Imperial College London

Imperial College London

Department of Electrical and Electronic Engineering

ELEC70103 - Signal Processing and Machine Learning for Finance

Financial Signal Processing Coursework Report

Author:
Riccardo El Hassanin

Lecturer:
Prof. Danilo Mandic

Contents

1 Regression Methods

1.1 Processing stock price data in Python

1.1.1 Logarithmic Prices

In Figure 1.1 the linear prices of the SPX Index data (from 1930 to 2017) and its logarithmic transform are plotted. The logarithmic time series provides a more compact representation range in which trends are usually represented more clearly. For instance, the general upward trend and the regional trend caused by the crash in the early 1930s, can be distinctly visualized. Moreover, this technique allows to preserve the relative order of the values (monotonic functions) and steady the variance of data, thus easing successive data manipulations.

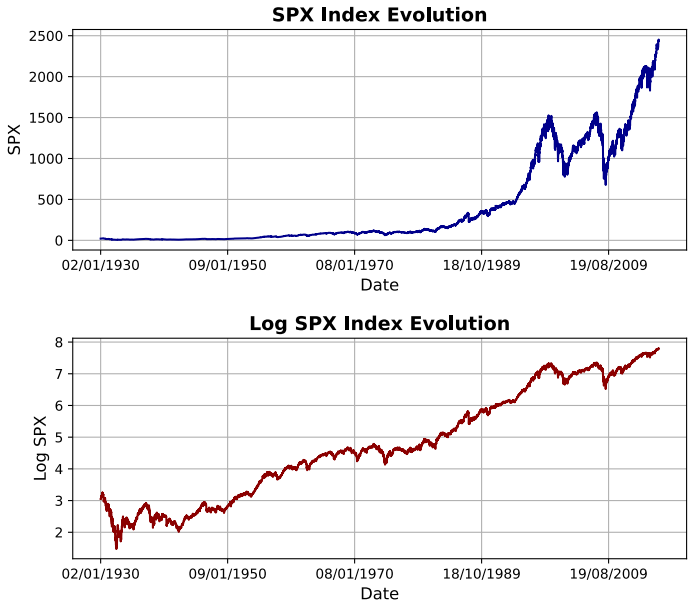


Figure 1.1: SPX Index time series and log prices

1.1.2 Stationarity analysis using rolling statistics

A stochastic process $X(t)$ is stationary if its autocorrelation function is time-invariant ($R_{xx}(t+\tau, t) = R_{xx}(t+\tau-t) = R_{xx}(\tau),\ \forall t \in \mathbb{R}$) and maintains its statistical framework over time, hence both the mean and standard deviation are not dependent on time shifts ($\mu_X(t) = \mu_X,\ \sigma_X(t) = \sigma_X,\ \forall t \in \mathbb{R}$). Figure 1.2 shows the evolution of the first and second-order statistics (mean and variance) of the price and log-price time series, using a sliding window of 252 days and 1-day increments. It is possible to observe that the sliding mean performs a smoothing operation on the data showing a clearer trend while the sliding standard variation gives an idea about the volatility of the signal. Therefore, as shown in the plot the moving average is upward trending with respect to time, and the standard deviations display strong oscillating behaviors - the log-price data is less volatile than the linear one as the oscillations are smaller indicating less instability. Nevertheless, both the linear and logarithmic prices are non-stationary stochastic processes as they do not have time-invariant statistics.

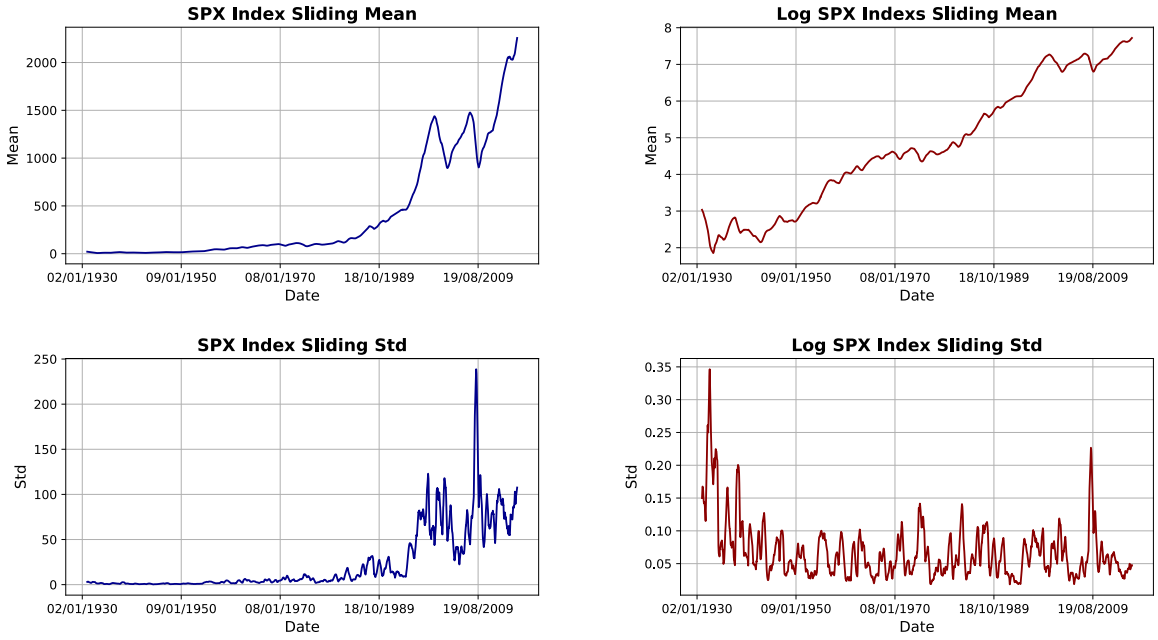


Figure 1.2: Mean and Standard deviation analysis of the SPX Index linear and log prices

