Se former autrement

HONORIS UNITED UNIVERSITIES

Unlocking Market Insights: A Study of Volatility and Investor Behavior in Finance

LAABIDI Haifa
Tutored by: Hedhli Amel

Délivrée par la Commission des Titres d'Ingénieur

CONFÉRENCE DES GRANDES ÉCOLES

Acknowledgement

My greatest appreciation goes out to Mme. HEDHLI for their crucial advice and assistance throughout the course of this project. Their knowledge, guidance, and commitment were crucial in determining the project's course and elevating its calibre. They made a significant impact on my academic journey with their perceptive criticism and persistent support, and I am sincerely appreciative of them.

Table of Contents

List of Abbreviations

1. ARCH : The autoregressive conditional heteroskedasticity
2. GARCH : Generalised AutoRegressive Conditional Heteroskedasticity
3. ^NYA: NYSE Composite (New York Stock Exchange)
4. ^IXIC: NASDAQ Composite
5. ^FTSE: FTSE 100 Index (Financial Times Stock Exchange, UK)
6. ^NSEI: Nifty 50 (National Stock Exchange of India)
7. ^BSESN: BSE SENSEX (Bombay Stock Exchange, India)
8. ^N225: Nikkei 225 (Japan)
9. 000001.SS: SSE Composite Index (Shanghai Stock Exchange)
10. ^N100: Euronext 100 (European Stock Exchange)
11. ^DJI: Dow Jones Industrial Average (USA)
12. ^GSPC: S&P 500 Index (USA)
13. GC=F: Gold Futures
14. CL=F: Crude Oil Futures
15. ADF: Augmented Dickey-Fuller
16. OLS: Ordinary Least Squares
17. LM: Lagrange Multiplier
18. AR : Autoregressive Model
19. MA : Moving Average
20. GDP: Gross domestic product
21. CPI : consumer price index
22. AAPL: apple stock market
23. TSLA: tesla stock market
24. RSI: Relative Strength Index
25. SMA: Simple Moving Average
26. EMA: Moving Average

List of Figures

List of Tables

General Introduction

The finance sector went through a significant transition in a time of unheard-of technological developments. The time when traditional financial models and intuition alone guided investing decisions is long past. The world of finance is currently becoming more and more entangled with the data-driven expertise of the twenty-first century: the field of data science. Data science skills are now essential for understanding, forecasting, and optimising investment results as financial markets become more complicated, global, and linked.

The purpose of this project is to explore the complex interrelationship between the fields of data science and finance. By highlighting how data science approaches and tools are essential in unravelling these intricacies, it aims to expose the subtle processes that underpin market volatility and investor behaviour. In doing so, we investigate the complex financial market environment where market trends, economic trends, and statistical methods all interact to influence investment choices.

The study is broken down into two main chapters, each of which discusses a crucial component of financial analysis. Detailed analysis of stock market volatility is provided in Chapter 1, along with statistical models that help us forecast and control market turbulence. The fascinating topic of investor behaviour is explored in Chapter 2, which also provides insight into the economic factors that influence investors' choices. Throughout our trip, we will make use of data science and statistical analysis, breaking down complex financial datasets and generating actionable insights using tools like R.

We will discover the significance of data science is in influencing investment strategies and financial decision-making as we move through this report. Additionally, we will develop a deeper understanding of the complex relationship between investor behaviour and market volatility, a knowledge that is essential for success in the always changing world of finance.

The robust data science methods that direct the study are at the core of it. As our analytical compass, we will make use of R, an effective statistical programming language. We may move through enormous financial datasets with the help of R, which will make data preparation, statistical analysis, and modelling easier.

In Chapter 1, we dig into the domain of market volatility. To solve the questions of volatility in the market, we use statistical models like GARCH (Generalised Autoregressive Conditional Heteroskedasticity) and ARCH (Autoregressive Conditional Heteroskedasticity). Our tools for comprehending, forecasting, and managing volatility will be these models. The area of investor behaviour is explored in Chapter 2 where we use data science tools to analyse the economic factors that influence investing choices. To shed light on how investors view and respond to market patterns, we will use techniques like moving averages and the Relative Strength Index (RSI).

Chapter 1

Volatility Study in the Stock Markets

1.1 Introduction

Volatility is a concept at the centre of every investment decision. It is frequently referred to as the restless pulse of financial markets. It is the invisible force that drives prices and shifts investor attitude, the heartbeat of the stock market. In this chapter, we set out on a study into the topic of market volatility in an effort to comprehend both its definition and its profound relevance in the financial industry.

The Concept of Volatility

Volatility essentially measures the size of price swings in financial assets across time. It gives insight into the ebb and flow of market sentiment and measures the uncertainty and risk that investors confront. But why is volatility a topic that is so crucial? It is the absolute personification of opportunity and risk, to put it simply.

significance of Volatility

Volatility has two sides to it. On the one hand, it offers traders and investors the chance to benefit significantly from price fluctuations. On the other hand, it adds a level of risk and unpredictability that could result in substantial losses. Therefore, it is essential for anyone navigating the financial markets to comprehend volatility and be able to measure, calculate, and predict it.

1.2 Business Introduction

Market volatility is a key notion in finance that influences how firms, investors, and financial institutions make decisions. Volatility, or more specifically, how much financial asset prices fluctuate, is not just a theoretical idea; it also directly affects how firms operate and how much money they make.

A statistical measurement of volatility is the dispersion of returns for a certain securities or market index. Most of the time, a security is riskier the more volatile it is. The standard deviation or variance of returns from the same securities or market index is frequently used to calculate volatility.

Volatility in the financial markets is frequently characterised by significant swings in either direction. For instance, the stock market is said to be volatile when it fluctuates by more than 1% over an extended period of time. When determining the price of an option contract, an asset's volatility is important.

Businesses increasingly use data science to evaluate and negotiate the unpredictable financial markets in today's data-rich economy. Organisations are given the analytical tools they need by data science to not only respond to market volatility but also proactively spot opportunities that might occur amid the uncertainty.

In this chapter, we will examine market volatility's practical importance for organisations while delving into its numerous dimensions. We will investigate patterns, evaluate risks, and pinpoint potential opportunities related to volatility for businesses, investors, and financial stakeholders using data science approaches and statistical models.

1.3 Data Preparation

It is crucial to create a solid foundation through efficient data preparation before navigating the complex world of market volatility analysis. The thorough procedure for data collection and preprocessing is covered in this chapter, shedding light on the crucial measures taken to guarantee the accuracy and integrity of the data used in our study.

The choice of reliable data sources forms the basis of any data analysis project. We have used data from trustworthy financial data suppliers [1] in our effort to comprehend market volatility. These sources cover a wide range of financial tools, such as indexes, stock prices, and pertinent economic indicators. Our analysis is thorough and representative of actual market conditions thanks to the breadth and depth of our data sources.

Our dataset encompasses daily price and volume data from several major global stock indices and commodities for the period of 2008 to 2023. The following financial instruments are included:

- ^NYA: NYSE Composite (New York Stock Exchange)
- ^IXIC: NASDAQ Composite
- ^FTSE: FTSE 100 Index (Financial Times Stock Exchange, UK)
- ^NSEI: Nifty 50 (National Stock Exchange of India)
- ^BSESN: BSE SENSEX (Bombay Stock Exchange, India)
- ^N225: Nikkei 225 (Japan)
- 000001.SS: SSE Composite Index (Shanghai Stock Exchange)
- ^N100: Euronext 100 (European Stock Exchange)
- ^DJI: Dow Jones Industrial Average (USA)
- ^GSPC: S&P 500 Index (USA)
- GC=F: Gold Futures
- CL=F: Crude Oil Futures

For each ticker symbol, the following columns are included:

- Date: The date of the data point.
- Open: The opening price for the given date.
- High: The highest price for the given date.
- Low: The lowest price for the given date.
- Close: The closing price for the given date.
- Adj Close: The adjusted closing price for the given date.
- Volume: The number of shares traded on the given date.

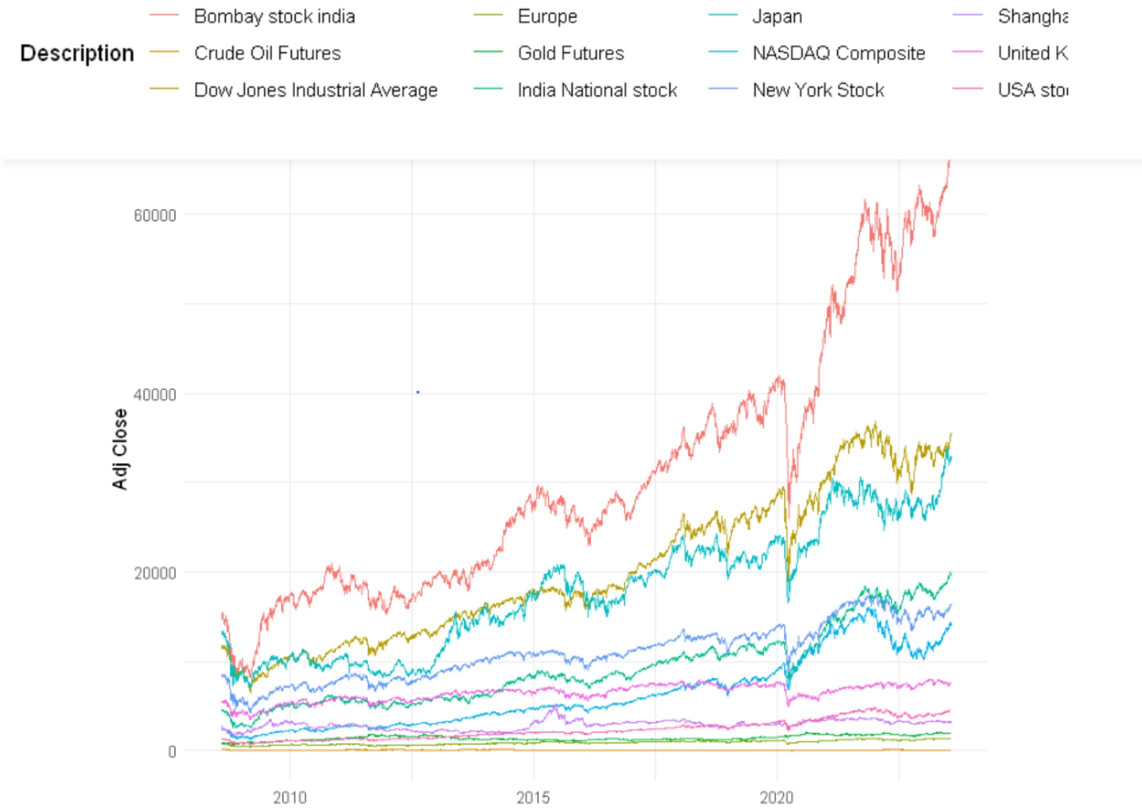

Fig 1.1: Stock price movement

For a volatility and investor behaviour analysis, the primary focus should typically be on the "Close" price column, which represents the closing price of the stock on each trading day and 'Volume' which represents The number of shares traded on the given date, in that case Open , High and Low columns will be deleted.

Close Price: The "Close" price is the final trading price of a security at the end of a trading session. It represents the last price at which a trade occurred before the market closed. The "Close" price is often used for general analysis and visualisation of stock prices.

Adjusted Close Price: The "Adj Close" price is the closing price adjusted for factors like dividends, stock splits, and other corporate actions. It accounts for these adjustments to provide a more accurate representation of the investment's value over time.

The "Adj Close" price is particularly important for long-term analysis and for tracking the actual returns an investor would have received. After financial research, it's common to use the "Adj Close" price when conducting studies that require precise return calculations, such as calculating historical volatility, risk-adjusted returns, or portfolio performance over extended periods.

After ensuring that the dataset is robust, accurate, and ready for the rigorous analysis that follows that dataset was handled in order to be clean and well aligned in time.

	Ticker	Date	Open	High	Low	Close	Adj.Close	Volume
2008.1	^NYA	2008-08-01	8438.71	8452.01	8356.43	8379.15	8379.15	4684870000
2008.2	^IXIC	2008-08-01	2326.83	2328.95	2286.41	2310.96	2310.96	2312140000
2008.3	^FTSE	2008-08-01	5411.90	5411.90	5321.30	5354.70	5354.70	1341947000
2008.4	^NSEI	2008-08-01	4331.60	4422.95	4235.70	4413.55	4413.55	0
2008.5	^BSESN	2008-08-01	14064.26	14682.33	14032.87	14656.69	14656.69	40200
2008.6	^N225	2008-08-01	13276.57	13294.17	13039.21	13094.59	13094.59	135000000

Table 1.1: A subset of the dataset

1.4 Market Correlation

The statistical relationship between the price changes of various financial assets, such as stocks, indices, or commodities, is measured by market correlation. It measures how much these assets move in concert or conflict with one another. The range of correlation values is -1 to 1:

- **When two assets have a positive correlation (0 to 1):** it indicates that they are likely to move in the same direction. The likelihood of the other asset increasing increases if the first one does.

- **Negative Correlation:** When two assets are negatively correlated, they are said to move in opposing directions. When one asset appreciates, the other typically declines.

- **No Correlation:** A correlation of zero indicates that there is no obvious connection between the movements of the two assets.

When two markets have a strong positive correlation, it means that they tend to move together in a very similar manner. In such cases, having both markets in the analysis or investment portfolio might not provide much diversification, as their movements are highly correlated. In order to reduce redundancy and simplify analysis, we can consider removing one of the markets with a strong positive correlation. This would allow us to focus on markets that provide more unique information or diversification potential.

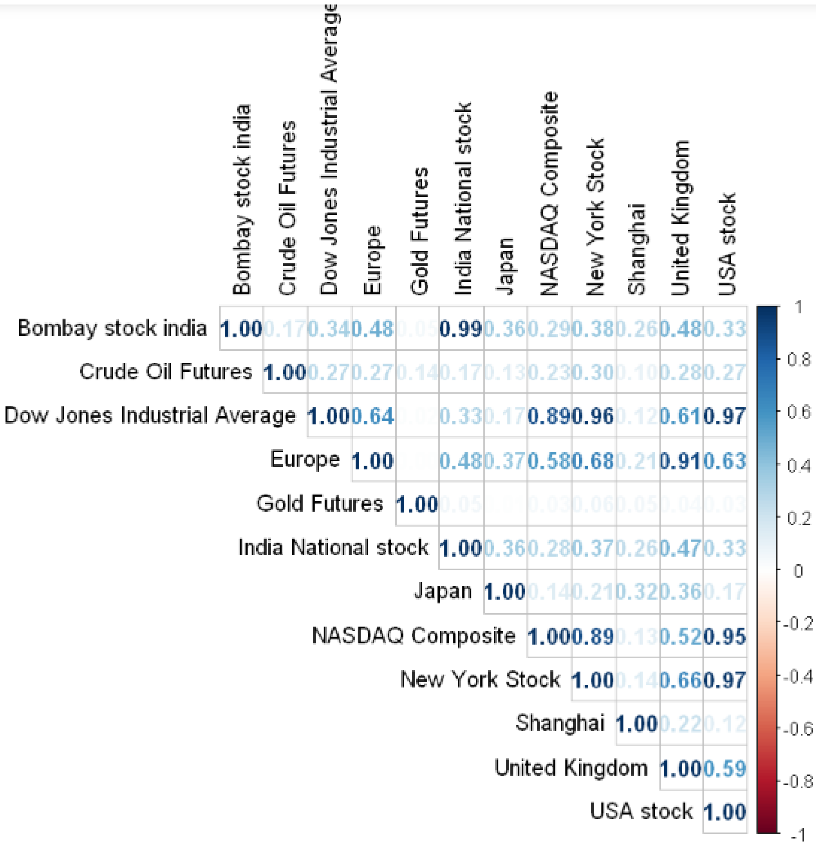

Fig 1.2: Markets Correlation Heatmap

- Bombay Stock Exchange and National Stock Exchange of India : 0.99
- Dow Jones Industrial Average and New York Stock: 0.96
- Dow Jones Industrial Average and USA stock: 0.97
- NASDAQ Composite and New York Stock: 0.89
- NASDAQ Composite and USA Stock: 0.95
- New York Stock and USA stock: 0.97

To reduce redundancy, we should remove one market from each pair.In terms of maintaining a diverse representation, markets with a lower significance or relevance in the analysis will be removed.

