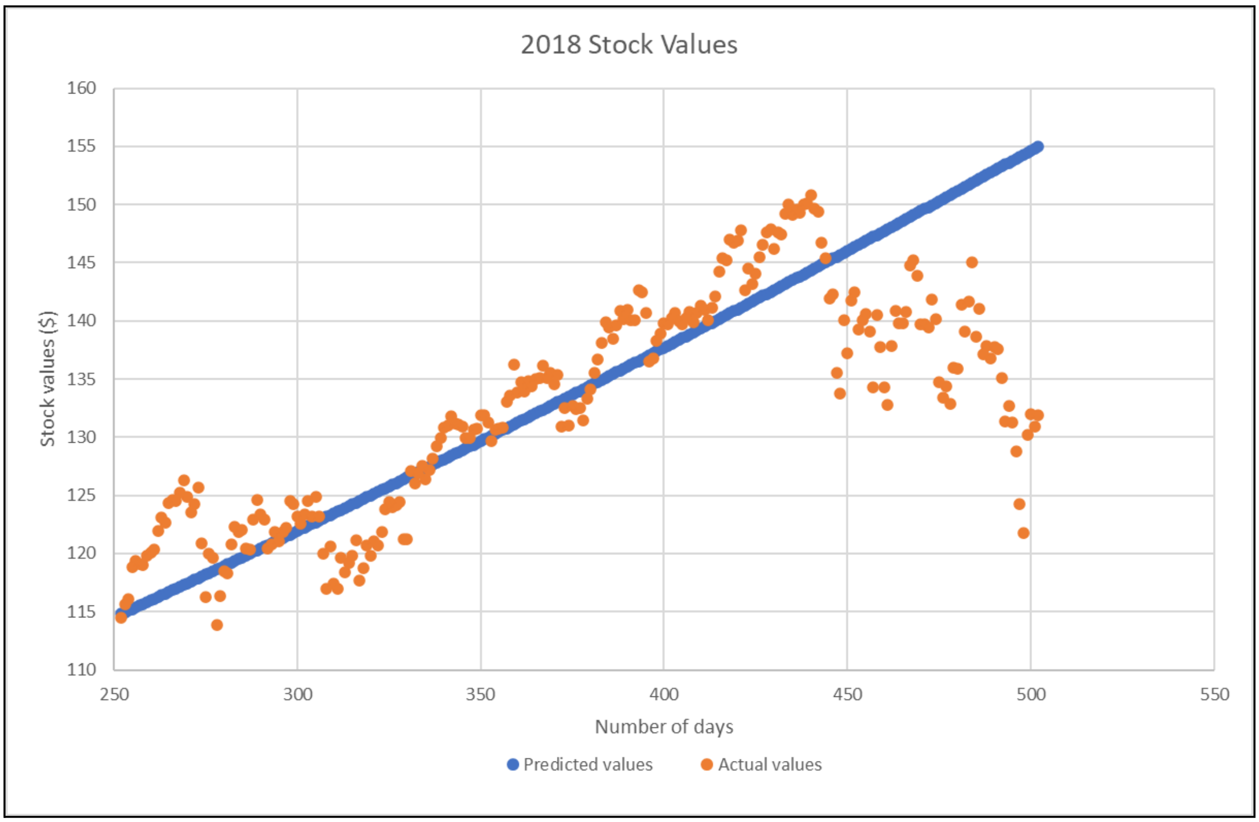
$$y = 6\times10^{-5} x^2 + 0.1156\ x + 81.861$$

Where x was the number of trading days

Equations are mathematical expressions that describe the relationship between variables plotted on a graph. We can determine the pattern or trend that the data is following by analysing the equation of a graph. We can deduce that the graph is parabolic because this function is quadratic.

We can predict future values of the variables by extending the line into the future based on the data trend.

The equation was extrapolated and therefore used to predict values for FY 2018. X was substituted with the number of days, the first trading day of 2018 was represented as the 252nd day in the equation. The trading days of FY 2018 included 250 trading days ($252 \leq x \leq 502$). The graph for the predicted values was plotted alongside the actual values from FY 2018 (graph 2), and the difference between the values was calculated in terms of percentage error.