1.1.3 Simple and Log Returns

In any financial analysis, it would be more desirable to deal with stationary signals rather than non-stationary ones (e.g. SPX Index in Section 1.1.2). Therefore the financial data can be processed into returns. The simple return is defined as: Simple Return:

$$R_t = \frac{p_t}{p_t - 1} - 1 \tag{1}$$

Where the p_t is the price of an asset at time instant t and R_t is the simple return at the same time instant. On the other hand, the logarithmic return r_t at time step t is defined as:

$$\begin{aligned} r_t &= \log(R_t) = \log\left(\frac{p_t}{p_t - 1} - 1\right) \\ &= \log(p_t) - \log(p_{t-1}) \end{aligned} \tag{2}$$

As seen in Figure 1.3, the sliding statics analysis on simple and log returns implies more stationarity than the plots in Figure 1.2. The moving average (sliding mean) of both simple and log returns show an oscillatory behavior around zero that suggests more stationarity than the standard price data.

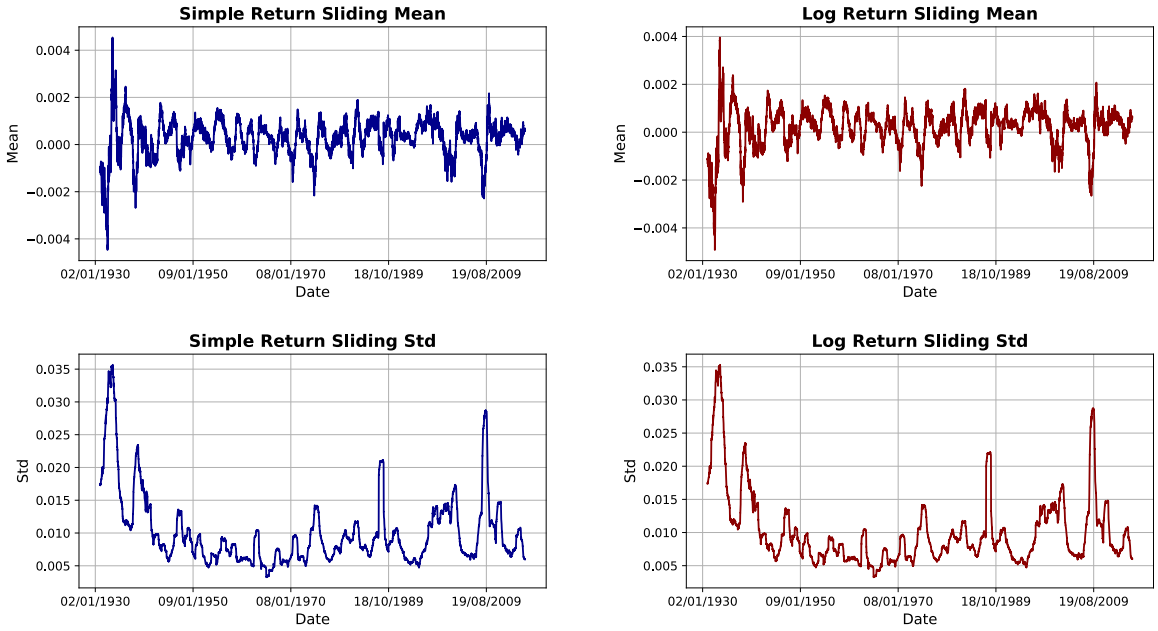


Figure 1.3: Mean and Standard deviation analysis of the simple and log returns for SPX Index data

1.1.4 Suitability of log-returns and "Jarque-Bera" test

In financial signal processing, log returns have several advantages over simple returns. Firstly, prices are considered to be log-normally distributed over short periods of time, therefore the log returns r_t are assumed to be also normally distributed. Gaussian behavior has a fundamental mathematical advantage as many statistical and signal processing techniques assume it. Moreover, logarithms, as mathematical operations, offer numerical stability since additions of small-valued numbers to the argument don't have significant effects, allow for easier manipulation of exponents with calculus and posses the time additive property (more on this in Section 1.1.5). From Figure 1.4 it is possible to observe the gaussian-like behaviour of returns from the simple and log return distribution plots.

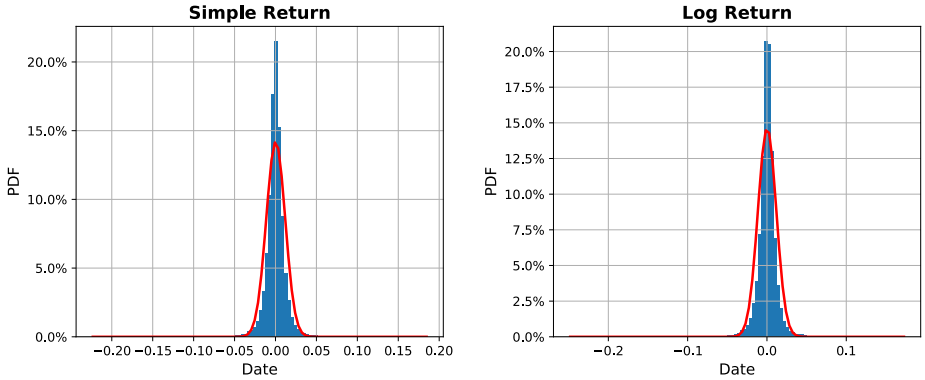


Figure 1.4: Probability distribution of simple and log returns

In order to test for Gaussianity on the data, the "Jarque-Bera" goodness-of-fit test can be used. This test checks whether the sample data have skewness (asymmetry of the distribution) and kurtosis - tailedness/shape of the curve: heavy-tailed or light-tailed - matching a normal distribution i.e. zero skewness and kurtosis equal to 3. As shown from Figure 1.5, as the number of samples increases, the log returns deviate more slowly from Gaussianity compared to simple returns, thus showing a more Gaussian behaviour.