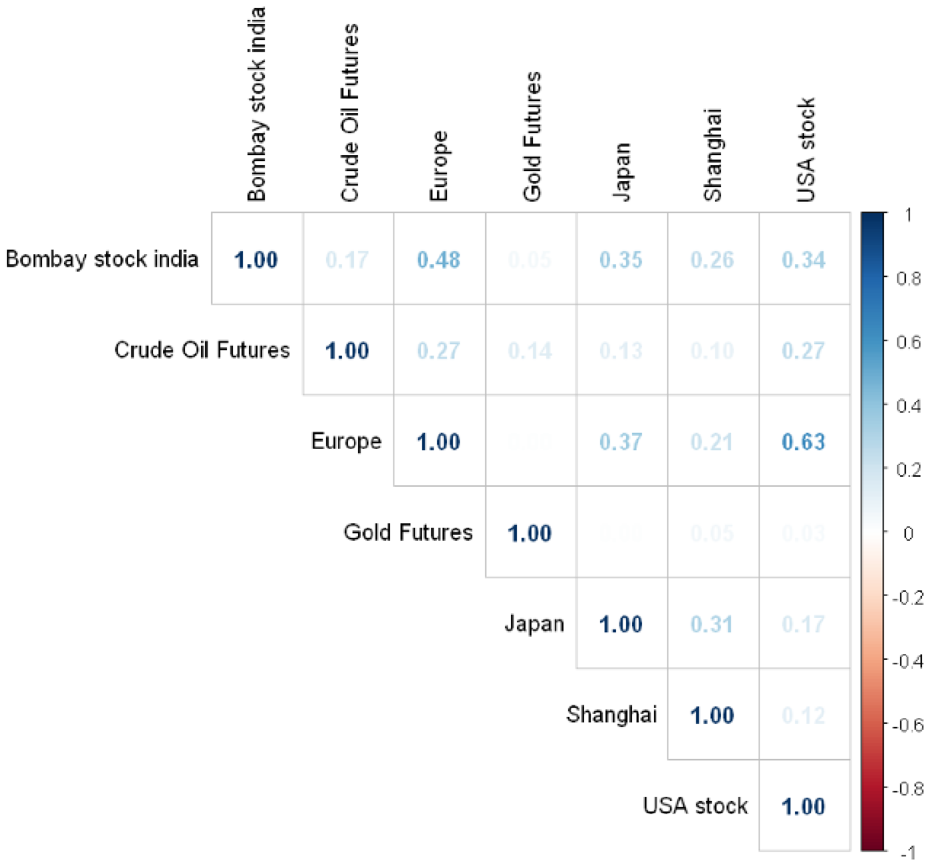

Fig 1.3: Markets Correlation Heatmap

- **Nifty 50:**it had a very high correlation (0.99) with BSE SENSEX. Since both Nifty 50 and BSE SENSEX are major Indian stock market indices and represent a similar basket of stocks from the Indian market, they provide redundant information. Removing one of them can help reduce multicollinearity in analysis.
- **Dow Jones Industrial Average:** This index was removed because it had strong correlations (0.89) with both NASDAQ Composite and NYSE Composite. Dow Jones Industrial Average, NASDAQ Composite, and NYSE Composite represent different aspects of the US stock market, but their strong correlations can lead to redundancy in analysis.
- **NASDAQ Composite:** We removed NASDAQ Composite due to its high correlation (0.89) with both Dow Jones Industrial Average and NYSE Composite. NASDAQ Composite, Dow Jones Industrial Average, and NYSE Composite all provide insights into the US stock market, but their high correlations could introduce multicollinearity issues.
- **FTSE 100:** FTSE 100 was removed because of its strong correlation (0.91) with Euronext 100. Both indices cover European markets, and their high correlation might result in redundant information in analysis.
- **NYSE Composite:** NYSE Composite was removed due to its high correlation (0.97) with S&P 500. Both indices represent the US stock market and have a very strong correlation, which could lead to multicollinearity problems in analysis.

1.5 Market Returns

Market returns act as a fundamental indicator in the field of financial analysis, providing information on the ever-changing structure of the financial markets. They give an investor a precise estimate of the gains or losses they might anticipate from keeping a financial asset over a specific time period. Understanding various measures of market returns is crucial in the context of our investigation of market volatility since they are essential for determining and deciphering market turbulence.

It is impossible to exaggerate the significance of these many market return indicators in the examination of volatility. Here's why they're important:

Granularity: Daily returns give analysts fine-grained information that enables them to spot minute changes in market sentiment as well as patterns of intraday volatility.

Stability Assessment: Weekly and monthly returns can be used to analyse the stability of returns over a wider time period, which is helpful in recognizing underlying market conditions and analysing longer-term patterns.

Risk Evaluation: Risk evaluation is aided by examining returns over various time frames. For long-term investors, short-term volatility may not be a major worry, but for day traders or other short-term investors, it can be very important.

Portfolio Management: Investors and portfolio managers build and oversee portfolios that balance risk and return using a variety of return measures.

We will use these market return measures in the parts that follow our analysis to examine the complex connection between market volatility and returns.

The "performance" column calculated represents the daily returns for the index based on the provided formula. Each value in the "performance" column represents the percentage change in the index's price from one day to the next, expressed as a logarithmic return.

Formula:

$$Ri,t \;=\; 100 \times [ln(Pi,t\;/\;Pi,t\;-\;1)]$$

Where:

$R_{i,t}$ is the daily return (performance) for the index on day t.

$P_{i,t}$ is the price of the index at the end of day t.

$P_{i,t-1}$ is the price of the index at the end of the previous day (day t - 1).

ln denotes the natural logarithm. The result is multiplied by 100 to express the return as a percentage.

```
# A tibble: 44,888 x 5
# Groups:   Description [12]
   Description        Date       Adj.Close Volume Performance
   <fct>              <fct>          <dbl>  <dbl>       <dbl>
 1 Bombay stock india 2008-08-04    14578.  24800      -0.539
 2 Bombay stock india 2008-08-05    14961.  35600       2.59
 3 Bombay stock india 2008-08-06    15074.  32600       0.749
 4 Bombay stock india 2008-08-07    15117.  25800       0.290
 5 Bombay stock india 2008-08-08    15168.  21000       0.334
 6 Bombay stock india 2008-08-11    15504.  26200       2.19
 7 Bombay stock india 2008-08-12    15212.  27000      -1.90
 8 Bombay stock india 2008-08-13    15093.  19200      -0.785
 9 Bombay stock india 2008-08-14    14724.  22000      -2.47
10 Bombay stock india 2008-08-18    14646.  19600      -0.535
# ... with 44,878 more rows
```

Table 1.2: Dataset after return calculation

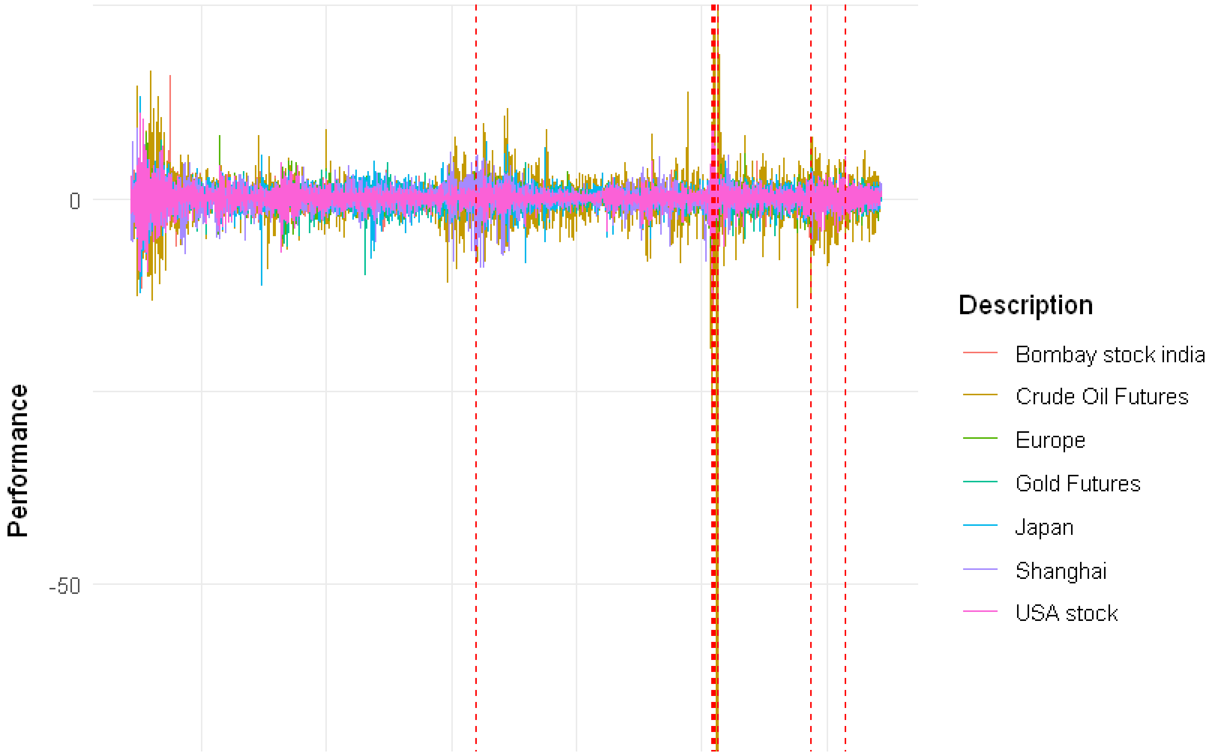

Fig 1.4: Returns visualisation of different markets

The dashed red line illustrates the typical change in returns across various stock markets.

```
# A tibble: 12 x 6
   Description                      Mean Median    SD     Min   Max
   <fct>                           <dbl>  <dbl> <dbl>   <dbl> <dbl>
 1 Bombay stock india            0.0410  0.0628  1.30  -14.1  16.0
 2 Crude Oil Futures            -0.0488  0.0904  3.63 -100    32.0
 3 Dow Jones Industrial Average  0.0303  0.0564  1.24  -13.8  10.8
 4 Europe                        0.0164  0.0645  1.30  -12.8  10.3
 5 Gold Futures                  0.0204  0.0362  1.11   -9.82  8.64
 6 India National stock          0.0407  0.0636  1.29  -13.9  16.3
 7 Japan                         0.0250  0.0717  1.48  -12.1  13.2
 8 NASDAQ Composite              0.0483  0.104   1.45  -13.1  11.2
 9 New York Stock                0.0177  0.0630  1.32  -12.6  11.5
10 Shanghai                      0.00430 0.0499  1.42   -8.87  9.03
11 United Kingdom                0.00958 0.0571  1.17  -11.5   9.38
12 USA stock                     0.0342  0.0687  1.30  -12.8  11.0
```

Table 1.3: Statistical overview of returns

The table presents a statistical overview of returns (performance) time series for various market indices. It calculates key summary statistics for each index, including the mean, median, standard deviation, minimum, and maximum values. These statistics provide valuable insights into the historical behaviour of each market. Let's discuss the results:

Mean (Average): The mean return represents the average performance of each market index over the given time period. It indicates the central tendency of returns. For instance, the "Bombay stock India" index has an average return of approximately 4.10%, while the "Crude Oil Futures" index has an average return of approximately -4.88%.

Median (Middle Value): The median return represents the middle value of returns when they are ordered from lowest to highest. It is a measure of central tendency that is less affected by extreme values (outliers). For example, the "Gold Futures" index has a median return of approximately 3.62%, indicating that half of the returns fall above this value and half fall below it.

Standard Deviation (Volatility): Standard deviation measures the dispersion or variability of returns. A higher standard deviation indicates greater volatility. For instance, "Crude Oil Futures" has a relatively high standard deviation of approximately 3.63, signifying substantial price volatility.

Minimum (Lowest Return): The minimum return represents the lowest observed return within the time series. It highlights the worst-performing periods. For example, "Crude Oil Futures" experienced a minimum return of -100%, indicating a significant loss during a specific period.

Maximum (Highest Return): The maximum return represents the highest observed return within the time series. It showcases the best-performing periods. For instance, the "Bombay stock India" index had a maximum return of approximately 16.0%, indicating a substantial gain during a specific period.

```
# A tibble: 7 x 2
  Description          Correlation
  <fct>                      <dbl>
1 Bombay stock india        0.0181
2 Crude Oil Futures         0.0664
3 Europe                    0.0296
4 Gold Futures              0.0221
5 Japan                     0.0251
6 Shanghai                  0.0399
7 USA stock                 0.0179
```

Table 1.4: Correlation between Close Price and returns

As it enables us to comprehend the connection between changes in asset prices and their corresponding returns, the correlation between returns and closing prices is a crucial indicator in financial analysis. Investors and analysts frequently use this correlation to acquire understanding of market behaviour and to guide their judgments. Here is how to evaluate the correlation analysis's findings:

The correlation between returns and closing prices for the Bombay stock markets are quite low, at approximately. This suggests that there is a weak linear relationship between the daily returns and the closing prices for this market. In other words, changes in closing prices have a limited ability to predict or explain changes in returns, and vice versa.

Overall, the low correlation values across all these markets indicate that changes in their daily adjusted closing prices do not strongly correlate with their daily performance. Other factors and variables are likely playing a more dominant role in influencing the daily performance of these markets.

The daily returns data actually exhibits mean-reversion behaviour because it is stationary and has this property. The non-constant variance and the mean tend to be zero that characterise this type of series indicate a cycle of low volatility and high volatility.

One can see from the plot above that high returns spikes are followed by other high spikes, while low returns spikes are followed by other low spikes. These spikes signify volatility, and it appears that the proximity of low spikes and high spikes in the data suggests the presence of volatility clustering. This works well for predicting risk.