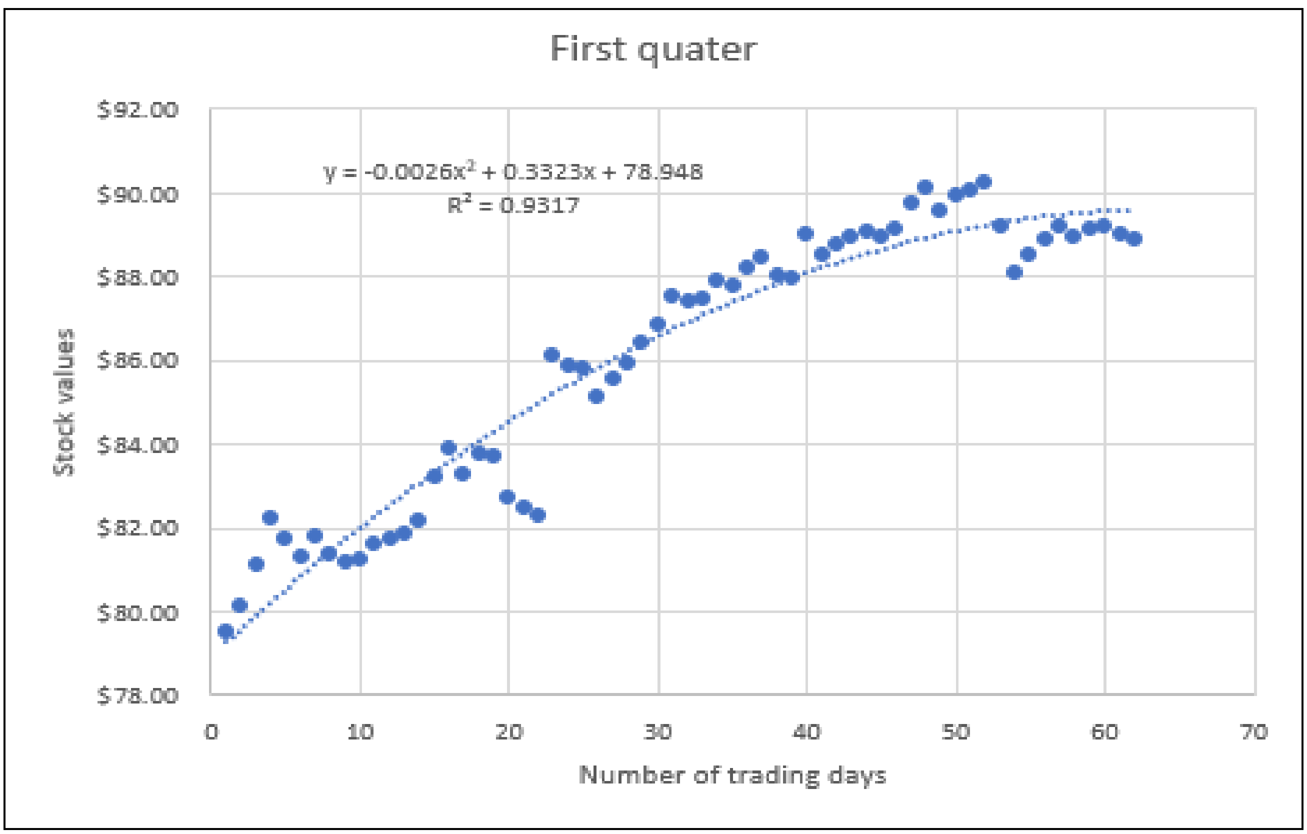


Graph 2 - predicted stock values vs actual stock values of 2018

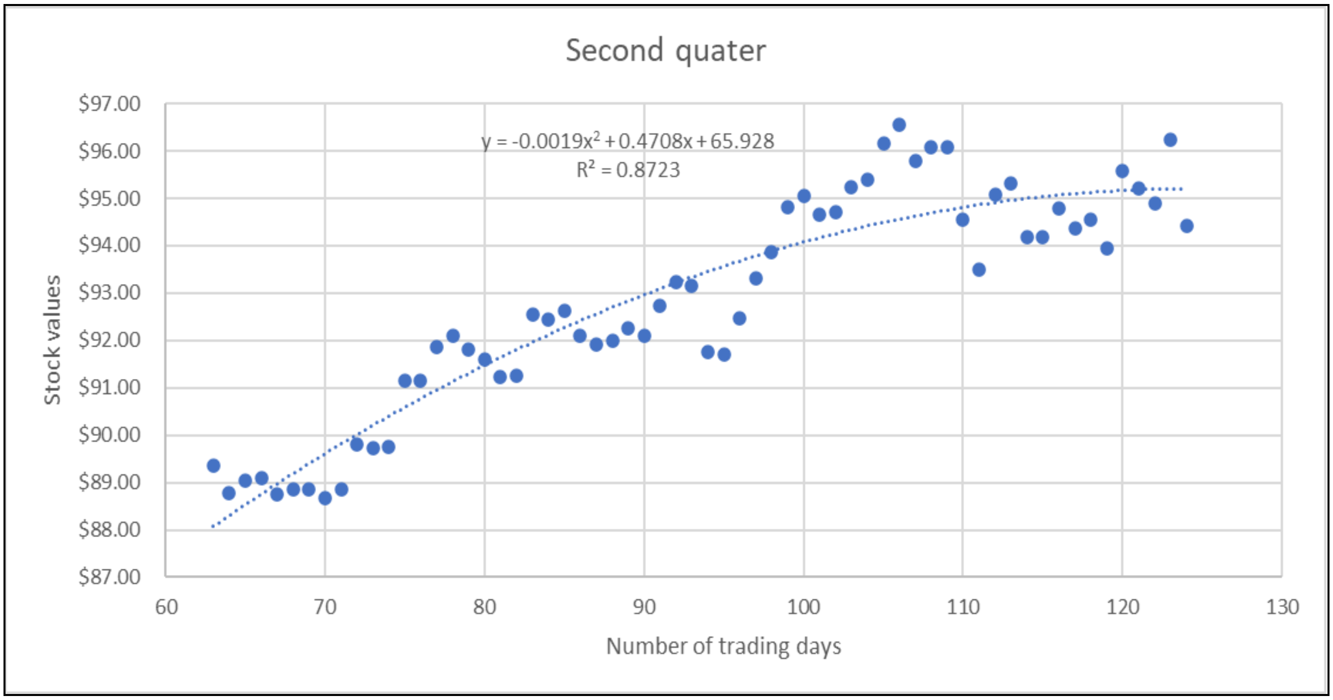
The two graphs showed some discrepancies, the predicted values depicted a relatively linear relationship whereas the actual values were more volatile. The percentage difference between the two values was calculated, and the maximum difference was 7.14%, the average percentage error was 1.27%.

To minimise the discrepancy between the predicted values and actual values, a