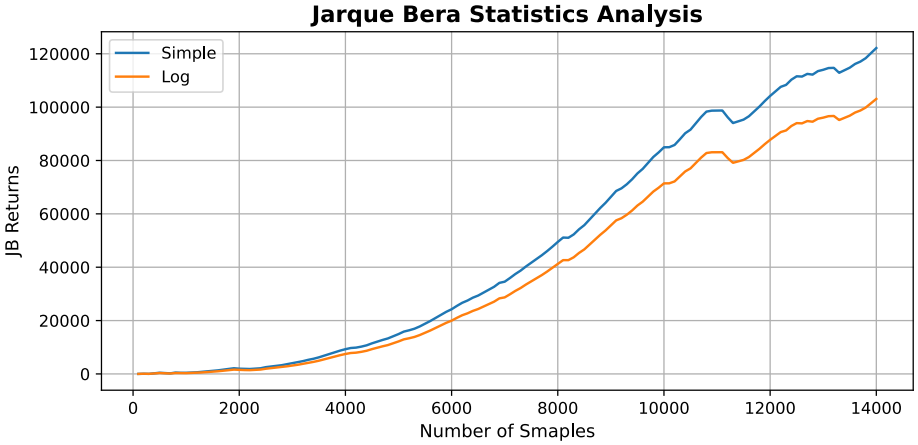


Figure 1.5: Jarque-Bera statistics evolution for simple and log returns

1.1.5 Simple and log return for Stock Purchase example

You purchase a stock for £1. The next day its value goes up to £2 and the following day back to £1. What are the simple and logarithmic returns over this period and what can you conclude about logarithmic returns on the basis of this example?

After the first day, the simple return would be 100% and the log return would be 69% while the following day, the simple return would now go down 50% and the log return would be -69%. It is possible to see that while the value of the stock has not changed as it returned to 1, the sum of the simple returns does not add to zero ($1 - 0.5 \neq 0$). However, the sum of the logarithmic returns gives zero ($0.69 - 0.69 = 0$) due to the time additive property of logarithms, hence providing an insight that the asset did not change in value. Therefore, using log-returns is advantageous when describing the changes in the value of financial assets over time.

1.1.6 Adavantages of simple return over log return

In certain circumstances, log returns are less suitable than simple returns. Firstly, the log-normality property vanishes over long-time periods. Log-normal distributions are assumed to be positively skewed but in reality, most financial time series might be negatively skewed due to financial crashes, therefore for long-term analysis it may not be suitable. Moreover, log-returns are not linearly additive across assets whereas simple returns are, thus making them more suitable when calculating the overall return of a multi-asset portfolio.

1.2 ARMA vs ARIMA Models for Financial Applications

1.2.1 Suitability of ARMA and ARIMA models

Both ARMA and ARIMA models are composed of Autoregressive (AR) and Moving Average (MA) components. An autoregressive model forecasts attempt future values based on past behavior data in a stationary time series, while a moving average model attempts to forecast future behavior based on past prediction errors. ARMA is the combination of the AR and MA models while ARIMA applies an extra integration step that removes elements of non-stationarity by differentiating.

For the case of the S&P 500 closing price time series (Figure 1.6), since the stochastic process was defined as non-stationary due to the lack of time-invariance in both the mean and standard deviation statistics, ARMA modeling would not be suited as ARMA models require stationarity (WSS) in the signal. Therefore, due to the non-stationary nature of this process ARIMA would be more appropriate.