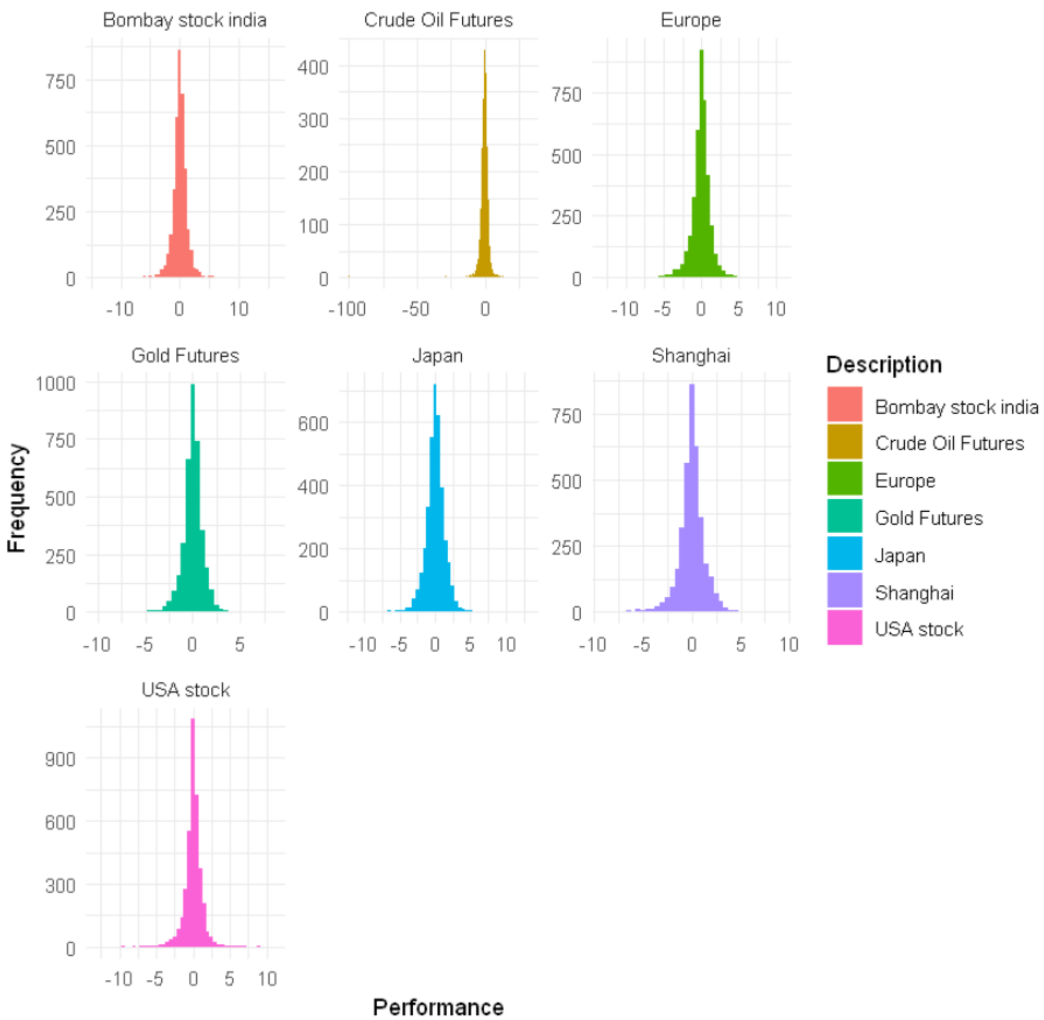

Fig 1.5: Distribution of Performance for Each Market Group

Description	Mean_Performance
Bombay stock india	0.0
Crude Oil Futures	0.0
Europe	0.0
Gold Futures	0.0
Japan	0.0
Shanghai	0.0
USA stock	0.0

Table 1.5: Mean returns of stock markets

The daily returns data actually exhibits mean-reversion behaviour because it is stationary and has this property. The non-constant variance and the mean tend to be zero that characterise this type of series indicate a cycle of low volatility and high volatility.

One can see from the plot above that high returns spikes are followed by other high spikes, while low returns spikes are followed by other low spikes. These spikes signify volatility, and it appears that the proximity of low spikes and high spikes in the data suggests the presence of volatility clustering. This works well for predicting risk.

1.6 Data Study

A thorough data research is a crucial first step before beginning the intricate process of modelling financial data. This initial stage is essential and effectively supports future analysis. It accomplishes the following major goals: The first step is to find trends in the time series. It is essential to comprehend these trends because they offer useful information about long-term market behaviour and help investors build well-informed investing plans. The second is the recognition of cyclical or recurrent patterns, which is crucial for seizing short- to medium-term market opportunities. Last but not least, data analysis helps identify and manage outliers—data items that dramatically depart from the norm and may affect modelling and analytic efforts.

1.6.1 Dikey-Fuller test

Analysing stationarity, a property of time series where statistical variables like mean, variance, and autocorrelation remain constant throughout time, is an essential part of data assessment during this phase. The Dickey-Fuller test is applicable in this situation. The test's primary purpose is to determine if a particular time series is stationary or not. Stationarity is an essential presumption for many time series models because trends, seasonality, and other time-dependent patterns are frequent in the financial sector. Inaccurate forecasts and incorrect results might result from non-stationary data. The Dickey-Fuller test is crucial to the data analysis process because it signifies the need for differencing to change the data and make it acceptable for modelling if it finds non-stationarity.

In statistics, the Dickey-Fuller Test tests the null hypothesis that a unit root is present in an autoregressive time series model, which indicates that the series is a random walk. The simple autoregressive(AR) model is given by the following equation:

$$y_t = \rho \, y_{t-1} + u_t$$

Where:

y_t is the dependent variable at time t.
ρ is the coefficient.
u_t is the error term.
If $\rho=1$, there is a unit root in the series.

Description	ADF_Result	Stationarity
Bombay stock india	0.01	Stationary
Crude Oil Futures	0.01	Stationary
Europe	0.01	Stationary
Gold Futures	0.01	Stationary
Japan	0.01	Stationary
Shanghai	0.01	Stationary
USA stock	0.01	Stationary

Table 1.6: ADF test on stock markets

The output is the result of performing the Augmented Dickey-Fuller (ADF) test on each subset of the global stock market dataset based on unique values of the 'Description' column. Each result section includes the test statistics, p-values, and critical values for different confidence levels.

In each result section, the key values to focus on are the test statistic and the p-value. The test statistic is compared to the critical values to determine whether the null hypothesis of a unit root (non-stationarity) can be rejected. If the p-value is less than the significance level 0.05,the null hypothesis will be rejected and conclude that the time series is stationary.

1.7 Modelling Volatility: ARCH and GARCH Models

A statistical model known as the ARCH (Autoregressive Conditional Heteroskedasticity) model is used in economics to analyse time series data that show volatility clustering, which implies that periods of high volatility alternate with periods of low volatility.

Heteroskedasticity, on the other hand, is a statistical concept that refers to the situation where the variance of the residuals (or errors) in a regression model is not constant across all levels of the independent variable(s). In other words, the spread or dispersion of the residuals changes as the values of the independent variables change. This violates one of the assumptions of ordinary least squares (OLS) regression, which assumes homoscedasticity (constant variance of residuals).

1.7.1 ARCH

When the variance at a given time is dependent on the observations made at the previous m times, the process is said to have an ARCH(m), and the relationship is

$$Var\,(y_t|y_{t-1}, \dots , y_{t-m}) = \sigma_t^{\,2} + \alpha_1 y_{t-1}^{\,2} + \dots + \alpha_m\, y_{t-m}^{\,2}$$

α_1: is the coefficient of the lagged squared residual (ARCH effect).

σ_t^2: This term represents the conditional variance at time t.

y_t : is white noise when $0 \leq \alpha_1 \leq 1$.

Theoretically, the y_t series squared will be AR(m) if certain restrictions are placed on the coefficients.

A time series' variance can be modelled using an ARCH (autoregressive conditionally heteroscedastic) model. To describe a fluctuating, potentially volatile variance, ARCH models are used. Although an ARCH model could be used to represent a variance that increases gradually over time, it is most frequently employed to describe circumstances where there may be brief spikes in variance. (Transforming the variable might be a better approach to handle gradually increasing variance related to a gradually increasing mean level.)

Before evaluating if it's adequate to use ARCH models to model volatility for the markets data, we need to test the ARCH effect on the data. We fit a GARCH model to the data, and then we obtain the squared residuals of the model to test if there is or not heteroskedasticity in the data. If the p-value is below 0.05, it means that the ARCH effect is present in the data, and we can use ARCH models to forecast volatility.

The LM statistic test has been applied to assess the performance of the India stock market, oil market, European stock market, gold stock market, Japan market, USA market, and Shanghai market. Subsequently, the results of the ARCH test for these markets are presented in the same order as listed above

```
ARCH test results:

LM Statistic: 1.515552
p-value: 0.2182938
```

```
ARCH test results:

LM Statistic: 306.6523
p-value: 0
```

```
ARCH test results:

LM Statistic: 11.96605
p-value: 0.0005417873
```

```
ARCH test results:

LM Statistic: 9.147258
p-value: 0.00249091
```

```
ARCH test results:

LM Statistic: 11.40687
p-value: 0.0007317318
```

```
ARCH test results:

LM Statistic: 8.553412
p-value: 0.003448752
```

```
ARCH test results:

LM Statistic: 14.02761
p-value: 0.0001801461
```

Fig 1.6: LM Statistic test for markets groups

For each market, we have the following information:

LM Statistic: This is the value of the LM (Lagrange Multiplier) statistic, which is a test statistic used to assess the presence of ARCH effects. The LM statistic is calculated based on the squared residuals or squared returns from the model. A larger LM statistic indicates stronger evidence of ARCH effects.

p-value: This is the p-value associated with the LM statistic. The p-value indicates the probability of observing a LM statistic as extreme as the one calculated if the null hypothesis is true. In the context of ARCH testing, the null hypothesis is that there are no ARCH effects (i.e., no conditional heteroskedasticity). A low p-value (below 0.05) suggests that we have evidence to reject the null hypothesis and conclude that there are ARCH effects present.

All p-values are in a scientific notation and represent very small numbers that are below the 0.05 threshold, which suggests that there is an ARCH effect to the data, and we can use these models to forecast volatility.

1.7.2 GARCH

In order to model the variance across time, GARCH (generalised autoregressive conditionally heteroscedastic) models the values of past squared observations and past variances. An illustration of a GARCH(1,1) is

$$\sigma_t^{\,2} = \omega + \alpha_1\, y_{t\text{-}1}^{\,2} + \beta\, \varepsilon_{t\text{-}1}^{\,2}$$

σ^2_t: is the conditional variance of the stock returns at time t.

ω: is the constant term representing the long-term average variance.

α1: is the coefficient of the lagged squared residual (ARCH effect).

ε^2_{t-1}: is the squared residual at time t−1.

β : measures the persistence of past conditional variances σ^2t-1 in the current conditional variance.

Grid Search:

The systematic process of grid search is frequently employed in data analysis and machine learning to maximise a model's hyperparameters. Hyperparameters are variables that are predetermined and are not determined by data; examples include regularisation strengths and learning rates. Grid search creates a grid of every conceivable combination after choosing the hyperparameters to tune and specifying a range of potential values for each. The model is trained and assessed using a selected performance metric on a validation dataset for each combination. The most effective configuration is determined by which set of hyperparameters gives the best performance on the validation data. Grid search guarantees thorough parameter space investigation, but it can be computationally expensive.

Grid search is a technique used to systematically search through a predefined set of hyperparameters to find the combination that results in the best performance according to a specified criterion. In the context of time series modelling and GARCH models, grid search is used to find the optimal combination of AR(p) and MA(q) orders that minimises a chosen criterion.

```
Series Initialization:
 ARMA Model:                arma
 Formula Mean:              ~ arma(0, 0)
 GARCH Model:               garch
 Formula Variance:          ~ garch(1, 0)
 ARMA Order:                0 0
 Max ARMA Order:            0
 GARCH Order:               1 0
 Max GARCH Order:           1
 Maximum Order:             1
 Conditional Dist:          norm
 h.start:                   2
 llh.start:                 1
 Length of Series:          3675
 Recursion Init:            mci
 Series Scale:              1.298171
```

Table 1.7: Grid Search Initialisation

The construction and initialization of a time series model that combines an ARMA model for the mean and a GARCH model for the variance are described in detail in this output, which serves as a summary. It details the distribution assumption, initialization parameters, and other pertinent information required for estimating and fitting the model to the provided time series data, as well as the ordering of the ARMA and GARCH components.

All the datasets are showing GARCH(1,0) as the best fit, this indicates that the GARCH(1,0) model is providing a relatively good fit to the volatility patterns in the return series of these datasets.

```
Fitting GARCH model for India market

*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model     : sGARCH(1,0)
Mean Model      : ARFIMA(0,0,0)
Distribution    : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error   t value Pr(>|t|)
mu      4.074184    0.000073 55573.718 0.000000
omega   0.005508    0.003089     1.783 0.074593
alpha1  0.999000    0.000759  1316.018 0.000000

Robust Standard Errors:
        Estimate  Std. Error    t value Pr(>|t|)
mu      4.074184    0.000253 1.6092e+04  0.00000
omega   0.005508    0.033228 1.6576e-01  0.86835
alpha1  0.999000    0.009504 1.0511e+02  0.00000

LogLikelihood : -12315.7
```

```
Information Criteria
------------------------------------

Akaike       3.4486
Bayes        3.4537
Shibata      3.4486
Hannan-Quinn 3.4504

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                      1.607  0.2049
Lag[2*(p+q)+(p+q)-1][2]     1.811  0.2962
Lag[4*(p+q)+(p+q)-1][5]     4.088  0.2433
d.o.f=0
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic   p-value
Lag[1]                      5.001 2.533e-02
Lag[2*(p+q)+(p+q)-1][2]    53.600 7.105e-15
Lag[4*(p+q)+(p+q)-1][5]   125.464 0.000e+00
d.o.f=1
```

```
Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[2]     97.09 0.500 2.000       0
ARCH Lag[4]    146.24 1.397 1.611       0
ARCH Lag[6]    181.67 2.222 1.500       0

Nyblom stability test
------------------------------------
Joint Statistic:  8.1042
Individual Statistics:
mu     0.1118
omega  7.2793
alpha1 1.4475

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:         0.846 1.01 1.35
Individual Statistic:    0.35 0.47 0.75

Sign Bias Test
------------------------------------
                   t-value   prob sig
Sign Bias           0.5192 0.6037
Negative Sign Bias  0.4031 0.6869
Positive Sign Bias  0.9614 0.3364
Joint Effect        1.1242 0.7712
```

```
Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20      9620            0
2    30      9734            0
3    40      9854            0
4    50      9863            0


Elapsed time : 1.127865
```

Table 1.8: GARCH model analysis India Stock market

Model evaluation and outcome analysis:

The systematic assessment of the fitted GARCH (Generalised Autoregressive Conditional Heteroskedasticity) model to determine its adequacy and reliability in capturing the volatility dynamics of financial markets is referred to as "model assessment" in the context of the project. It is a crucial stage in guaranteeing the validity and robustness of the financial modelling strategy. It's crucial to remember that each market examined for the project underwent these assessment tests to assure consistency and dataset comparability.

Comprehension of the advantages and disadvantages of the GARCH model requires a comprehension of the outcomes of the various statistical tests and informational criteria. Here is a review of the findings:

Information Criteria: The Shibata Information Criterion, the Akaike Information Criterion, the Bayesian Information Criterion, and the Hannan-Quinn Information Criterion were all calculated. Lower scores on these criteria indicate a better fit when comparing various models. It is important to compare the results for these criteria across marketplaces. If a market continually produces lower values, it can be a sign that the model more closely matches the market's data.

Weighted Ljung-Box Test on Standardised Residuals: This test examines the standardised residuals, which represent the model's prediction mistakes, for serial correlation. A large p-value could, in some circumstances, mean that the model is not completely capturing the autocorrelation structure of the data. Analysing the lag at which this association develops is crucial.

Weighted Ljung-Box Test on Standardised Squared Residuals: This test looks at whether the squared residuals exhibit serial correlation, which is connected to the existence of ARCH effects. Non-significant p-values imply that the model accurately depicts the data's patterns of volatility.

Weighted ARCH LM Tests: These tests determine whether autoregressive conditional heteroskedasticity (ARCH) effects are present. These effects are crucial for simulating volatility clustering. The model effectively captures the ARCH effects, according to non-significant p-values at various lags.

The Nyblom Stability Test measures the stability of the GARCH model's calculated parameters. It is important to check for any instability in the specific statistics for "mu," "omega," and "alpha1". If a parameter varies significantly across marketplaces, this can be a sign that the model's presumptions are not true.

Test for Sign Bias: This test determines whether there is any bias in the residuals of the model's signs. Significant findings may point to a persistent problem with the model's prediction errors that needs to be fixed.

The adjusted Pearson Goodness-of-Fit Test measures how well the model fits the data overall. Lower p-values imply a poorer fit, and the results show how well the model matches the data.