smaller set of data was used. While larger datasets may seem like they would provide more information for forecasting stocks, a smaller dataset that contains more relevant and recent data may be more effective in providing accurate and useful predictions.

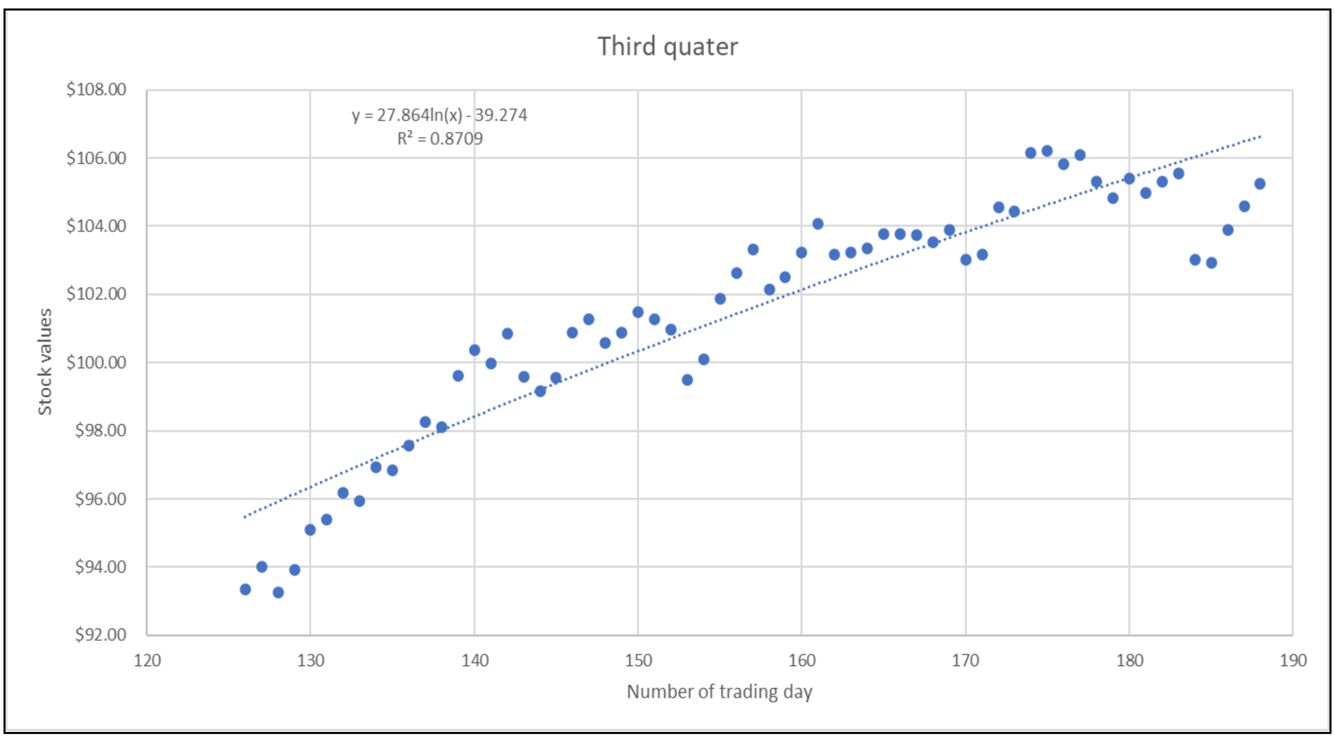
To do so, the stock values from FY 2017 were divided quarterly, and the data from the first three quarters were analysed to predict the stock values for the fourth quarter. Each of the three quarters comprised 62 trading days.