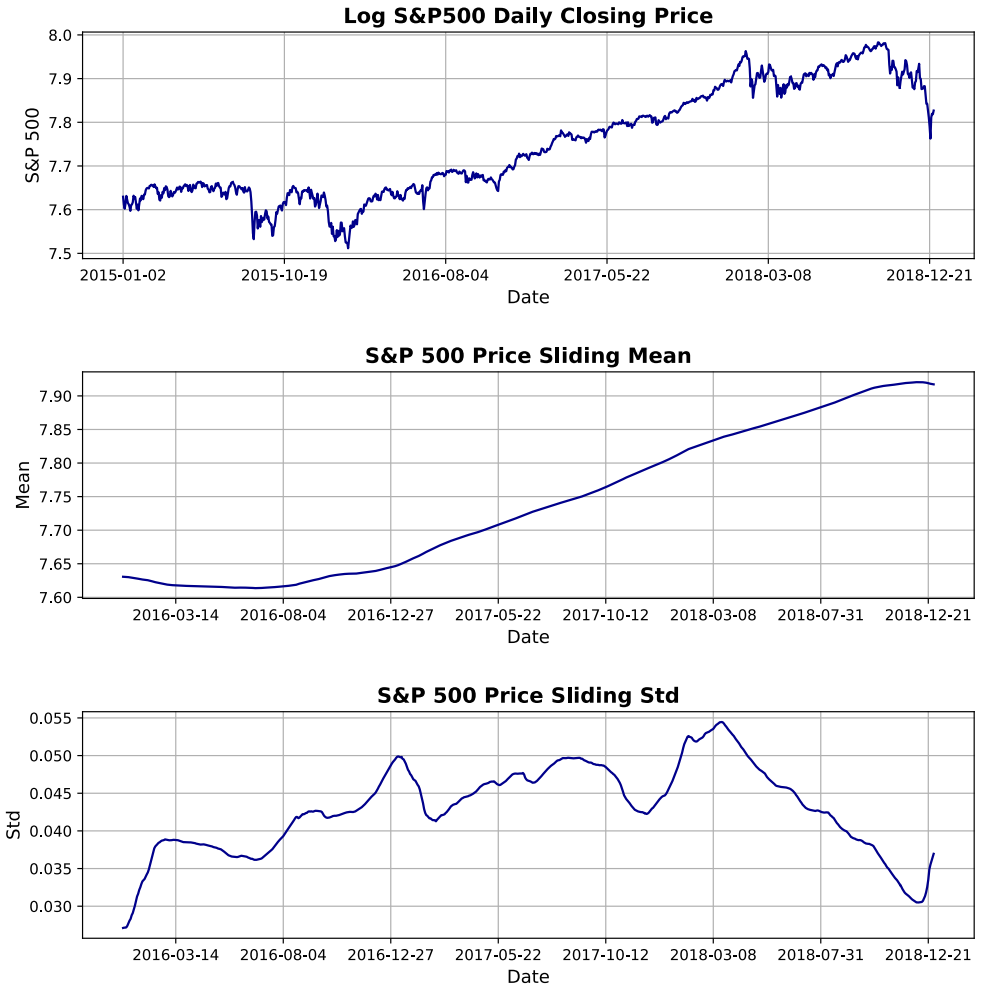


Figure 1.6: S&P 500 log prices, sliding mean and standard deviation analysis

1.2.2 ARMA (1, 0) Model

An ARMA(p,q) model is defined as follows:

$$x_t = \underbrace{\sum_{i=1}^{p} a_i x_{t-i}}_{autoregressive} + \underbrace{\sum_{i=0}^{q} b_i x_{t-i}}_{moving\ average} + \epsilon_t \tag{3}$$

Where a is the autoregressive coefficient, b is the moving average coefficient and ϵ is the noise. Hence for an ARMA(1,0) the equation becomes:

$$x_t = a_1 x_{t-1} + \epsilon_t \tag{4}$$

Which essentially is a first-order autoregressive model AR(1). As seen in Figure 1.7, predicted log prices display an overall decent prediction of the data representing a lagged version of the signal (shown clearly in the 100-days zoomed version). This is due to the fact that the autoregressive model of AR(1) is a function depending only on the previous time-step with coefficient $a_1 = 0.99736$, thus explaining the lag of the prediction. Despite the mean absolute residual is 0.006, the residual curve shows oscillatory behaviors in the high volatility sections due to the absence of a moving average part that models the shocks of external events which can cause high volatility.

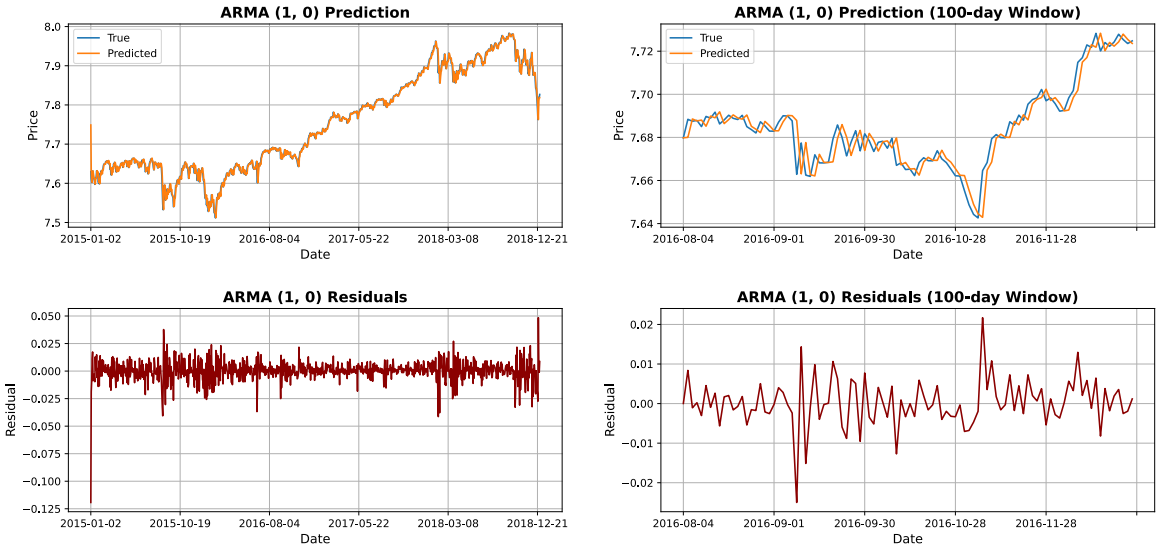


Figure 1.7: S&P 500 log prices prediction with an ARMA(1,0) model (left). A zoom on a 100-day window is provided (right)

1.2.3 ARIMA (1, 1, 0) Model

An ARIMA(p,d,q) model has an initial differentiating of order d applied to the data to remove elements of non-stationarity. The ARIMA(1,1,0) is similar to the ARMA(1,0) model and is given by

$$x_t = a_1 x_{t-1} + \epsilon_t \tag{5}$$

Where the autoregressive coefficient is $a_1 = -0.00816978$. It is possible to notice that the coefficient is small in magnitude and closer to zero indicating little to no correlation between the prediction and the previous time instance (non-stationary). The plots in Figure 1.8 are very similar to those for ARMA(1,0), although the mean absolute error slightly dropped to 0.0058. However, ARIMA model provides a more meaningful analysis since it focuses on maximising an objective function represented by the difference of the log prices, thus maximising the log returns can have a physical meaning as maximising the predicted earnings.