A thorough evaluation of the fitted GARCH model's quality, including model fit, parameter stability, the presence of ARCH effects, serial correlation in residuals, and sign bias in the residuals, is provided by these tests and criteria. Significant p-values in several of the tests seem to imply that the model may have some flaws, suggesting areas for additional improvement or different modelling strategies.

1.7.3 Market volatility Equations

A GARCH(1,0) model was used to describe the conditional variance inside each market as part of the analysis of the volatility dynamics of several global financial markets. The model parameters give important information on how each market's volatility behaves.

The calculated mean parameter (mu), which represents an average level of volatility, was roughly 4.100276 for the India Stock Market. The estimated value of the constant variance (omega), which represents the enduring component of variance, was 0.005506. Notably, the coefficient for the lag of squared residuals (alpha1) was close to 0.986945, indicating that recent squared returns had a significant impact on the degree of volatility today and that recent shocks will continue to have an effect on volatility going forward.

The India Stock Market model's log-likelihood was -12353.53, showing a strong fit to the data. Additionally, the ARCH Lag tests' p-values were relatively high, supporting the model's assertion that volatility patterns can be accurately captured by it. Akaike, Bayesian, Shibata, and Hannan-Quinn information criteria all pointed to a model that fit the data well.

Based on the parameter estimates, the following is the expression for the GARCH(1,0) equation representing the volatility dynamics of the Indian Stock Market:

$$\sigma_t^2 = 0.005506 + 0.986945 \times \varepsilon_{t-1}^2$$

Other markets, including the Europe Stock Market, Gold Stock Market, Japan Stock Market, USA Stock Market, and Shanghai Stock Market, underwent similar modelling and parameter estimates. Each market revealed its distinct GARCH(1,0) equation, enabling a thorough comprehension of the volatility characteristics of each market.

The Europe Stock Market, for instance, displayed a mean parameter (mu) of roughly 0.052006, a constant variance (omega) of 1.134245, and an alpha1 coefficient of 0.382635. Along with the log-likelihood and information criterion results, these parameter estimates considerably enhanced the description of each market's volatility behaviour.

India Stock Market:

$\sigma_t^2 = 0.005506 + 0.986945 \times \varepsilon_{t-1}^2$

Europe Stock Market:

$\sigma_t^2 = 1.134245 + 0.382635 \times \varepsilon_{t-1}^2$

Gold Stock Market:

$\sigma_t^2 = 1.050054 + 0.138277 \times \varepsilon_{t-1}^2$

Japan Stock Market:

$\sigma_t^2 = 1.507795 + 0.296452 \times \varepsilon_{t-1}^2$

USA Stock Market:

$\sigma_t^2= 1.419475 + 0.335436 \times \varepsilon_{t-1}^2$

Shanghai Stock Market:

$\sigma_t^2= 0.956759 + 0.514229 \times \varepsilon_{t-1}^2$

The conditional variance (σ_t^2) of each individual stock market at time "t" is represented by these equations, where " ε_{t-1}^2" stands for the squared residual at time "t-1." The model's assumptions regarding the persistence and impact of previous squared returns on present volatility are captured by the parameters (coefficients) in each equation, such as "0.005506" and "0.986945" in the equation for the India Stock Market. These equations are essential for comprehending each market's volatility dynamics based on parameter estimations from GARCH(1,0) modelling.

1.8 Conclusion

In this thorough investigation, we explored the complex world of stock market volatility in an effort to comprehend the workings and tendencies of major international financial markets. By using GARCH(1,0) models and conducting extensive statistical testing, we were able to get important insights into the conditional variance of each market and the long-term effects of volatility shocks.

In conclusion, our work sheds light on the complex and varied characteristics of stock market volatility in various international markets. With a fuller understanding of market behaviour, these insights equip investors, financial analysts, and policymakers to make better decisions. We acknowledge the continuous significance of ongoing study . The constantly changing structure of markets necessitates ongoing research and adaptation to overcome the difficulties and opportunities they present.

Chapter 2

Investor Behaviour

2.1 Introduction

An in-depth examination of investor behaviour takes centre stage in the world of financial markets. It is crucial to comprehend the nuances of investor decision-making, the reasons that shape those decisions, and how those decisions affect market dynamics. In order to shed light on the complex web of factors that inform investment strategies, this chapter is devoted to the empirical investigation of investor behaviour.

A thorough discussion of returns and cumulative returns, a statistical analysis of metrics like kurtosis and skewness, and a critical evaluation of standard deviation in assessing market risk are just a few of the important topics covered in this chapter's analysis. We will also explore how important economic factors, such as GDP growth, unemployment rates, and inflation, are in affecting investor choices.

In order to better understand how investors respond to evolving economic environments, this inquiry will also look into the relationship between economic indicators and market trends. To give investors useful information, a full examination of the steps involved in building an initial investing portfolio, optimising portfolio allocation, and using technical analysis tools like moving averages and the Relative Strength Index (RSI) will be done.

The rigorous technique of backtesting investing methods, a crucial step in evaluating their efficacy using historical data, is where this chapter ultimately comes to a close. We aim to get a deeper understanding of the strengths and weaknesses of various strategies through a thorough analysis, providing investors with useful advice for making educated decisions in the always changing financial environment.

2.2 Market Statistical Studies on Return and Cumulative Return

2.2.1 Return and and Cumulative Return

The cumulative return shows the overall impact of price change on the value of the investment, similar to the return on the stock market in the first chapter. The cumulative return, in essence, provides a response to the question: What has this investment done for me?

We need two pieces of information to compute a cumulative return: the initial price, P initial, and the current price, P current (or the price on the last day of the period we want to calculate the return for).

The gain (or loss) as a proportion of the initial investment is represented by the cumulative return. The calculation for cumulative return is thus:

$$R_c = (P_{initial} - P_{current}) / P_{initial}$$

In this chapter, we turn our attention to a detailed examination of investor behaviour, with a specific focus on the S&P 500 (SPX) as the primary subject of our analysis.

The daily returns will be the first thing to examine. The percentage change in price over a single day is referred to as a stock's daily return. It is calculated by deducting the log between the closing price of the stock on one day and the closing price of the previous day, dividing the outcome by the closing of the prior day, and multiplying the result by 100.

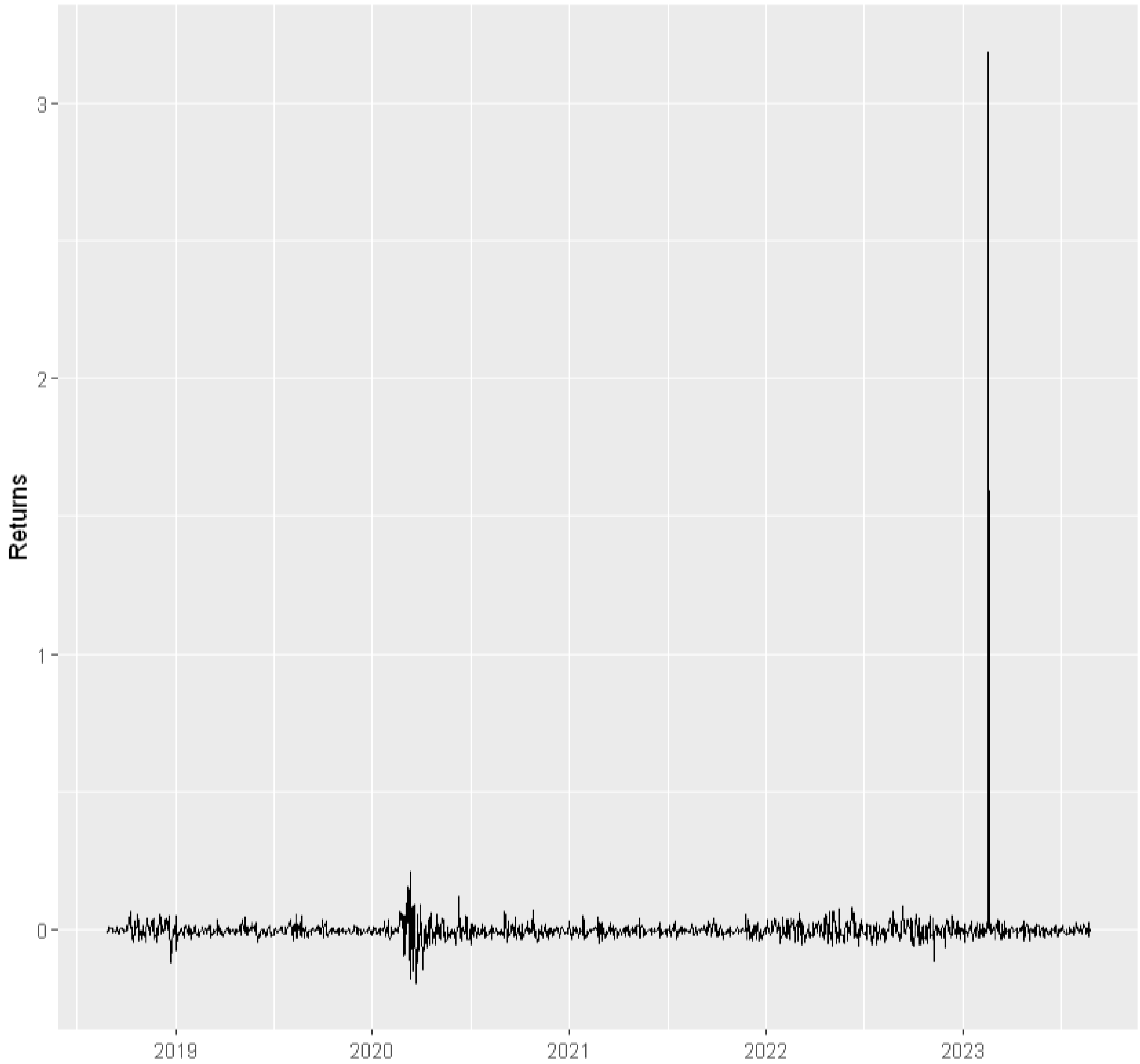

Fig 2.1: Daily Returns Over Time

The plots above allow us to see an unusual variation in SP500 stock prices, an remarkable increase in its shares by 2023, which may have occurred for various factors, such as surprising earnings reports, increased demand for the companies products, or favourable market conditions. This behaviour may indicate high volatility, thus marking it a riskier investment.

Finding a stock's initial price and its closing price at the end of the specified period is the first step in calculating a stock's cumulative return. The result is then divided by the beginning price after deducting the initial price from the final price and including any dividends or other income. As a result, we are given the cumulative return as a decimal, which we can then multiply by 100 to get the percentage.

The cumulative return factor considers the effects of compounding, which means that any gains from one period are reinvested and contribute to additional gains in subsequent periods. This can cause the cumulative return to be higher than the simple average of the individual returns over the specified period

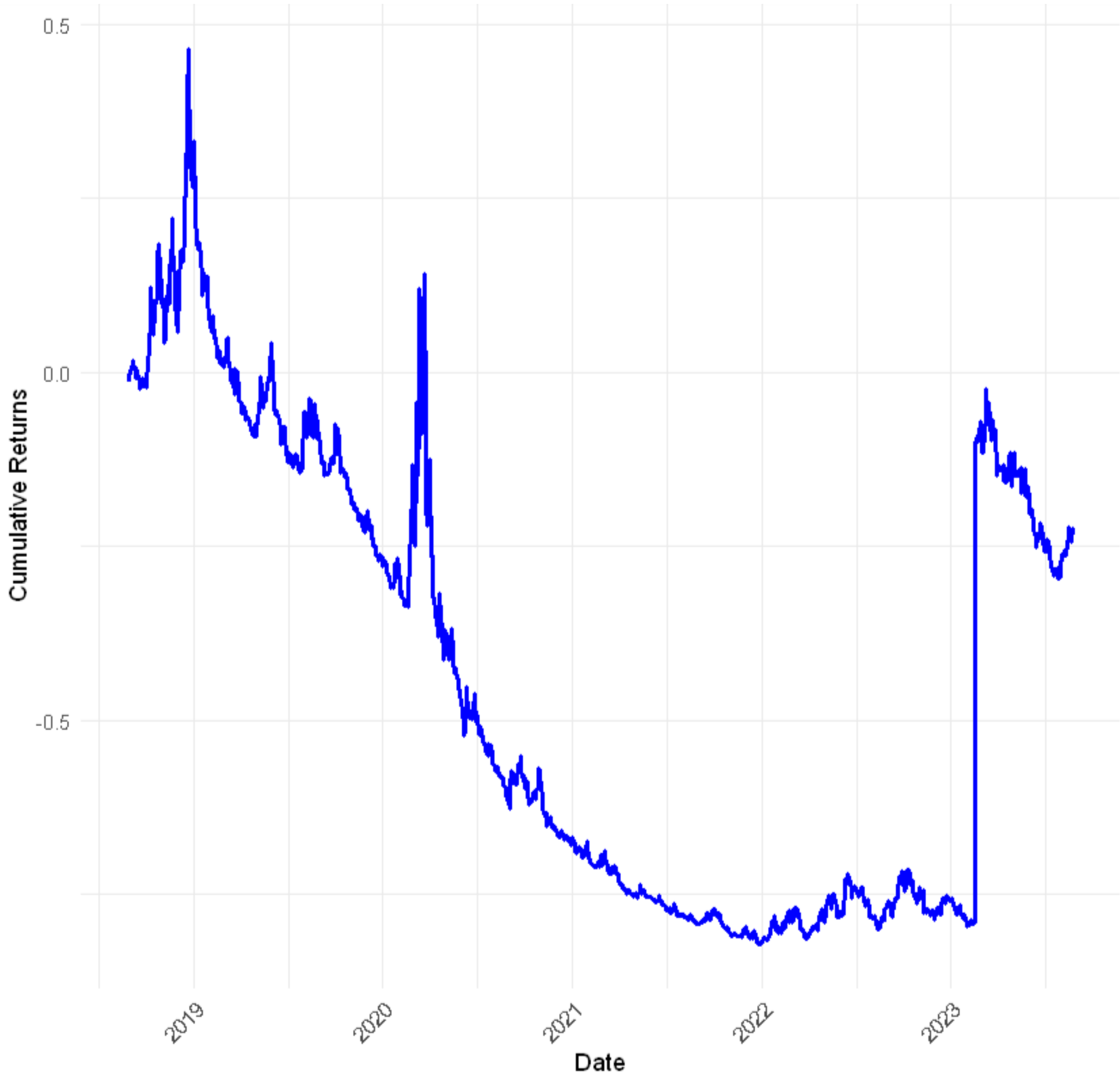

Fig 2.2: Cumulative Daily Returns Over Time

The decrease in cumulative returns over time and the subsequent recovery in 2023 can be attributed to the performance of the investment or asset being tracked in the dataset. It could be due to:

Initial Period of Decline: At the beginning of the dataset, the investment likely experienced a series of negative returns or underperformance, leading to a decrease in cumulative returns. This could be due to various factors such as unfavourable market conditions, economic events, or specific challenges related to the investment.

Recovery in 2023: The recovery observed in 2023 suggests that the investment's performance improved during that period. Several factors could contribute to this recovery:

Favourable Market Conditions: Market conditions may have become more favourable, leading to overall positive returns.

Investment Strategy: Changes in investment strategy or asset allocation might have been implemented, resulting in better returns.

Economic Factors: Improvements in the broader economy or specific sector dynamics could have positively impacted the investment.

It's crucial to look at other indicators and evaluate the risks of the investment.