Graph 3.1 - stock prices of Visa data ($1 \leq$ 62 trading days)



Graph 3.2- stock prices of Visa data ($63 \leq 125$ trading days)



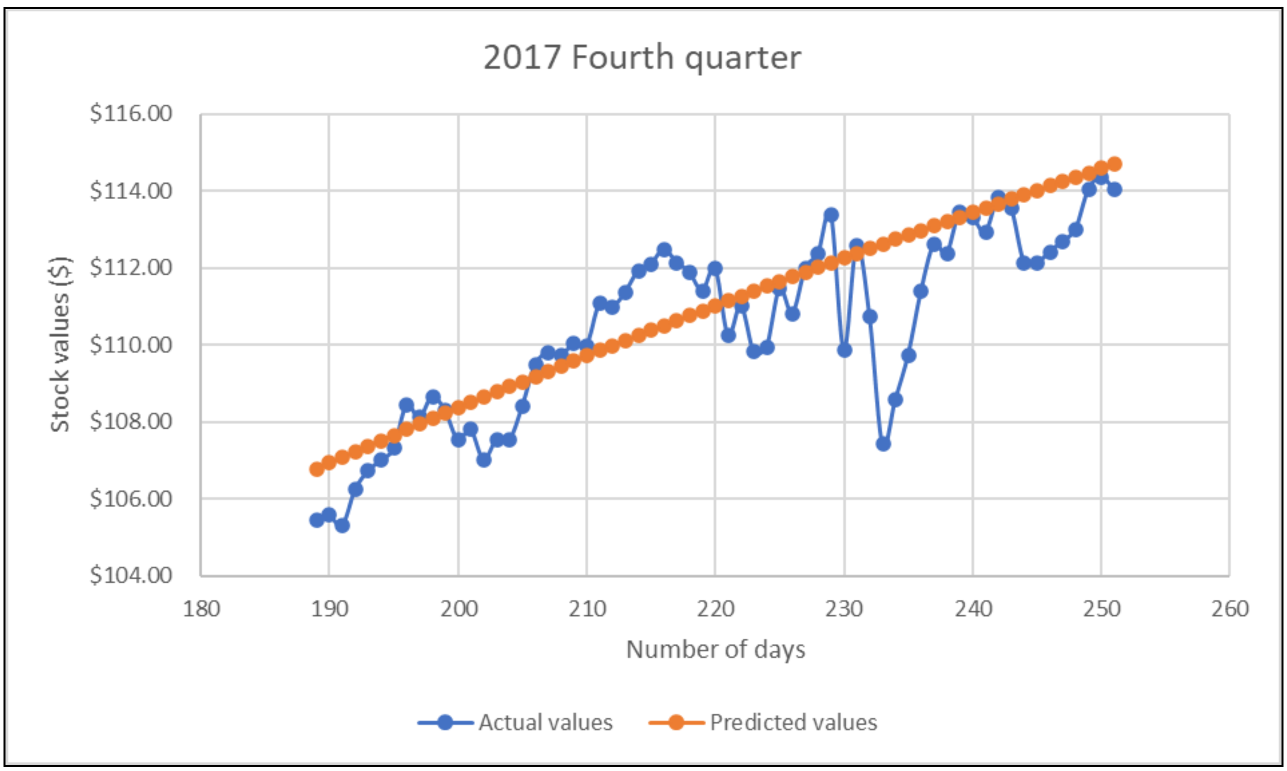
Graph 3.3 - stock prices of Visa data ($126 \leq 188$ trading days)

The logarithmic model was selected since it was the best fitting model that was closest to the trend. Using the function derived from the third quarter, the stock values for the fourth quarter were predicted. The logarithmic equation produced for the third quarter was as follows:

$$y = 27.864\ln(x) - 39.274$$

Where x is the number of days

The stock value 189th to the 252nd trading day of the FY 2017 was calculated using the formula. The percentage error was calculated between the predicted and actual values. The maximum percentage error was 4.83%. The average percentage error was 0.97%. The percentage error was unexpectedly higher. The percentage error suggests that the model isn't precise but relatively accurate.