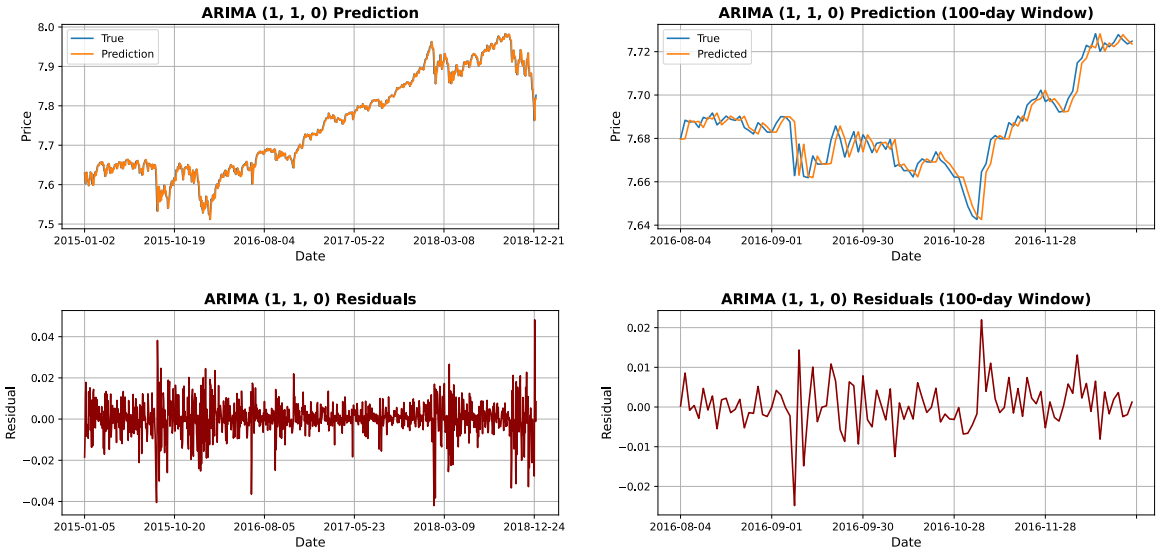


Figure 1.8: S&P 500 log prices prediction with an ARIMA(1,1,0) model (left). A zoom on a 100-day window is provided (right)

1.2.4 ARIMA analysis with log-prices

As discussed in the previous section, for an ARIMA analysis, taking the log of prices allows to maximise the log returns by using logarithmic price changes. If the ARIMA modeling is conducted on linear price changes, the outcome would not be interpreted it as modeling simple returns. Therefore, the predictions would not describe the maximum profit. This is due to the mathematical properties of the logarithms, which allow for easier manipulation of data for modeling and eventually better results.

1.3 Vector Autoregressive (VAR) Model

1.3.1 Concise VAR matrix form

Vector Autoregressive Models, Var(p), are multivariate extentions of AR models, AR(p), and are given by

$$\begin{aligned}\mathbf{y}_t &= \mathbf{c} + \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{e}_t \\ &= \mathbf{c} + \sum_{i=1}^{p} \mathbf{A}_i\mathbf{y}_{t-i} + \mathbf{e}_t\end{aligned} \tag{6}$$

Where $\mathbf{c} \in \mathbb{R}^{k\times 1}$, $\mathbf{A}_i \in \mathbb{R}^{k\times k}$ $\forall i$, $\mathbf{y}_n \in \mathbb{R}^{k\times 1}$ $\forall n \in [t-1, t-p]$ and $\mathbf{e}_t \in \mathbb{R}^{k\times 1}$ $\forall t$. This equation can also be extended in matrix form:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{k,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} + \sum_{i=1}^{p} \begin{bmatrix} a^i_{1,1} & a^i_{1,2} & \dots & a^i_{1,k} \\ a^i_{2,1} & a^i_{2,2} & \dots & a^i_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ a^i_{k,1} & a^i_{2k,2} & \dots & a^i_{k,k} \end{bmatrix} \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \\ \vdots \\ y_{k,t-i} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \\ \vdots \\ e_{k,t} \end{bmatrix} \tag{7}$$

This above equation can be represented in a concise matrix form as

$$\mathbf{Y} = \mathbf{BZ} + \mathbf{U} \tag{8}$$

where:

$$\begin{aligned} \mathbf{Y} &= \mathbf{y}_t \\ \mathbf{B} &= \begin{bmatrix} \mathbf{c} & \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_p \end{bmatrix} \\ \mathbf{Z} &= \begin{bmatrix} 1 \\ \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-p} \end{bmatrix} \\ \mathbf{U} &= \mathbf{e}_t \end{aligned} \tag{9}$$

This representation can also be generalized for a multi-period case, where T time instances are modeled:

$$\begin{aligned} \mathbf{Y} &= \begin{bmatrix} \mathbf{y}_t & \mathbf{y}_{t+1} & \mathbf{y}_{t+2} & \cdots & \mathbf{y}_T \end{bmatrix} \\ \mathbf{B} &= \begin{bmatrix} \mathbf{c} & \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_p \end{bmatrix} \\ \mathbf{Z} &= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \mathbf{y}_{t-1} & \mathbf{y}_t & \cdots & \mathbf{y}_{T-1} \\ \mathbf{y}_{t-2} & \mathbf{y}_{t-1} & \cdots & \mathbf{y}_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{t-p} & \mathbf{y}_{t-p+1} & \cdots & \mathbf{y}_{T-p} \end{bmatrix} \\ \mathbf{U} &= \begin{bmatrix} \mathbf{e}_t & \mathbf{e}_{t+1} & \mathbf{e}_{t+2} & \cdots & \mathbf{e}_T \end{bmatrix} \end{aligned} \tag{10}$$

Where $\mathbf{Y} \in \mathbb{R}^{K\times T}, \mathbf{B} \in \mathbb{R}^{K\times(KP+1)}, \mathbf{Z} \in \mathbb{R}^{(KP+1)\times T}$ and $\mathbf{U} \in \mathbb{R}^{K\times T}$.