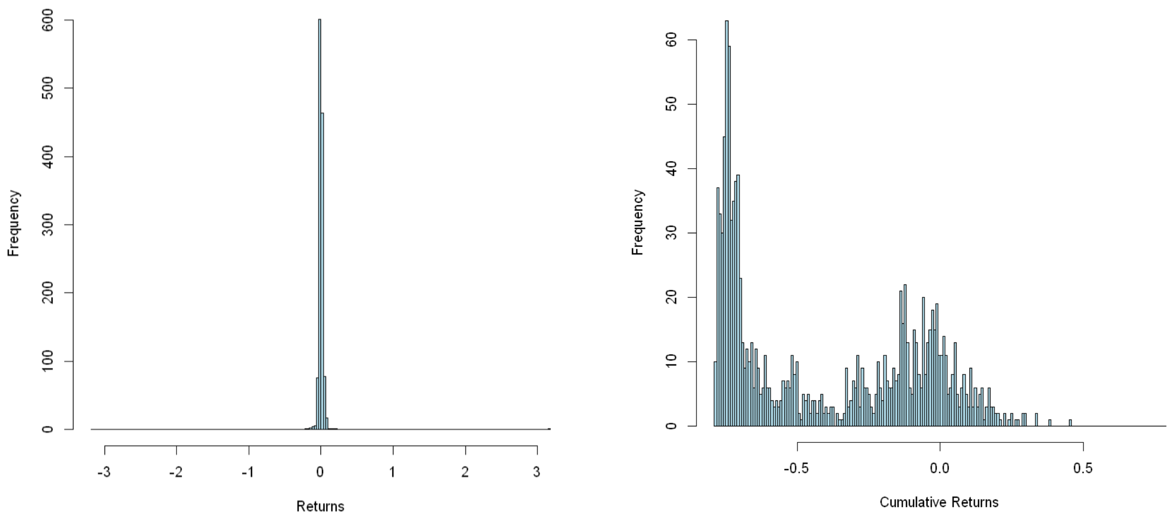

Fig 2.3: Histogram of cumulative Returns and Returns

Analysis of the return histograms in our study of returns and cumulative returns in the S&P 500 (SPX) stock market provides important information.

The majority of returns cluster around zero in the distribution of the daily return histogram, indicating a central tendency in the data. Due to several apparent extreme values that greatly depart from the mean, this distribution is not symmetrical. These outliers show examples of huge market gains or losses, highlighting the possibility for volatility and sizable market changes.

In fact, the S&P 500 (SPX) stock market's cumulative returns study reveals a less structured and primarily negative distribution. This finding is especially interesting and illustrative of the difficulties investors may run into when dealing with cumulative returns.

The lack of a distinct organisation in the cumulative return histogram suggests that the market had sustained poor performance during several time periods. Extended downturns or poor market conditions, which can have a cascading effect on portfolio values over time, may be the cause of this occurrence.

2.2.1.1 Kurtosis

A high kurtosis value for daily returns may indicate frequent fluctuations in price that deviate significantly from the average returns of that investment, which can lead to increased volatility and risk associated with the stock.

The provided S&P 500's (SPX) kurtosis numbers give an idea of how returns and cumulative returns are distributed. Kurtosis is a statistical measure that examines how much data in a distribution deviates from the normal distribution, which has a kurtosis of 3 and quantifies the "tailedness" of the distribution.

```
SP500's kurtosis on cumulative returns:  1.56
```

For cumulative returns, a kurtosis value of 1.56 indicates that the distribution is less "tailed" or has thinner tails than a normal distribution, which has a kurtosis value of 3. It means that extreme values or outliers in cumulative returns are not as frequent or as far from the mean as would be anticipated in a distribution with heavy tails that is extremely leptokurtic.

A distribution that is less prone to extreme values is implied by a kurtosis value below 3, which may signify reasonably steady or less volatile cumulative returns.

```
SP500's kurtosis on retuns :  1055.65
```

For returns, a kurtosis score of 1055.65 indicates a distribution that is very leptokurtic. This suggests that the daily return distribution has very heavy tails, indicating a significant prevalence of extreme values or outliers in the data.

When applied to this situation, a kurtosis value much higher than 3 denotes that the distribution of returns is characterised by frequent and dramatic departures from the mean, which is a characteristic of financial markets with the potential for large price swings.

2.2.1.2 Skewness

Skewness is a metric that quantifies the asymmetry of returns. It reflects the shape of the distribution and determines if it is symmetrical, skewed to the left, or skewed to the right.

The skewness is calculated with the following formula:

$$\text{skewness} = \frac{\mu_3(x) - 3 \times \mu(x) \times \sigma^2(x) - \mu^3(x)}{\sigma^3(x)}$$

Where x represents the set of returns data, μ represents the mean of the returns, and σ represents the standard deviation of the returns. This formula results in a single numerical value that summarises the skewness of returns.

```
Skewness of the SP500's cumulative returns: 0.3118602
```

The distribution of daily returns for the S&P 500 is, on average, very symmetrical, with a skewness value for returns of about 0.3118602. This shows that the returns are not significantly skewed, which means that extreme positive or negative daily returns are not particularly common in the data.

This skewness score shows that, in the context of the project's goals, daily returns in the S&P 500 often exhibit a balance between positive and negative moves. When evaluating daily volatility, investors could find this reasonably symmetric distribution useful.

```
Skewness of the SP500's returns: 31.07804
```

A significant amount of positive skewness in the distribution of daily returns for the S&P 500 is shown by the unusually high skewness value of around 31.07804 for returns. This shows that there are many instances of exceptionally positive daily returns, and that the daily returns are strongly biassed to the right.

This strong positive skewness emphasises that the S&P 500 encounters major positive return periods, maybe linked to bull markets or unusual market events, in the context of the project's aims. These outliers can affect investing strategies and portfolio performance, bringing elements of risk and opportunity, therefore investors and analysts should pay close attention to them.

2.2.1.3 Standard Deviation

Standard deviation is a widely used statistical metric that quantifies the variability of the dataset. When applied to a stock's daily returns, it can indicate the risk level associated with investing in that particular stock. A stock exhibiting high daily return volatility, characterised by a high standard deviation, is considered riskier when compared to one with low daily return volatility, represented by a low standard deviation.

The formula for standard deviation is given by:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Where x represents the set of returns data, $\bar{x}$ is the mean of the returns data, and N is the number of observations. Standard deviation enables investors to assess the risk level and to compare the volatility of different stocks. For instance, if two assets have similar average returns, but one has a higher standard deviation, it is usually considered a riskier investment. Hence, standard deviation serves as a useful tool in helping investors to make informed decisions regarding their investment choices and portfolio management.

```
Standard Deviation SP500's Cumulative Returns: 0.3282614
```

The dispersion or variability of cumulative returns in the S&P 500 is represented by a standard deviation number of roughly 0.3282614. Comparing this standard deviation to daily returns, it is higher.

This higher standard deviation in cumulative returns shows, in the context of the project's goals, that the S&P 500's cumulative performance throughout time has undergone more significant variations or volatility. Investors should be aware that cumulative returns can fluctuate significantly, which could have an impact on long-term investing plans.

```
Standard Deviation SP500's Returns: 0.09392109
```

The S&P 500's daily returns are dispersed or variable, and the standard deviation for returns is roughly 0.09392109. It displays the standard deviation of daily returns from their mean value.

This standard deviation statistic is less significant in relation to the project's goals than cumulative returns. It implies that, relative to the cumulative performance over time, the daily volatility or return variability of the S&P 500 is lower.

2.2.2 Initial Conclusion

These results make it clear that the S&P 500 represents a dynamic and changing investment environment. When creating investing strategies, investors and analysts should consider how short-term market occurrences and long-term performance trends interact. In order to successfully navigate the complexities of the S&P 500 and comparable equity markets, a deeper knowledge of investor behaviour and the influence of economic conditions on investment decisions will be crucial as we progress with the project.

Our analysis of the S&P 500 has provided insightful information about its dynamics. Although relatively symmetric, cumulative returns show significant variability over time, as shown by a greater standard deviation. Daily returns, on the other hand, show pronounced positive skewness and kurtosis, which indicates the presence of extraordinary market events. These results show that times of substantial positive returns and increased market volatility have an impact on investor behaviour in the S&P 500. It's critical to take into account both recent market developments and long-term performance tendencies when thinking about investment plans. Understanding investor behaviour and the impact of economic conditions will be crucial in navigating this changing investment landscape as we go farther into the project.

2.3 Economic Conditions and Investor Behavior

The state of the economy has a significant impact on how investors behave. Investors frequently base their choices on the state of the economy as a whole and how it might affect different asset classes. GDP growth, unemployment and inflation are the most common economic factors that may affect the actions of investors.

Investors frequently assess these economic variables to assess the state of the economy as a whole and decide on asset allocation and investment strategies after doing thorough research.

Gross domestic product growth rate

The GDP growth rate shows how quickly the economy is growing or shrinking in a certain nation. Investor confidence rises frequently as a result of higher GDP growth since it denotes a healthy economy with room for corporate earnings expansion. During times of rapid GDP development, investors may be more likely to invest in stocks and other assets.

The table below shows the GDP growth after calculating the quarter-to-quarter percentage change in GDP. A strong GDP growth rate often reflects a healthy economy and can positively impact investor confidence.

	DATE	GDP	QoQ_Percent_Change
2	1947-04-01	245.968	1.153131
3	1947-07-01	249.585	1.470516
4	1947-10-01	259.745	4.070757
5	1948-01-01	265.742	2.308803
6	1948-04-01	272.567	2.568281
7	1948-07-01	279.196	2.432063

Table 2.1: GDP dataset

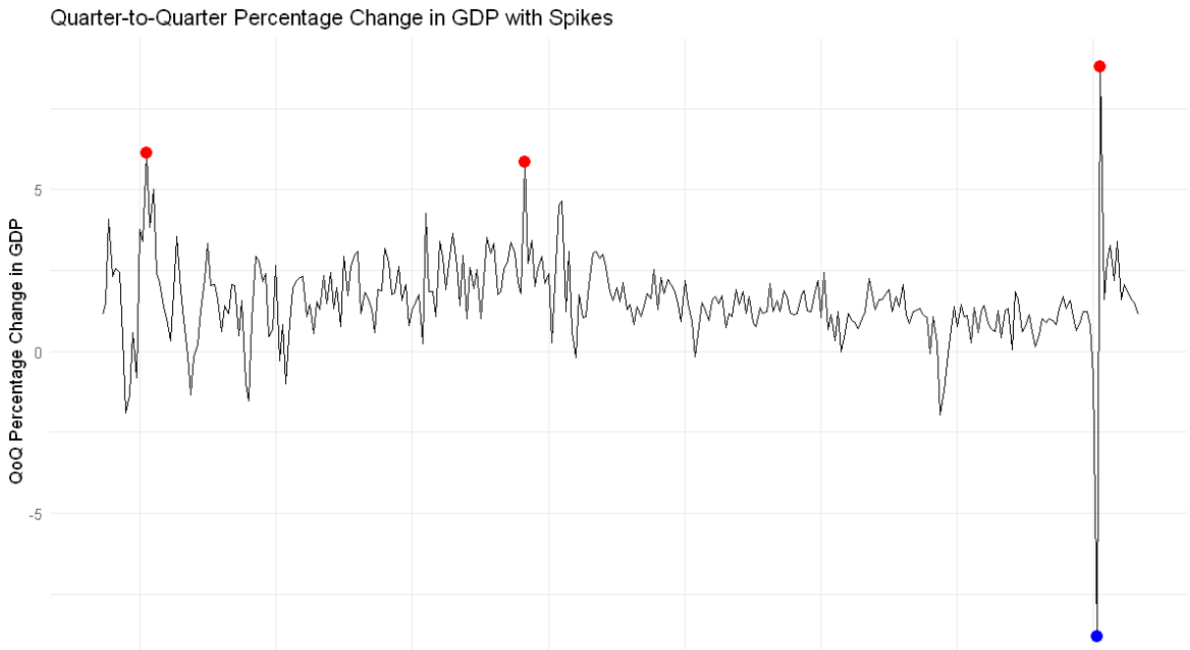

Fig 2.4: Quarter-to-Quarter Percentage Change in GDP with highlighted Spikes

The plot shows the existence of significant spikes, which are extreme points that frequently denote periods of rapid growth or collapse. We can compute the magnitude of the changes and establish a threshold for what we consider to be a "big spike" using statistical approaches to identify huge spikes or major swings in the QoQ percentage changes in GDP.

	DATE	GDP	QoQ_Percent_Change
294	2020-04-01	19636.73	-8.827645

Table 2.2: Dates mentioning negative significant spike

Economic contraction or recessionary times are often indicated by negative spikes in QoQ GDP percentage changes. These increases could be related to:

Economic Shocks: Unexpected occurrences like financial crises, geopolitical tensions, or natural calamities that impede economic activity can cause negative surges.

Market Volatility: Negative spikes, which indicate uncertainty and diminished investor confidence, may appear during times of high market volatility.

Disruptions in the supply chain can result in negative spikes in economic growth, such as those brought on by pandemics or trade disputes.

Consumer expenditure Reduction: Reductions in consumer expenditure brought on by things like high inflation or economic ambiguity might cause negative spikes.

	DATE	GDP	QoQ_Percent_Change
15	1950-07-01	308.153	6.119504
126	1978-04-01	2331.633	5.850524
295	2020-07-01	21362.428	8.788107

Table 2.3: Dates mentioning positive significant spikes

Positive spikes in QoQ GDP percentage changes typically signal times of accelerated economic expansion or growth. These surges could be related to things like:

Economic Stimulus: Positive increases in economic activity could be the result of government policies or economic stimulus programs that promote consumer spending, company investment, and all aspects of the economy.

Technical Innovation: Positive spikes may occur during times of significant technical advancement, which will boost output and spur economic expansion.

Global Demand: Positive spikes can be caused, especially for nations that are export-oriented, by strong export demand or a good global economic situation.

Low interest rates: When interest rates are low, borrowing and investment are encouraged, which helps the economy thrive.

Unemployment Rate

The state of the labour market is indicated by the unemployment rate. An improving labour market and prospective increases in consumer spending can be shown in a declining unemployment rate. When unemployment is low, investors may be more upbeat since it implies a healthy economy with prospects for rising consumer demand.

	DATE	UNRATE	QoQ_Percent_Change
2	1948-02-01	3.8	0.05882353
3	1948-03-01	4.0	-0.02631579
4	1948-04-01	3.9	-0.10000000
5	1948-05-01	3.5	0.02564103
6	1948-06-01	3.6	0.00000000
7	1948-07-01	3.6	0.08333333

Table 2.4: Unemployment dataset

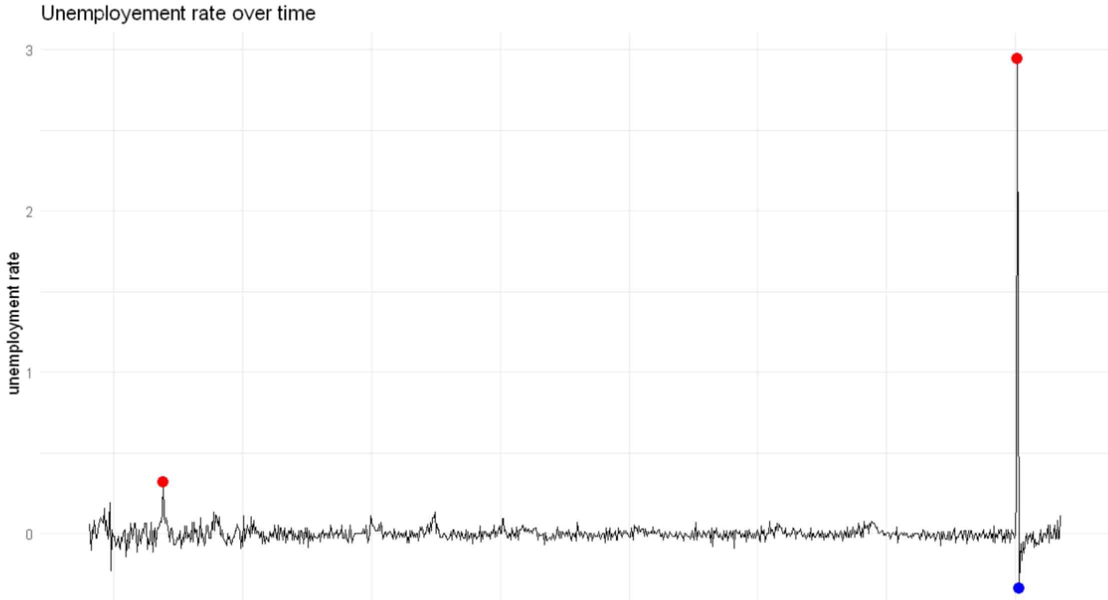

Fig 2.5: Quarter-to-Quarter Percentage Change in Unemployment with highlighted Spikes

	DATE	UNRATE	QoQ_Percent_Change
71	1953-11-01	3.5	0.3225806
867	2020-03-01	4.4	2.9428571

Table 2.5: Dates mentioning positive significant spikes

	DATE	UNRATE	QoQ_Percent_Change
868	2020-04-01	14.7	-0.3409091

Table 2.6: Dates mentioning negative significant spikes

Inflation:

Inflation measures the increase in the general price level of goods and services. High inflation erodes purchasing power and can affect investor decisions. Moderate inflation is generally healthier for the economy, as it supports economic growth without causing excessive uncertainty.