Graph 4 - actual and predicted values for the fourth quarter of FY 2017 (189 to 252 days)

The graph for the predicted values were plotted alongside the actual values for the fourth quarter of FY 2017 (graph 4). We can infer that the model isn't taking into account minor trends and fluctuations, which may be the cause of the discrepancy between the predicted and actual values.

1.2 Simple moving average

An additional analytical tool used to predict stock values are time series analysis like the simple moving average (SMA). A moving average is a statistical calculation used to analyze data points by creating a series of averages of different subsets of the full data set. In stock analysis, a moving average is used to smooth out the fluctuations in stock prices over time and to identify trends which is why it may produce more precise predictive values. The limitations of the earlier model, particularly the minor trends and fluctuations, might be considered in this model and hence, can be presumed that it is a more accurate model.

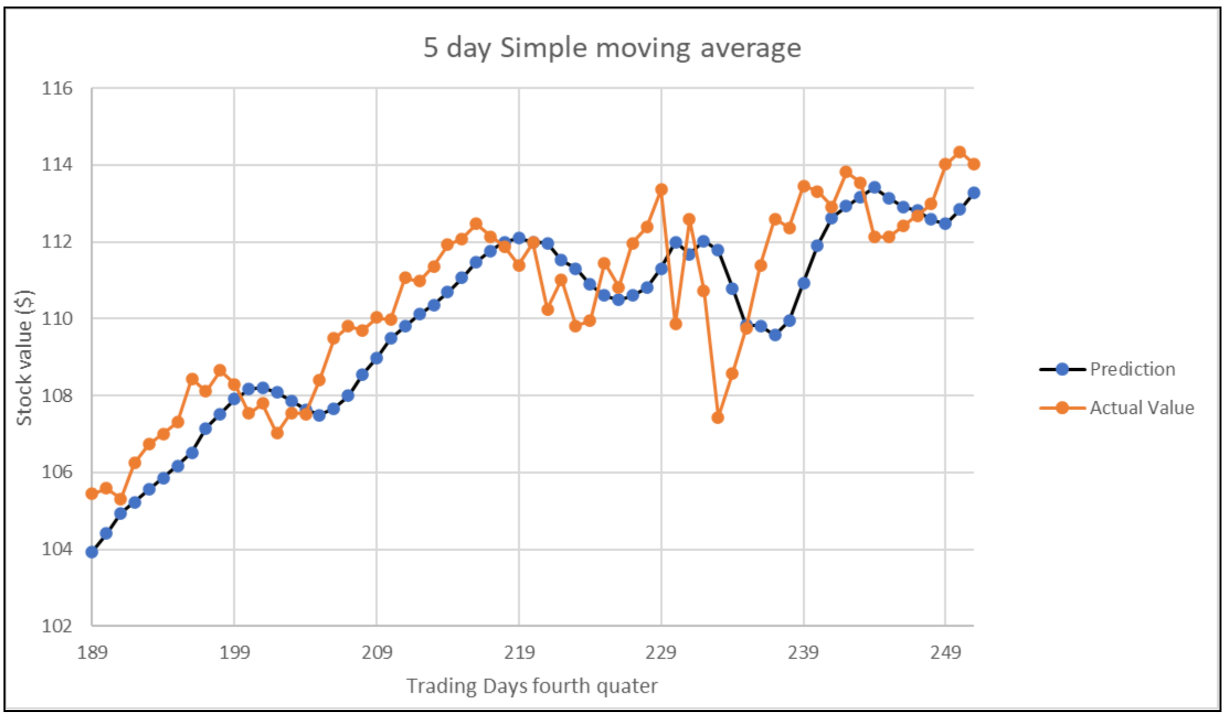
A simple moving average is calculated by adding up the prices of a stock over a certain number of time periods and then dividing the sum by the number of periods. An exponential moving average gives more weight to recent data points and less weight to older data points. The moving average is calculated using the most recent data and changes as new data becomes available. Unlike the previous approach, the simple moving average can not be used to determine the stock values for a long period of time. The SMA is expected to produce more accurate values than the earlier approach.

Moving averages can be used to predict stock values by analyzing the relationship between the stock price and its moving average. For this investigation, a 5-day simple moving average for the fourth quarter of FY 2017. Calculating the 5-day simple moving average (SMA) for Visa's stock would involve adding up the closing prices of Visa's stock over the past 5 days, and then dividing by 5 to get the average. This was computed using technology.

The SMA for the 189th trading day would be calculated as such:

$$\frac{(105.24+104.58+103.89+102.94+103.02)}{5} = 103.934$$

The predictions for the fourth quarter using the simple moving average were then graphed against the graph of the actual values.