1.3.2 Optimal VAR coefficients

In order to find the optimal coefficients $\mathbf{B}_{\text{opt}}$, the error J of the VAR model has to be minimized. By defining $J = \mathbf{U}\mathbf{U}^\top$ where $\mathbf{U} = \mathbf{Y} - \mathbf{B}\mathbf{Z}$, the least squares method can be used.

$$\begin{aligned} J(\mathbf{B}) &= \mathbf{U}^\top\mathbf{U} \\ &= (\mathbf{Y} - \mathbf{B}\mathbf{Z})^\top(\mathbf{Y} - \mathbf{B}\mathbf{Z}) \\ &= \mathbf{Y}^\top\mathbf{Y} - \mathbf{Y}^\top\mathbf{B}\mathbf{Z} - (\mathbf{B}\mathbf{Z})^\top\mathbf{Y} + (\mathbf{B}\mathbf{Z})^\top\mathbf{B}\mathbf{Z} \\ &= \mathbf{Y}^\top\mathbf{Y} - 2\mathbf{Y}^\top\mathbf{B}\mathbf{Z} + \mathbf{Z}^\top\mathbf{B}^\top\mathbf{B}\mathbf{Z} \end{aligned} \tag{11}$$

To find the optimal solution, the gradient of the above function is set to zero

$$\frac{\partial J}{\partial \mathbf{B}} = -2\mathbf{Y}\mathbf{Z}^\top + 2\mathbf{B}\mathbf{Z}\mathbf{Z}^\top = 0 \tag{12}$$

Then solving for $\mathbf{B}_{\text{opt}}$

$$\begin{aligned} 2\mathbf{Y}\mathbf{Z}^\top &= 2\mathbf{B}\mathbf{Z}\mathbf{Z}^\top \\ \mathbf{B}_{\text{opt}} &= \mathbf{Y}\mathbf{Z}^\top\left(\mathbf{Z}\mathbf{Z}^\top\right)^{-1} \end{aligned} \tag{13}$$

1.3.3 Stability of VAR

Consider a VAR(1) process

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \mathbf{e}_t \tag{14}$$

The previous instants can be written as

$$\begin{aligned} \mathbf{y}_{t-1} &= \mathbf{A}\mathbf{y}_{t-2} + \mathbf{e}_{t-1} \\ \mathbf{y}_{t-2} &= \mathbf{A}\mathbf{y}_{t-3} + \mathbf{e}_{t-2} \\ &\vdots \end{aligned} \tag{15}$$

By recursively substituting equations (12) for earlier instants, the equation for VAR(1) becomes:

$$\mathbf{y}_t = \mathbf{A}^n\mathbf{y}_{t-n} + \sum_{i=0}^{n-1}\mathbf{A}^i\mathbf{e}_{t-i} \quad \text{for } n \in [1, 2, \ldots, t] \tag{16}$$

The eigenvalues λ for a diagonisable matrix $\mathbf{A} \in \mathbb{R}^{k\times k}$, satisfy the equation $\det(\mathbf{I}\lambda - \mathbf{A})$. Therefore the eigenvalues correspond to the reciprocal of the roots of $\det(\mathbf{I} - \mathbf{A}\mathbf{z})$. This is because if we look at the z-transform of Equation (11), we obtain:

$$\begin{aligned} Y[z] &= \mathbf{A}Y[z]z^{-1} + E[z] \\ Y[z](\mathbf{I} - \mathbf{A}z^{-1}) &= E[z] \\ \frac{Y[z]}{E[z]} &= \frac{1}{\mathbf{I} - \mathbf{A}z^{-1}} \end{aligned} \tag{17}$$

From Equation (17) it is possible to notice that VAR(1) is stable if $\det(\mathbf{I} - \mathbf{A}\mathbf{z}) \neq 0$ for $|z| \leq 1$. Which means that for a diagonisable matrix, $\mathbf{A} = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^{-1}$, stability requires all eigenvalues to be less than 1 in absolute value.

1.3.4 VAR portfolio analysis with Moving Average

In this subsection, a VAR(1) model has been utilized to construct a portfolio formed of the following stocks with tickers: CAG, MAR, LIN, HCP, and MAT. Firstly the time series are detrended using a moving average model of order 66 (Figure 1.9) and then fitted to a VAR(1) model.

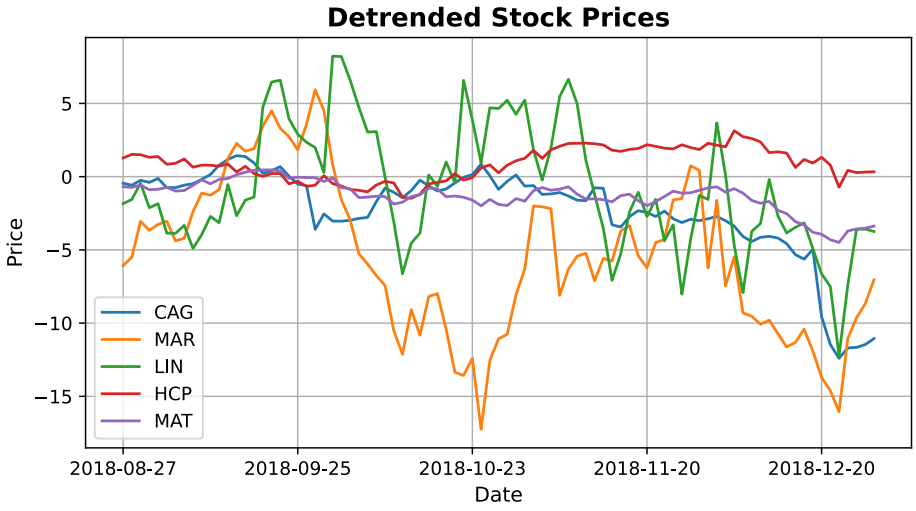


Figure 1.9: Detrended time-series data with MA(66)

Table 1 represents the parameters of the fitted VAR(1) represented and they correspond to the values of the regression matrix **A**. It is possible to observe that while the assets show positive autocorrelation with their own past values (diagonal components in **A**), most of the off-diagonal elements are small in magnitude, suggesting weak correlations between other assets. This suggests that the assets are uncorrelated as they operate in different sectors (Table 2), leading to a diversified portfolio which would potentially lower the risk associated with market fluctuations and sector-specific events.