The table below shows the calculation of the percentage change in consumer price index (CPI). High inflation can erode purchasing power and affect investor decisions.

	DATE	T10YIE	Percent_Change
2	2018-08-21	2.08	0.000000000
3	2018-08-22	2.08	0.004807692
4	2018-08-23	2.09	0.000000000
5	2018-08-24	2.09	0.009569378
6	2018-08-27	2.11	0.000000000
7	2018-08-28	2.11	0.004739336

Table 2.7: Inflation Dataset

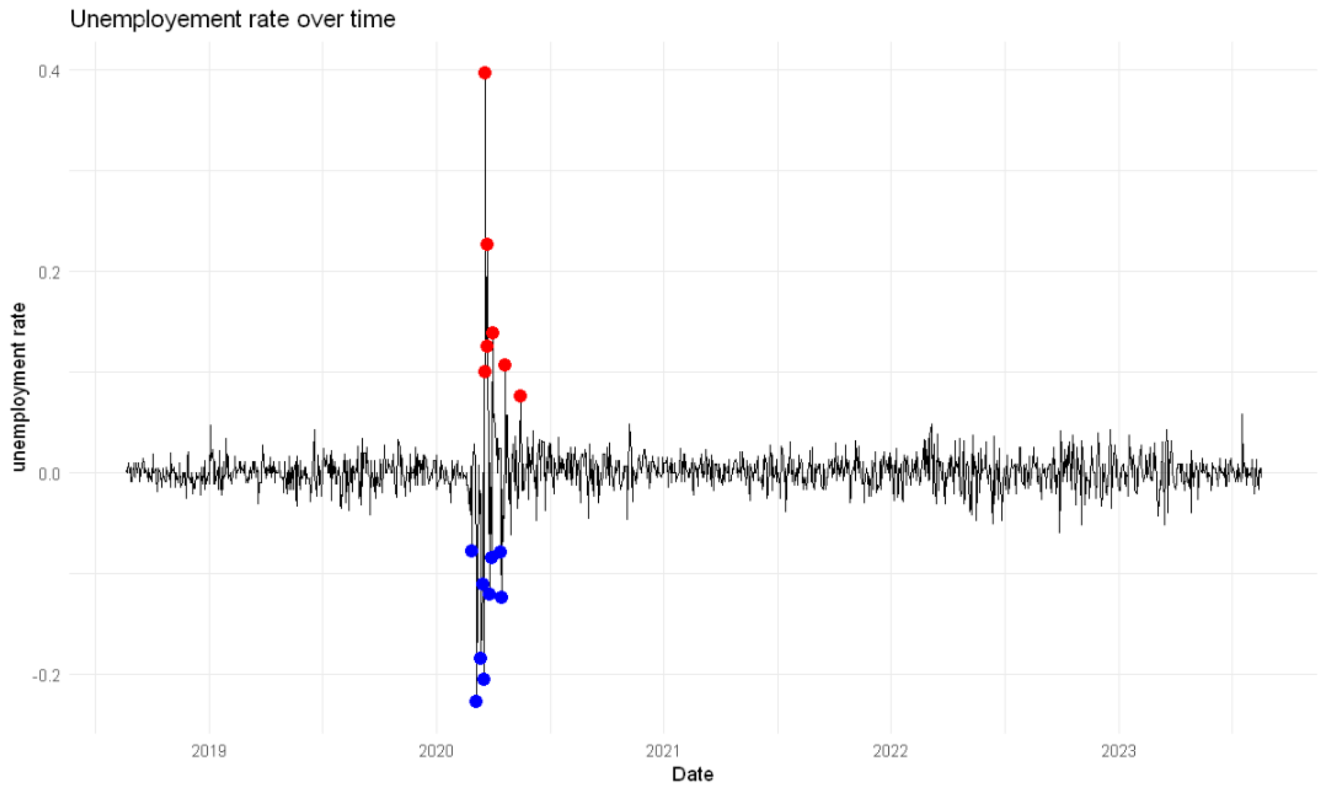

Fig 2.6: Quarter-to-Quarter Percentage Change in inflation with highlighted Spikes

	DATE	T10YIE	Percent_Change
399	2020-02-27	1.53	-0.07792208
405	2020-03-06	1.31	-0.22857143
410	2020-03-13	0.90	-0.18478261
411	2020-03-16	0.73	-0.11111111
413	2020-03-18	0.63	-0.20634921
419	2020-03-26	1.07	-0.12149533
421	2020-03-30	0.95	-0.08510638
432	2020-04-14	1.29	-0.07936508
433	2020-04-15	1.19	-0.12403101

	DATE	T10YIE	Percent_Change
399	2020-02-27	1.53	-0.07792208
405	2020-03-06	1.31	-0.22857143
410	2020-03-13	0.90	-0.18478261
411	2020-03-16	0.73	-0.11111111
413	2020-03-18	0.63	-0.20634921
419	2020-03-26	1.07	-0.12149533
421	2020-03-30	0.95	-0.08510638
432	2020-04-14	1.29	-0.07936508
433	2020-04-15	1.19	-0.12403101

Table 2.8: Dates mentioning positive and negative significant spikes

Our investigation has demonstrated that economic indicators frequently display common data points that signify positive or negative peaks. These increases may be a harbinger of important changes in the economy or in market sentiment. It is critical for investors and analysts to understand and analyse these typical data patterns because they offer insightful information about the broader economic environment.

It is crucial in this situation to understand the relationships between different economic indices. The use of correlation analysis can reveal dependencies and correlations between various indicators. For instance, if two economic indicators have a positive correlation, it means that their trends tend to move in the same direction, whereas a negative correlation means that their trends tend to move in the opposite directions. These correlations can be quite helpful for determining the overall state of the economy and for selecting investments.

We acquire a better understanding of how various aspects of the economy are interrelated by looking at the connections among economic data. For example, the relationship between GDP growth and unemployment rates might reveal how sensitive the labour market is to economic expansion. Similar to this, we can learn about the central bank's monetary policy decisions from the relationship between inflation and interest rates.

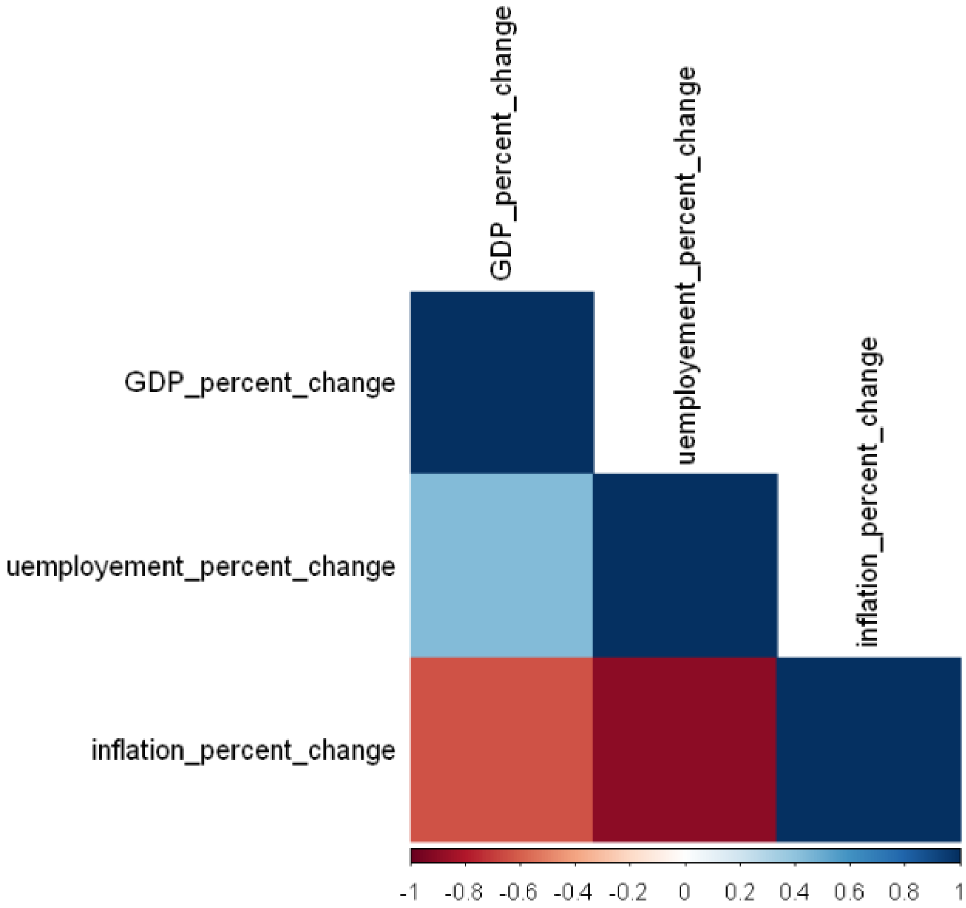

Fig 2.7: Correlation heatmap between economic indicators

GDP and Unemployment Correlation: The relationship between rising GDP and falling unemployment rates shows that times of economic boom may also be accompanied by better job market conditions. This may have a beneficial effect on investor sentiment because an expanding economy typically translates into better company earnings, possible stock market gains, and more consumer spending. During these times, investors might feel more upbeat, which might encourage them to buy more stocks and other assets.

The inverse relationship between GDP growth and inflation suggests that economic expansion may be accompanied by decreased inflation. Investors may benefit from this situation since it shows that economic growth is taking place without experiencing major price rises that reduce purchasing power. During times of decreased inflation, investors could feel more secure about the stability of their investments.

The high negative association between increases in inflation and drops in the unemployment rate suggests that there is a trade-off between these two variables. Consumer spending usually increases when the job market gets better, which could lead to an increase in inflation. When making investment decisions, investors should take this relationship into account because shifting inflation rates might affect the real return on their investments.

Investors can gain important insights into the broader economic environment by analysing connections between GDP, unemployment, and inflation. Rising GDP and declining unemployment are inversely correlated, which indicates a healthy economy, bolstering investor confidence and promoting asset allocation. Additionally, the fact that GDP growth and inflation have an inverse relationship gives investors confidence in long-term investments by demonstrating that economic growth is possible without decreasing purchasing power. The trade-off between inflation and unemployment, however, emphasises the necessity for investors to stay alert and change their tactics to successfully deal with shifting economic situations. These correlations help investors connect their portfolios with the current state of the economy and are vital tools for making informed investment decisions.

2.4 Building an Initial Portfolio

What is a portfolio?

A mixture of financial assets, including stocks, bonds, commodities, and other investments, are referred to as a portfolio in the financial markets. By employing portfolios, investors can diversify their holdings, reduce risk, and increase returns.

Portfolios are essential for investor behaviour and strategy because they serve many important functions. They make risk management easier by assisting in reducing risk exposure by distributing investments over a number of assets. By allocating assets according to risk appetite and financial goals, portfolios can aid in maximising earnings. They also encourage long-term thinking, reduce behavioural biases, and increase tax efficiency. The selection of investments, generation of income, and asset allocation within such portfolios are all components of effective portfolio management. In essence, portfolios give investors a methodical plan for reaching their financial objectives while navigating the complexity of the financial markets.

To build a portfolio, investors must select a variety of assets that are expected to perform well under diverse economic and market conditions. The sum of money allocated to each asset relies on the risk appetite and investing goals of the investor. This process involves examining the investor's financial situation, goals, time horizon, and risk tolerance in addition to studying and analysing particular assets and market movements. Due to the dynamic nature of portfolios, they should be regularly reviewed and modified to reflect changes in the market, the investor's financial situation, or their aspirations.

The weights in a portfolio show what portion of the overall value is allocated to each particular asset. The distribution of weights, which sets the portfolio's level of risk and return characteristics, is an important phase in the creation of a portfolio. The weight assigned to an asset reveals the investor's confidence in the asset's yield potential and willingness to accept the risk it implies. The historical performance, potential for future growth, sector exposure, and benefits of diversity can all be taken into consideration when determining a weight. Portfolio managers can use a range of methodologies, such as factor-based investing and modern portfolio theory, to determine the appropriate weightings. The weightings must be accurate for the expected outcomes and the success of any investment strategy.

We should create a portfolio with initial weightings of 50% each for the aapl and tsla stock markets so that we may begin investigating portfolio design and optimization.

We have determined that Apple Inc. (AAPL) and Tesla, Inc. (TSLA) are two interesting datasets for our portfolio analysis after closely examining the S&P 500. The following are some strong arguments in favour of choosing AAPL and TSLA:

Growth and Innovation: Both AAPL and TSLA have established a track record of innovation and significant growth potential. These characteristics make them appealing options for investors hoping to gain exposure to businesses known for innovative technologies and futuristic business models.

Market Capitalization: With sizable market capitalizations, AAPL and TSLA are positioned as reasonably secure investments in the market. Compared to smaller businesses, their scale and established presence might give a sense of security.

Liquidity: These stocks have high levels of liquidity, which results in regular trading activity and small bid-ask spreads. Investors benefit from this liquidity since it makes it simple to purchase and sell shares, enabling portfolio modifications as needed.

We'll put PerformanceAnalytics and quantmod to use to build a thorough portfolio analysis report. With the help of this report, we will be able to assess the performance and risk of our portfolio in relation to a benchmark, in this case the S&P 500 index.