Graph 5 - simple moving average in the fourth quarter

The difference between the predicted values and actual values was considerably small. The average percentage error between the actual and predicted values using the SMA approach was approximately 0.4%. The lowest percentage error was 0.01%, whereas the highest percentage error was 4.05%. The percentage error between the predicted and actual values was found to be lowest for the Simple Moving Average. Therefore since it produced the most accurate values, we can conclude that the simple moving average is the most reliable analytical method to forecast stock values.

1.3 Normal distribution

Since stock returns play an important role in portfolio optimisation. Stock returns are changes in the value of a stock. This calculation is based on the idea that the return on an investment is the profit or loss you make on that investment over a certain period of time. The formula for stock returns used in this exploration is as such:

Stock Return = opening stock price for a day - opening price for the previous day

Statistical tools were used to determine the probability of earning a profit by investing in Visa's stock during the financial year of 2017. Determining the distribution of a set of data is important in calculating probability because it provides valuable information about the characteristics of the data and how likely different outcomes are. If the data follows a particular distribution, the properties of that distribution can be used to estimate the likelihood of different outcomes and make informed decisions about risk and return.

The stock returns of the Visa stock are assumed to follow a normal distribution. This is because The data seems to be symmetric around the mean when the set of data was plotted on a histogram (figure 1).