	CAG	**MAR**	**LIN**	**HCP**	**MAT**
L1.CAG	0.87278568	0.11317853	-0.2812651	0.01191212	0.05877585
L1.MAR	-0.0637455	0.89581964	-0.1848196	-0.005004	0.02291671
L1.LIN	0.00013412	-0.1116779	0.70402268	0.00498159	-0.0255574
L1.HCP	-0.084776	-0.0838309	-0.4014168	0.93170759	-0.0464062
L1.MAT	0.64307204	0.09493095	2.03303615	-0.0128839	0.80297387

Table 1: Regression matrix **A**: Parameters for VAR(1)

The eigenvalues of **A** where found to be: $\lambda_0 = 0.715 + 0.129j$ ($|\lambda_0| = 0.726$), $\lambda_1 = 0.715 - 0.129j$ ($|\lambda_1| = 0.726$), $\lambda_2 = 1.006$, $\lambda_3 = 0.861$ and $\lambda_4 = 0.911$. The eigenvalues of the VAR matrix provide insights into the diversification and stability of the portfolio, which in turn affect the variance of the portfolio. The eigenvalues of the covariance matrix represent the amount of variance explained by each principal component. A large eigenvalue indicates that the corresponding principal component explains a significant proportion of the total variance in the data, whereas a small

eigenvalue suggests a small to the overall variance, implying that the risk is spread across multiple assets or combinations of assets, leading to a well-diversified and stable portfolio.

It is possible to notice that the magnitude of λ_2 is greater than one ($|\lambda_2| = 1.006$), which suggests that there might be some instability in the portfolio. It could be an indication that there's some concentration of risk in one or more assets or combinations of assets. This concentration of risk can contribute to the variance of the portfolio, making it more volatile and susceptible to market fluctuations and sector-specific events. However, the other eigenvalues have magnitudes less than 1, indicating that the other components of the portfolio are relatively stable and diversified. This diversification can help reduce the overall variance of the portfolio.

In summary, the eigenvalues of the covariance matrix affect the variance of the portfolio by indicating the level of diversification and stability within the portfolio. In the given case, while the weak correlations between the assets in the VAR matrix pointed towards a well-diversified portfolio, the presence of an eigenvalue slightly greater than 1 suggests that there might be some instability and increased variance in the portfolio. To minimize the overall risk, a further diversifying the portfolio is suggested by including additional uncorrelated assets.

Symbol	**GICS Sector**
CAG	Consumer Staples
MAR	Consumer Discretionary
LIN	Materials
HCP	Real Estate
MAT	Consumer Discretionary

Table 2: Sectors for each asset ticker

1.3.5 Sector-based VAR portfolio analysis

As seen in Table 3, all sectors except for "Financials" are stable since the magnitudes of the maximum eigenvalues do not exceed 1, thus providing the right conditions for a stable VAR(1) model. Even though constructing a portfolio by grouping stocks by sector may seem beneficial since it reduces the instability when modeling the portfolio, it may lead to higher correlations between the stocks, as they are likely to be affected by similar economic factors and industry trends. This results in a higher overall risk, as the assets in the portfolio are not diversified across different sectors. Therefore, it is advised to construct a more robust portfolio using uncorrelated assets from different sectors to minimise the risk. This can be achieved by carefully selecting the assets and understanding their correlations, which directly influence the overall variance of the portfolio.

On a mathematical level, this can be seen by the following equation for variance of a portfolio.

$$\begin{aligned} \sigma_p^2 &= \mathbf{w}^\top \mathbf{C}\mathbf{w} \\ \mathbf{C} &= E\left\{(\mathbf{r}[t] - \boldsymbol{\mu})(\mathbf{r}[t] - \boldsymbol{\mu})^T\right\} \\ \boldsymbol{\mu} &= E\{\mathbf{r}[t]\} \end{aligned} \tag{18}$$

where $\mathbf{C}$ is the covariance matrix the returns of assets, the individual covarinces between assets i and j are given by $[\mathbf{C}]_{ij} = \sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$ (where ρ_{ij} is the correlation coefficient). Therefore variance of the portfolio depends on both the individual variances of the assets and their pairwise correlations. Therefore, the total risk of the portfolio can be reduced by selecting assets with low correlations.

	Max Eigenvalue	**Min Eigenvalue**	**Stability**
Industrials	0.99172087	0.37124614	Yes
Health Care	0.99415312	0.09215734	Yes
Information Technology	0.99273756	0.37408106	Yes
Communication Services	0.98226301	0.75248826	Yes
Consumer Discretionary	0.99064993	0.44756257	Yes
Utilities	0.98564768	0.04211543	Yes
Financials	1.00433989	0.15257544	No
Materials	0.99174409	0.13783772	Yes
Real Estate	0.98278518	0.76356312	Yes
Consumer Staples	0.99150805	0.54645834	Yes
Energy	0.98557732	0.8257071	Yes

Table 3: Eigenvalues statistics for each stock sector

2 Bond Pricing

2.1 Examples of bond pricing

2.1.1 Percentage return per annum

We consider the case that an investor receives USD 1,100 in one year in return for an investment of USD 1,000 now. In order to compute the percentage return per annum, compounding can be used:

$$A = P\left(1 + \frac{r}{n}\right)^{nt} \tag{19}$$

where A is the final amount returned, P is the initial principal investment, r is the interest rate (expressed as decimal),t is the time period (in years) and n is the compounding frequency per time period. The previous equation can be rearranged to obtain the compound return:

$$r_n = n\left[\left(\frac{A}{P}\right)^{\frac{1}{nt}} - 1\right] \tag{20}$$

It is important to note that for the considered example $t = 1$ since it is over a period of a year.