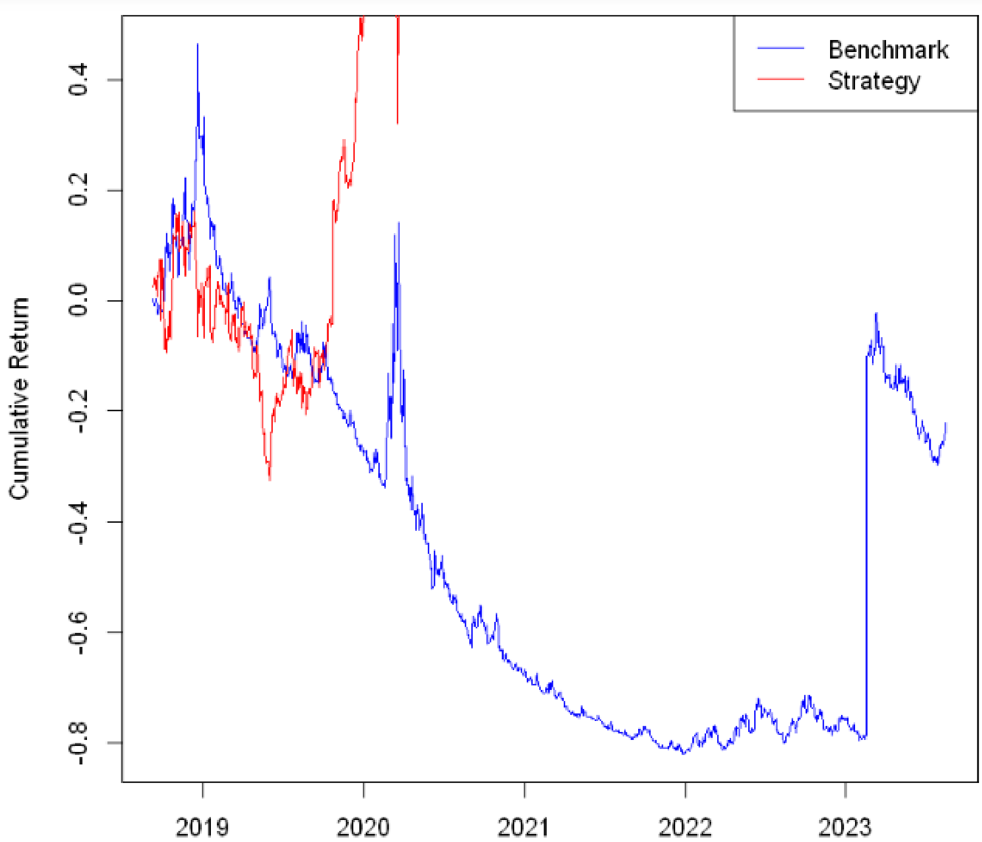

Fig 2.8: Cumulative Return Comparison

```
        Welch Two Sample t-test

data:  df$cumulative_returns and strategy$cumulative_returns
t = -47.953, df = 1253.1, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.023563 -5.550061
sample estimates:
 mean of x  mean of y 
-0.4489849  5.3378272 
```

Table 2.9: Welch Two Sample t-test

The plot and the result of the t-test strongly suggests that there is a significant difference between the cumulative returns of the benchmark and strategy.

Fig 2.9: Cumulative Return Difference between strategy and benchmark

The plot illustrates a substantial disparity in cumulative returns between our investment strategy and the benchmark.

2.5 Optimising the Portfolio

The process of choosing the ideal mix of assets and weights for a portfolio to optimise returns and reduce risk. By considering the assets' past performance, their correlations with one another, and other pertinent elements like the state of the market and the prognosis for the economy, this procedure involves choosing the weights that are the most appropriate for each asset. The fundamental objective is to build a well-diversified portfolio that balances risk and returns and fits the risk appetite of the investor.

We use a loop that iterates through various weight combinations and figures out the correlation for each combination in order to optimise the portfolio by discovering the best weights for AAPL and TSLA stocks that maximise the correlation between the cumulative return of the strategy and the cumulative return of the benchmark.

```
Best Weights (AAPL, TSLA): 0.6958  0.3042
Maximum Correlation: 0.72585
```

Table 2.10: optimised weights

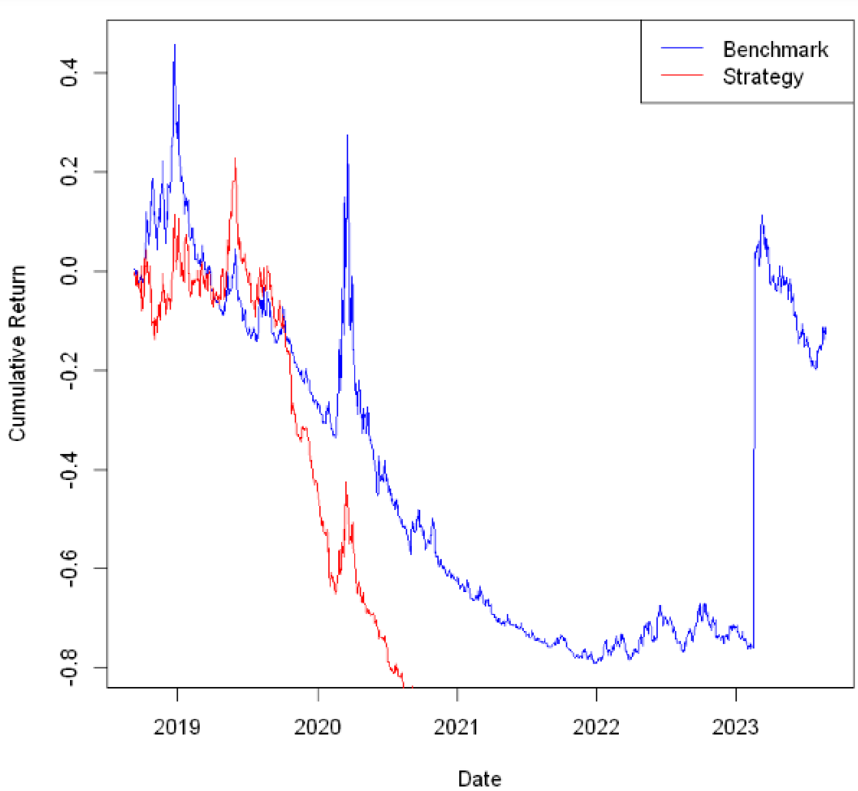

Fig 2.10: Cumulative Return Comparison after optimising

The comparison of cumulative returns before and after optimization offers insightful information about how well our investing approach has been improved. There was a major difference between our technique and the benchmark before optimization. However, following optimization, we noticed a notable convergence as our plan more successfully reacted to market circumstances. This improvement in our strategy's capacity to capitalise on market opportunities and reduce risks is indicated by our alignment with the benchmark. It is evidence of our dedication to improving our investment strategy for better success. The optimization approach has helped to produce a performance that is more dependable and coordinated, which is a crucial accomplishment in our quest for investing excellence.

This convergence provides a promising trajectory for our strategy going ahead and emphasises the importance of ongoing optimization and adaptation to shifting market conditions.

2.6 Technical Analysis

2.6.1 Simple Moving Averages

A moving average (MA) is a stock indicator that is frequently employed in technical analysis in finance. The purpose of generating a stock's moving average is to create a continuously updated average price in order to assist smooth out the price data.

By calculating the arithmetic mean of a given set of data over a certain period, a simple moving average (SMA) is created. Stock prices are calculated by adding up a group of numbers and dividing the result by the total number of prices in the group. The following formula can be used to determine a security's simple moving average:

$$SMA = (A_1+A_2+...+A_n)/n$$

A= Average in periode n

n= Number of time periods

The method involves computing different moving averages for the close price of the stock market of SP500 and experimenting with different moving average time periods to capture both short- and long-term trends.

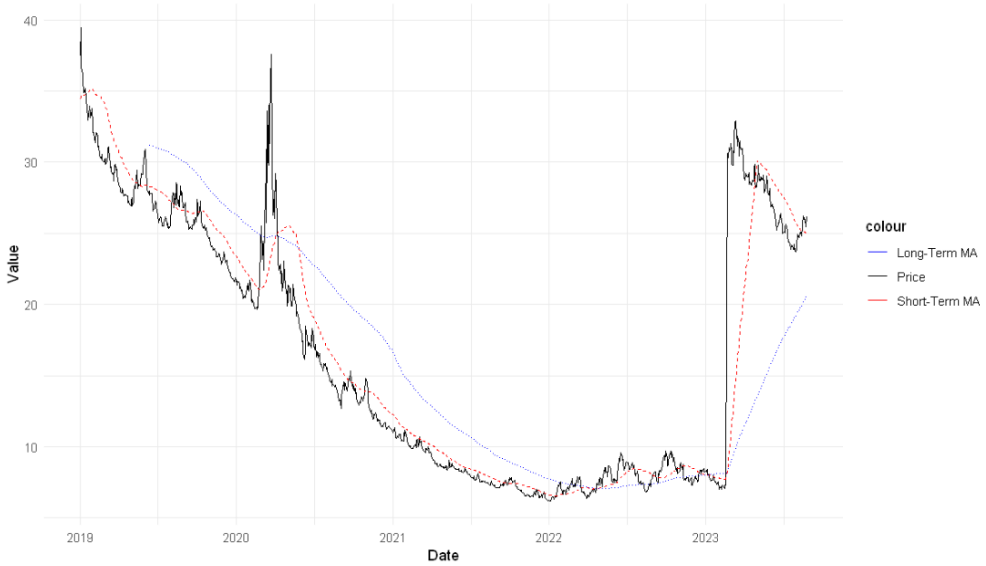

Fig 2.11: Long and short SMA representation

In fact, a frequent technique in technical analysis is to use two moving averages and base trading signals on the intersection of the two. It is common to refer to this method as a "moving average crossover strategy." Using the relationship between two moving averages, a short-term moving average, such as the 50-day moving average, and a long-term moving average, such as the 200-day moving average, this method aims to spot changes in trend direction. Investment decisions can be based on the indications produced by the crossover of these moving averages. This is how it goes:

When the short-term moving average crosses above the long-term moving average, an uptrend signal (also known as a bullish crossover) is generated. It is seen as a possible signal that the price of the stock or asset is starting to move upward, pointing to a potential buy or long position.

When the short-term moving average crosses below the long-term moving average, a downtrend signal (also known as a bearish crossover) is generated. It is seen as a potential sign that the price of the stock or asset is starting to move lower, pointing toward the possibility of taking a sell or short position.

2.6.2 Relative Strength Index

Technical analysis uses the relative strength index (RSI), a momentum indicator. To assess whether a security's price is overvalued or undervalued, RSI evaluates the speed and amplitude of recent price fluctuations.

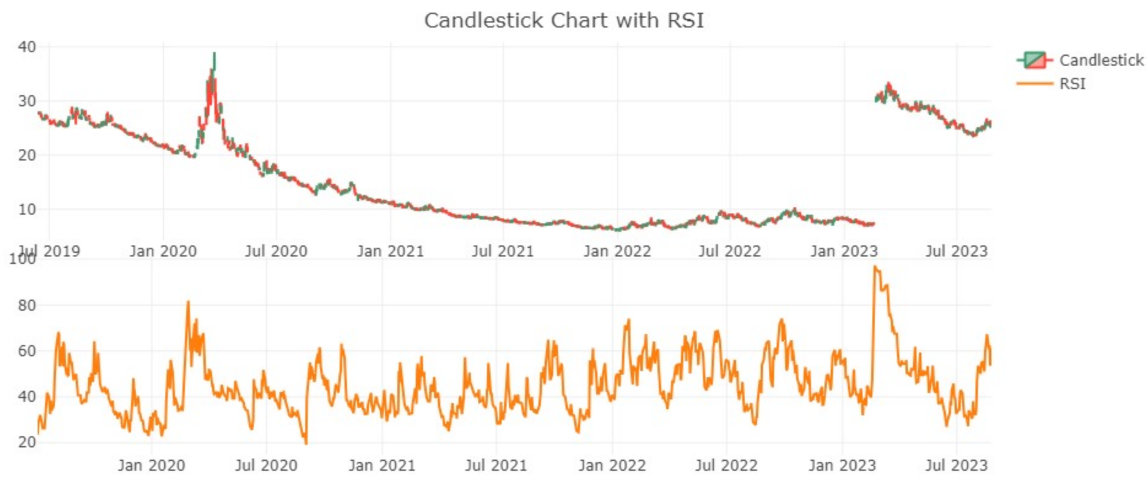

Fig 2.12: Candlestick chart of RSI

A candlestick plot, also known as a candlestick chart, is a widely used visualisation tool in finance and investment analysis. It presents price movements for a given financial asset, typically over a specified time frame, such as daily, weekly, or monthly intervals. Candlestick charts offer a visual representation of price data, making it easier for investors and analysts to interpret market sentiment and make informed trading decisions.

RSI is often used to gauge the level of buying or selling pressure in the market. Extremely high RSI values (above 70) might indicate that an asset is overbought, potentially leading to profit-taking or trend reversal. Conversely, extremely low RSI values (below 30) might suggest oversold conditions and a potential rebound.

Investors may use RSI to identify potential entry and exit points. For example, a trader might wait for RSI to drop below 30 before considering a long position, anticipating a bounce.

The RSI is calculated using the following formula:

$$RSI = 100 - [100 / (1 + RS)]$$

Where:

RS= Average Gain / Average Loss

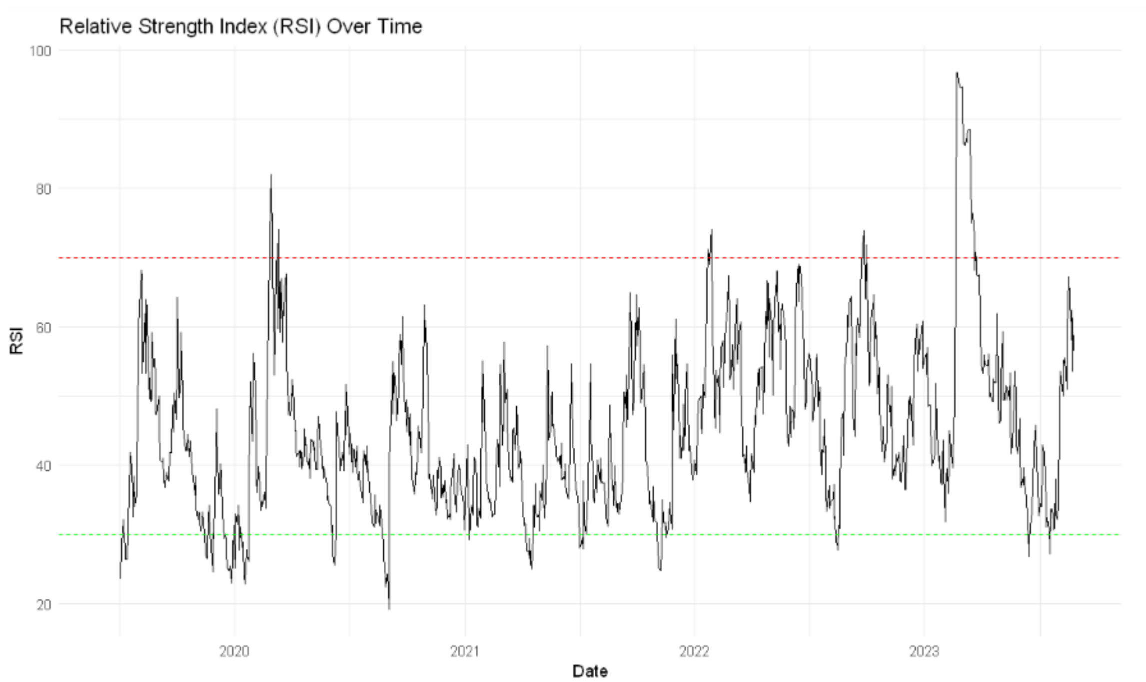

Fig 2.13: Relative RSI over time

2.7 Backtesting

Using past market data, a trading strategy's success is evaluated through the process of backtesting. In other words, it involves testing a trading strategy by examining historical data to evaluate how it would have performed if it had been applied in the past. This helps traders and investors to assess the effectiveness of their trading plans before applying their trading techniques in live trading.

Below, two different trading strategies—the RSI method and the moving average crossover method—will be back tested on the SP500 stock market.

Traders can sell or buy when the Relative Strength Index (RSI) is over 70 or below 30, which can be used to spot overbought and oversold situations.

This type of approach can be thought of as a counter trend strategy as traders look for successful entry opportunities to open a position that goes counter to the current trend.

Whenever the RSI falls below 30 and over 70, we will buy and sell, respectively, for this backtesting.

Let's assume a starting capital of $100 for this backtesting. The initial capital, the total number of trades, and the final capital after the backtest will then be printed.