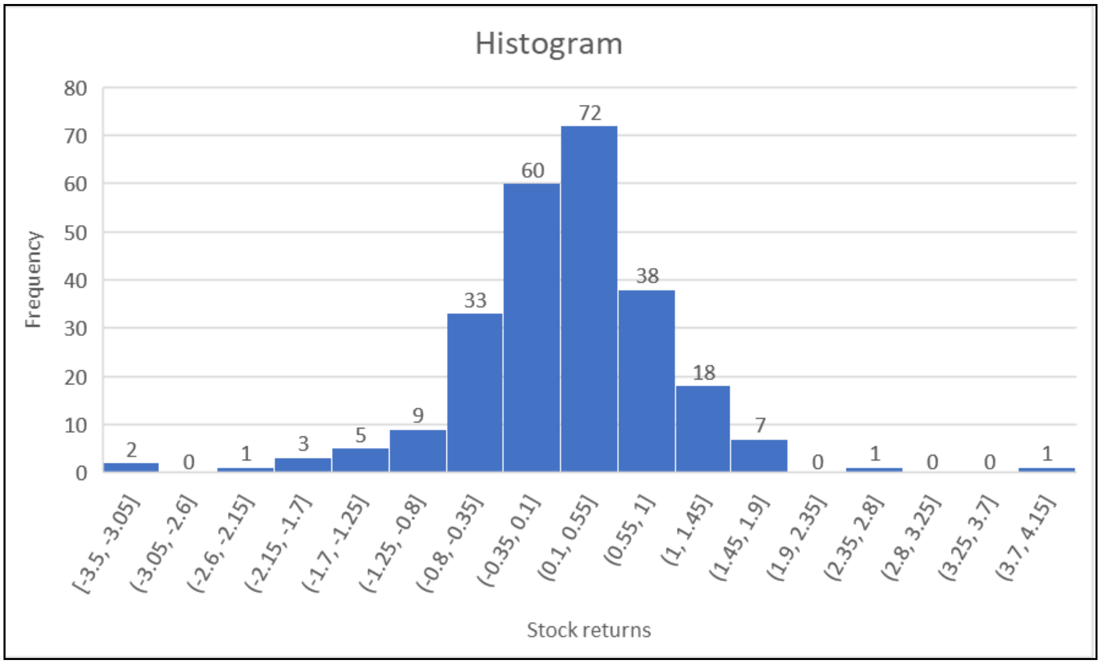


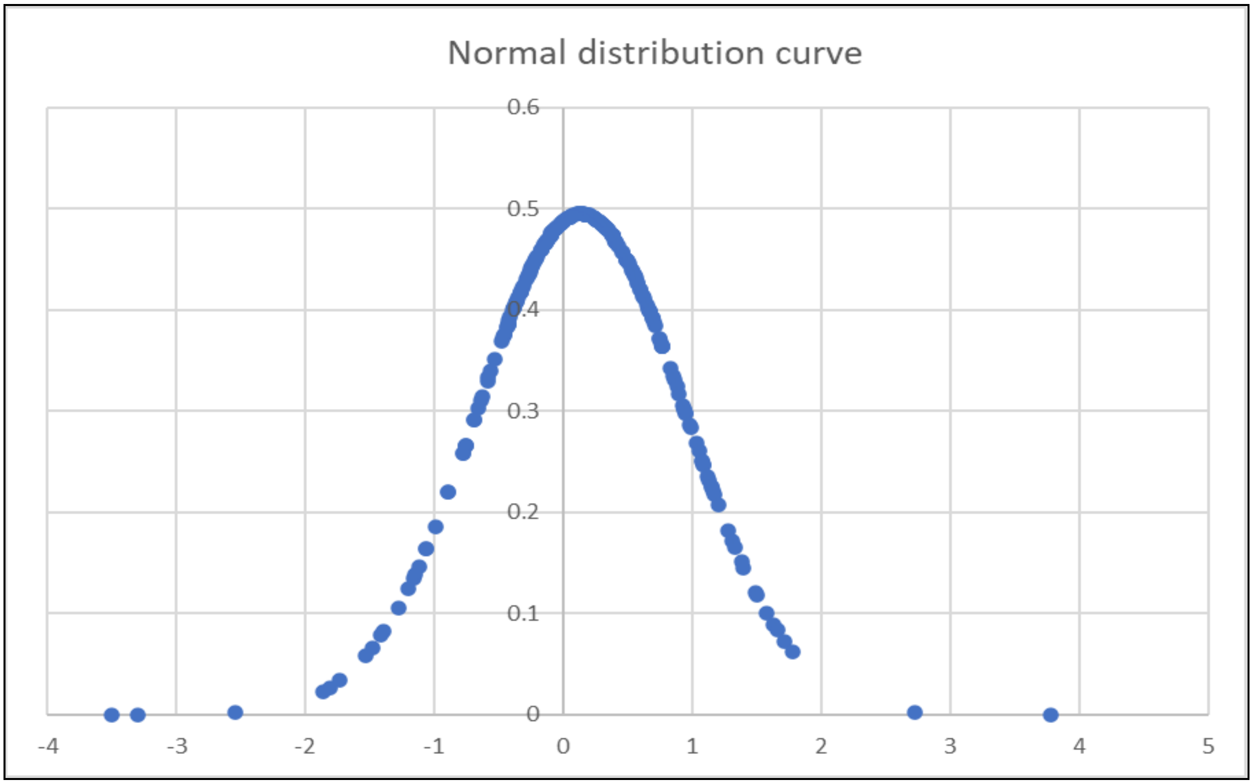
Figure 1 - Histogram for the 2017 stock returns

The mean and standard deviation are the two parameters that establish the normal distribution. The normal distribution can be used to calculate the probability of stock returns by standardising the returns and then using the standard normal distribution to calculate the probability of different returns or ranges of returns. The stock returns were calculated for 2017 by finding the difference between the latest value and the stock value from the former trading day. The mean and standard deviation for the stock returns were calculated (table 2).

Mean	**Standard deviation**
0.138	0.806

Table 2 - mean and standard deviation for stock returns

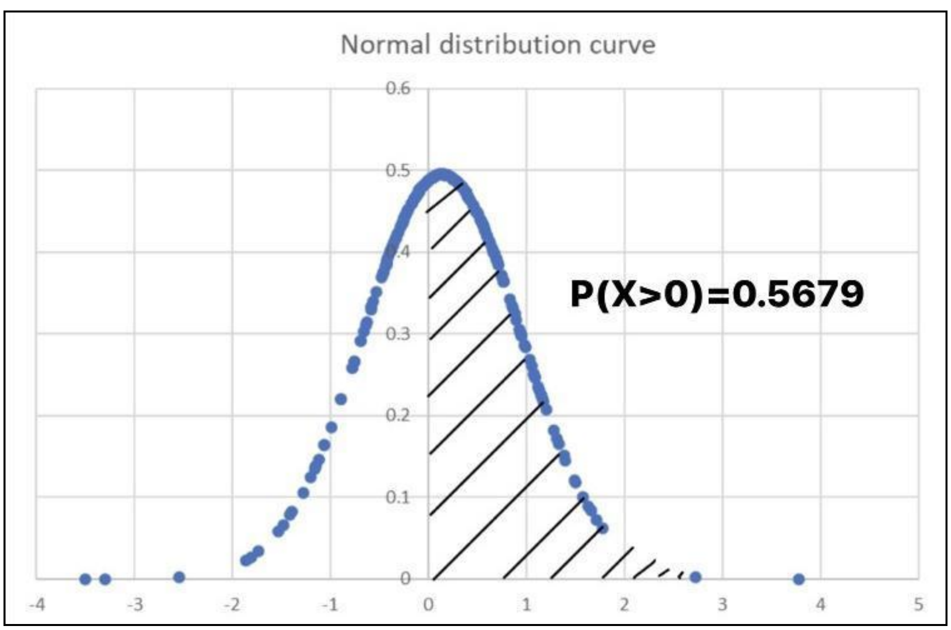
The normal distribution curve for the stock returns of the FY 2017 is as follows:



Graph 6 - normal distribution curve

The normal distribution is symmetrical around the calculated mean. The shaded region represented by the graph below illustrates the probability of the stock returns being positive.