2.7.1 Relative Strength Index backtesting

Backtesting part by defining the criteria for our Relative Strength Index (RSI) technique. As the next crucial phase in our trading system, we generated entry signals based on these factors. The Signal line, which is a 9-day Exponential Moving Average (EMA) of the Moving Average Convergence Divergence (MACD) line, was used to generate our signals. This line helped to amplify the MACD and was a key indicator of prospective trend changes. Traders frequently use the MACD to inform their trading decisions, interpreting crossings above and below the Signal line as bullish and bearish signals, respectively. We then determined the portfolio value from the initial returns and calculated the cumulative returns for our strategy.

EMA-12 (Exponential Moving Average - n periods): EMA is a type of moving average that gives more weight to recent prices, making it more responsive to recent price changes. In this case, EMA-12 represents the n-period Exponential Moving Average. It's calculated as the average of the closing prices over the last 12 days, with more weight given to recent days

As a last step, we gave thorough details on the quantity of transactions carried out, the starting capital put up, and the total capital earned during our backtesting procedure.

```
Number of trades: 27
Initial capital: $ 100
Final capital: $ 123.744537177615
Total return:  23.7445371776154 %
```

Table 2.11: Market backtesting - RSI strategy

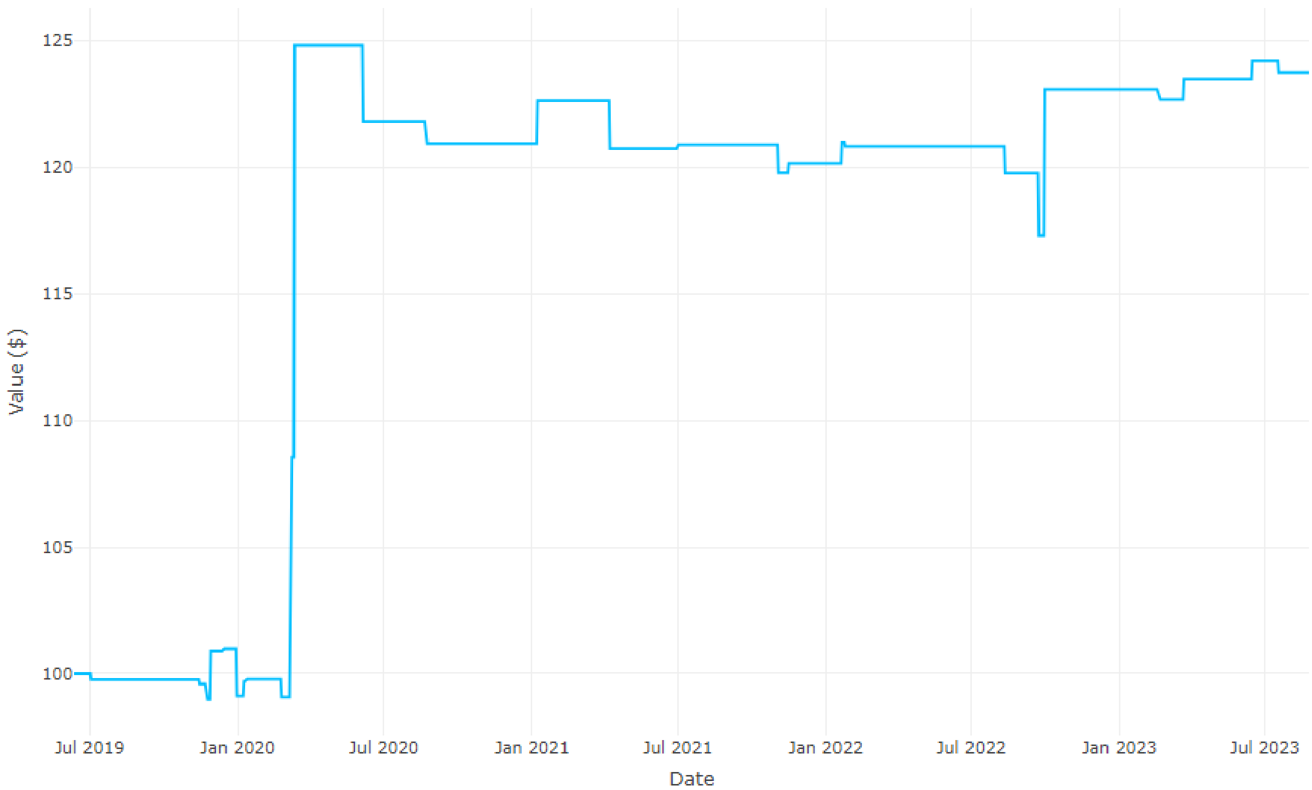

Fig 2.14: Portfolio evolution - RSI Strategy on Daily Data

Our RSI approach carried out 27 trades altogether during the course of our backtesting exercise. In order to evaluate the effectiveness of the approach in actual market circumstances, we started this test with a $100 starting capital. After our backtesting period was through, we saw a commendable result with our total capital amounting to $123.74. This indicates the strategy's capacity to produce positive returns on our original investment with a significant total return of about 23.74%. This result shows the potential profitability of our RSI-based trading technique and is encouraging. While previous performance is illustrative, it's crucial to remember that actual market conditions could differ. Therefore, it will be crucial to maintain and possibly even improve the strategy's efficiency to monitor it constantly and make changes as needed.

2.7.2 Moving Average Crossover Backtesting

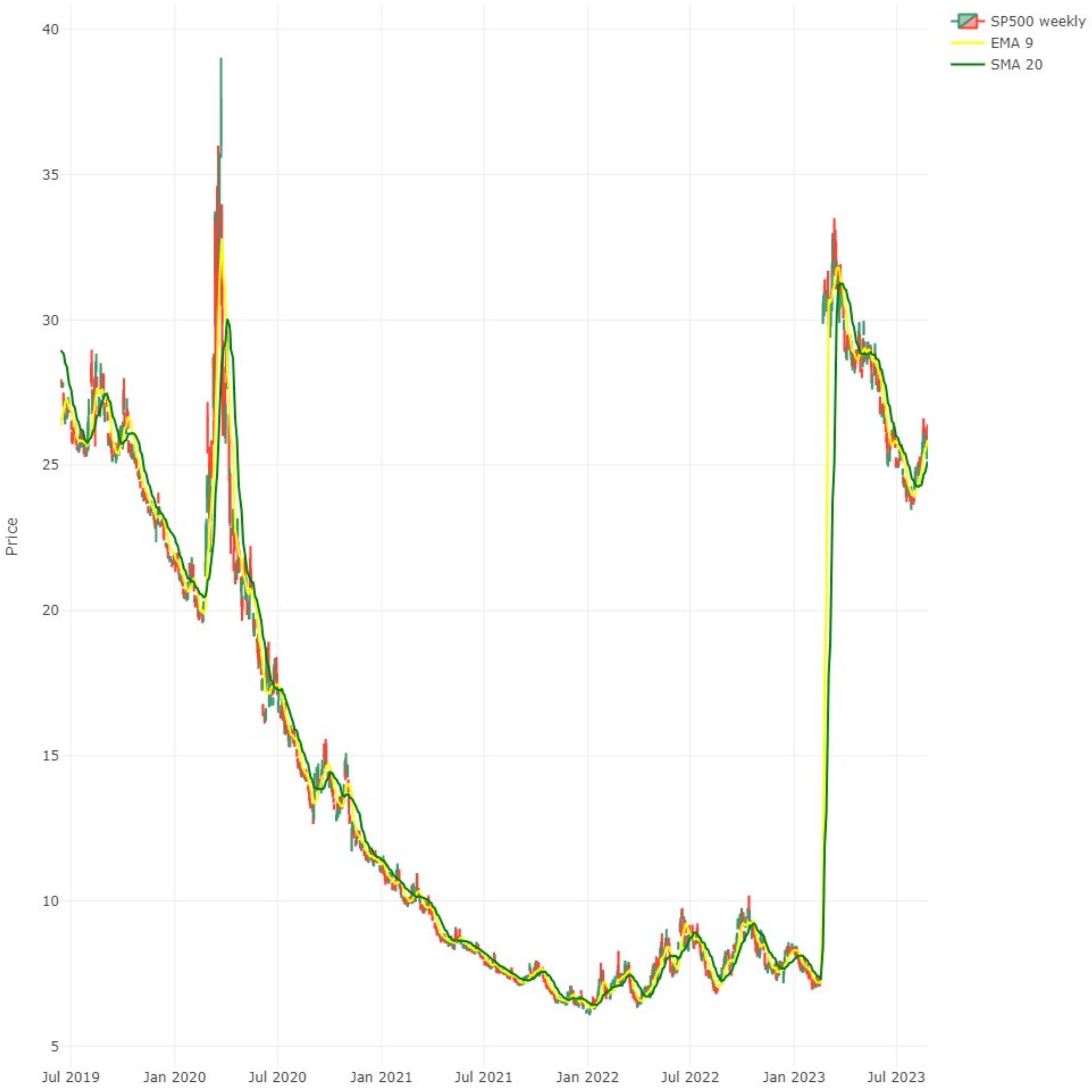

Fig 2.15: SP500 Candlestick Chart from 2021 to 2023

The moving average crossover strategy is a trend-following method compared to the RSI strategy. The foundation of this method is two moving averages, one representing a shorter time and the other a longer term.

It is a buy signal when the shorter moving average crosses above the longer moving average, and a sell signal when it crosses below the longer moving average.

We employed a simple moving average of 20 periods for the longer moving average and an exponential moving average of 9 periods for the shorter moving average in our backtesting. Because it lends more weight to recent prices and is more responsive to market fluctuations than the basic moving average, an exponential moving average was chosen for the shorter time.

For this backtesting, we'll once more assume an initial capital of $100.

```
Number of trades: 41
Initial capital: $ 100
Final capital: $ 119.846354277159
Total return:  19.846354277159 %
```

Table 2.12: Market backtesting - MACD strategy

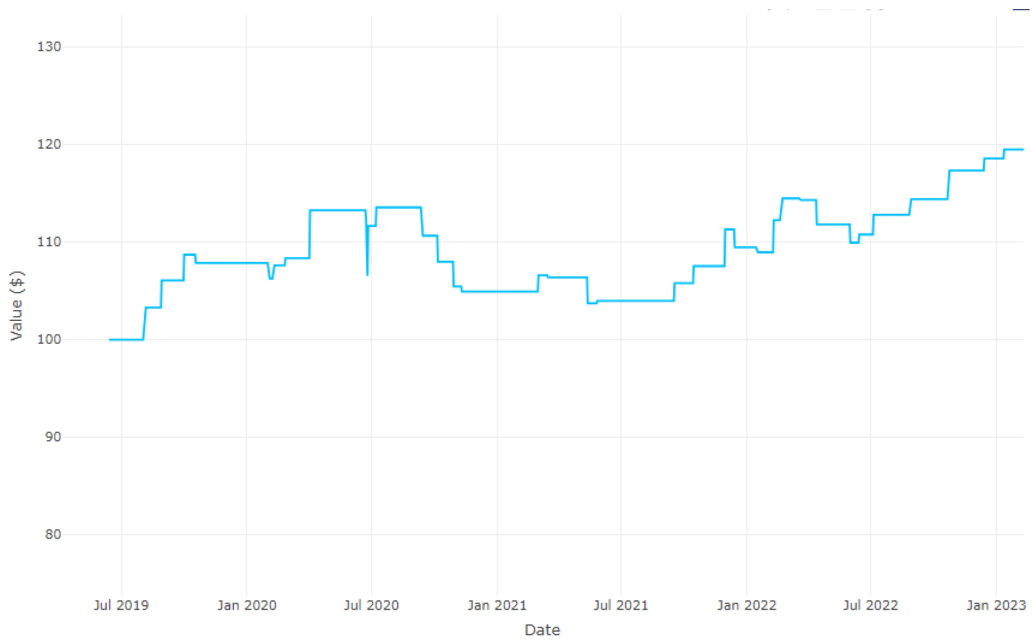

Fig 2.16: Portfolio evolution - Moving Average Crossover Strategy

A total of 41 trades were completed as a result of the extensive backtesting of the MACD (Moving Average Convergence Divergence) method. The method produced an end capital of $119.85 on an initial capital investment of $100. This translates to a total return of roughly 19.85%, indicating success for the trading strategy based on the MACD.

This outcome and the earlier RSI (Relative Strength Index) technique are comparable, and both have shown the ability to provide positive returns on the initial investment. However, compared to the MACD strategy's 19.85% total return, the RSI approach had a marginally greater total return of about 23.74%. Although both approaches have shown promise, it's vital to keep in mind that they may perform differently depending on the state of the market. When deciding between these two techniques or combining them to create a diversified approach, traders and investors should take into account their risk appetite, market outlook, and personal goals. The performance of these methods in actual trading settings must also be maintained and optimised by constant observation and prospective adjustments.

2.8 Conclusion

In this second section of our examination, we focused on the subject of investor behaviour in an effort to better understand the forces that influence financial market decision-making. We made the decision to investigate and examine the complex links between various economic indicators and their effects on investor sentiment and market movements at the outset of our journey. Through our investigation, we found connections that illuminated the interdependence of economic elements and offered crucial information for investors trying to make their way through the intricacies of the financial scene.

Additionally, using both the Moving Average Convergence Divergence (MACD) and Relative Strength Index (RSI) techniques, we set out on a systematic quest to optimize trading tactics. We were able to evaluate these strategies' performance under actual market conditions thanks to our backtesting efforts, which provided important insights into their potential to produce returns on investment.

As we take stock of our investigation into investor behaviour, it becomes clear that both market fundamentals and technical analysis are essential in determining how to allocate capital. A complex picture of investor behaviour, one in which well-informed choices can produce favourable results, is painted by the interplay between economic indicators and investing methods. However, it is equally important to understand that market dynamics are constantly changing and that flexibility continues to be an essential component of successful investing.

The practical implementation of these insights will be discussed in the chapters to follow, using our improved trading techniques and a more in-depth examination of the S&P 500 market. The journey continues as we work to combine the strengths of finance and data science to meet our investment objectives and successfully traverse the fascinating world of financial markets.

General Conclusion

We have successfully combined the disciplines of investor behaviour and market volatility in this extensive statistical project, providing key insights into the specifics of financial markets. Our meticulous modelling and research have shed light on the principles underlying the dynamics of the stock market, including correlations, returns, and price changes. This fundamental investigation has improved our understanding of market behaviour while also laying the groundwork for a more thoughtful approach to investing.

In the next stage of our research, we focused on investor behaviour, realising the crucial influence of both economic indicators and human psychology on market behaviour. The significance of a comprehensive approach to investment has been highlighted by a thorough analysis of these elements and the development of trading techniques.

When these two areas are combined, an intriguing narrative becomes apparent: investor behaviour is both a result of and a product of market volatility. Within the financial environment, this symbiotic relationship appears as a dynamic feedback loop.

Both knowledgeable investors and financial analysts should take note of the consequences of our research. It emphasises the importance of an all-encompassing approach to decision-making, where investing strategies can be guided by a thorough comprehension of market volatility and perceptions of investor behaviour. Our approaches provide financial analysts with a solid basis for using data science tools to analyse market complexity and produce useful insights.

As we look to the future, our initiative acts as a starting point for additional study and improvement. To improve forecast accuracy, future research might take into account incorporating extra data sources, sophisticated modelling methods, or real-time sentiment analysis. Further research can be done by looking into how geopolitical concerns and global events affect market volatility and investor sentiment.

In conclusion, our statistical study serves as an example of how data science and finance can work together. It highlights how we might open up new possibilities for investment strategies and decision-making by embracing the interaction between market volatility and investor behaviour. The financial environment is always changing, providing countless opportunities for research and innovation.

References:

[1] *yahoofinance.con*