The normal distribution curve for the stock returns of the FY 2017 is as follows:



Graph 7 - normal distribution curve P(X > 0)

The probability for the stock returns being positive was found using the probability density function (PDF). The PDF can be used to quantify the likelihood of a random variable falling within a specific range of values.

The probability density function:

$$P(X = x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\times\left(\frac{X-\mu}{\sigma}\right)^2}$$

$$P(X = x) = \frac{1}{0.806197576\sqrt{2\pi}} e^{-\frac{1}{2}\times\left(\frac{X-0.13808}{0.806197576}\right)^2}$$

For investigating stock returns, the probability of the stocks growing was calculated, hence, the PDF for the stock returns to be positive $P(X>0)$ was found. Using technology, the probability of stock returns being positive for the Visa stock for the year of 2018 was calculated to be **0.5679**, through this, we can infer that approximately **57%** of the time, the stock returned profitable returns.

Conclusion and evaluation

Through the use of mathematics, appropriate models for stock forecasting were devised, which each yielded their respective percentage errors by comparing the percentage difference between the actual predicted values generated from each model, the most reliable model was established. The table below is a compilation of the percentage error.

Model	Average percentage error
Non-linear regression long-term analysis	1.27%
Non-linear regression short-term analysis	0.97%
Simple moving average	0.4%

Table 3 - percentage error for all the models

The simple moving average generated the least percentage difference whereas the non-linear regression for the long term showed the most percentage error. Therefore, we can infer that the SMA model produced the most accurate values in comparison to the other two models. Hence, we can conclude that it is the most reliable method for forecasting stocks in this investigation.

However, while the models are derived and tested on real-world data, I can not be certain of their accuracy when applied as a regular tool to forecast stocks. This is largely due to a number of assumptions made over the course of the investigation:

Assumption 1: The stock price will follow its past performance.

Limitations: This assumption is based on the premise that stocks grow at a constant rate which is untrue. While there are some patterns and trends that can be observed in the performance of individual stocks and the stock market as a whole over time, there are many factors that can influence the future price of a stock, and these factors are often unpredictable.

Assumption 2: The model is not overfitting the data

Limitation: Overfitting occurs when a model is too complex and fits the historical data too closely, which can lead to inaccurate predictions when applied to new data. Equations derived from historical data assume that the model is not overfitting the data and that it can be applied to new data accurately.

These assumptions limit the validity of the models. Further, as an improvement to the existing model, mathematical concepts like the Fibonacci sequence could be used that is applied in existing algorithms. The Fibonacci ratio helps in producing levels that can be used to identify potential areas of support or resistance where a stock's price may reverse direction.

Additionally, a number of assumptions made while analysing stock returns could also produce inaccurate probabilities:

Assumption 3: independence of stock returns

Limitations: The normal distribution probability density function assumes that the observations in the data set are independent of each other. In other words, the value of one observation does not affect the value of another observation. However, in many real-world scenarios, observations may not be independent, which can affect the accuracy of the probability estimates.

Assumption 4: The data follows normal distribution

Limitations: The normal distribution probability density function assumes that the data follows a normal or bell-shaped distribution. However, many real-world datasets do not follow a normal distribution, and using the normal distribution PDF to estimate probabilities in such cases may not be accurate.

The probability calculated for the stock returns of FY 2017 can not be representative of future stock returns since they're dependent on stock prices which do not grow at a constant rate.

Bibliography

1. Haese, Michael, et al. *Mathematics: Applications and Interpretation HL*. 2019.
2. “What Is SMA? - Simple Moving Average - Fidelity.” *What Is SMA? - Simple Moving Average - Fidelity*, www.fidelity.com/learning-center/trading-investing/technical-analysis/technical-indicator-guide/sma.
3. “The True Probability Distribution of Stock Market Returns.” *TradeOptionsWithMe*, 20 Mar. 2020, tradeoptionswithme.com/probability-dsitribution-of-stock-market-returns.
4. “Using the Nonlinear Regression to Predict the Prices of Stocks and Its Impact on the Investor Decisions in the Financial Market.” *ResearchGate* , May 2020, www.researchgate.net/publication/341617753_Using_the_nonlinear_regression_to_predict_the_prices_of_stocks_and_its_impact_on_the_investor_decisions_in_the_financial_market.