- Annual Compounding ($n = 1$):

$$r_1 = \left[\left(\frac{1100}{1000}\right) - 1\right] = 0.1 \qquad \Rightarrow \qquad r_1 = 10\% \tag{21}$$

- Semi-annual Compounding ($n = 2$):

$$r_2 = 2\left[\left(\frac{1100}{1000}\right)^{\frac{1}{2}} - 1\right] = 0.976 \qquad \Rightarrow \qquad r_2 = 9.76\% \tag{22}$$

- Monthly Compounding($n = 1$):

$$r_{12} = 12\left[\left(\frac{1100}{1000}\right)^{\frac{1}{12}} - 1\right] = 0.957 \qquad \Rightarrow \qquad r_{12} = 9.57\% \tag{23}$$

- Continous Compounding ($n \to \infty$):

$$A = \lim_{n\to\infty} P\left(1 + \frac{r}{n}\right)^{nt} \tag{24}$$

Let $k = \frac{n}{r}$, then

$$A = \lim_{k\to\infty} P\left(1 + \frac{1}{k}\right)^{krt} = P\underbrace{\lim_{k\to\infty}\left[\left(1 + \frac{1}{k}\right)^{k}\right]^{rt}}_{Euler's\ Number} = Pe^{rt} \tag{25}$$

Therefore the rate of return (with $t = 1$) is determined by

$$r_\infty = \ln\left(\frac{A}{P}\right) = \ln\left(\frac{1100}{1000}\right) = 0.953 \qquad \Rightarrow \qquad r_\infty = 9.53\% \tag{26}$$

2.1.2 Monthly to continuous compounding

The rate of interest with continuous compounding that is equivalent to 15% per annum with monthly compounding, is obtained in the following steps.

$$\begin{aligned} \left(1 + \frac{r_n}{n}\right)^{nt} &= e^{r_\infty t} \\ r_\infty t &= \ln\left(\left(1 + \frac{r_n}{n}\right)^{nt}\right) \\ r_\infty &= n\ln\left(1 + \frac{r_n}{n}\right) \end{aligned} \tag{27}$$

Solving the above equation given $n = 12$, $r_{12} = 15\%$:

$$r_\infty = 12 \times \ln\left(1 + \frac{0.15}{12}\right) = 14.91\% \tag{28}$$

2.1.3 Quarterly to continuous compounding

The continuous compounding interest equivalent to 12% per annum with quarterly compounding is obtained using Equation (27) with $n = 4$, $r_4 = 12\%$:

$$r_\infty = 4 \times \ln\left(1 + \frac{0.12}{4}\right) = 12.18\% \tag{29}$$

Given an initial deposit of USD $10,000$, the interest paid in each quarter will be:

$$\begin{aligned}
Q_1 &= \$10,000\left(e^{\frac{0.1218}{4}} - 1\right) = \$304.5 \\
Q_2 &= \$(10,000 + Q_1)\left(e^{\frac{0.1218}{4}} - 1\right) = \$313.8 \\
Q_3 &= \$(10,000 + Q_1 + Q_2)\left(e^{\frac{0.1218}{4}} - 1\right) = \$323.4 \\
Q_4 &= \$(10,000 + Q_1 + Q_2 + Q_3)\left(e^{\frac{0.1218}{4}} - 1\right) = \$333.2
\end{aligned} \tag{30}$$

2.2 Foward Rates

Suppose that the one–year interest rate, r_1 is 5%, and the two–year interest rate, r_2 is 7%. If you invest USD 100 for one year, your investment grows to $100 \times 1.05 = USD\ 105$; if you invest for two years, it grows to $100 \times 1.07^2 = USD\ 114.49$. The extra return that you earn for that second year is $1.07^2/1.05 - 1 = 0.090$, or 9.0%.

2.2.1.a Would an investor be happy to earn that extra 9% for investing for two years rather than one?

An investor might be willing to lock their money for an extra year to achieve this extra 9% return as it represents a higher overall return on their investment. However, some investors may prioritize liquidity and flexibility over the additional return. They might prefer to invest for only one year to maintain the option to access their funds sooner or to invest in other opportunities as they arise. Another factor that might play a role in their decision-making process is an investor's risk tolerance. Those with a higher risk tolerance may be more willing to invest for a longer period, while those with a lower risk tolerance might prioritize shorter-term investments.

2.2.1.b 5%, 7%, and 9% investment strategies analysis

The 5% investment strategy provides a lower return but offers greater flexibility and liquidity. An investor can access their funds after one year, which can be beneficial in case of emergencies or if they want to reinvest in other opportunities that might arise. The two-year investment at 7% provides a higher overall return compared to the one-year investment, however, the investor's money will be locked for a longer period, which could limit their ability to access funds or invest in other opportunities. The foward rate of 9% return for the second year of the two-year investment can be attractive to some investors who are willing to commit their funds for a longer period in exchange for a higher return.

If liquidity and flexibility are top priorities, the one-year investment at 5% may be more suitable. On the other hand, if the investor is willing to commit their funds for a longer period to earn a higher return, the two-year investment at 7% could be more appealing.

2.2.1.c Advantages and disadvantages of the forward rate of 9%

The foward rate of 9% represents an advantage as it provides a higher return in the second year compared to the one-year interest rate of 5%. Moreover, by investing in a two-year deal the investor locks in their return for two years, providing stability and avoiding potential fluctuations in the one-year interest rate during that time. This can offer some level of protection against negative changes in the near future.

However, longer-term investments generally carry higher risks than shorter-term investments (unforeseen factors can potentially affect the investment). It can be disadvantageous if the investor needs access to their funds sooner or wants to change strategy to take advantage of other investment opportunities that may arise during that time.

2.2.1.d How much would you need to go from 1y investment to 2y investment and what does it depend upon?

The amount needed to go from a 1-year investment to a 2-year investment strategy can be computed by the following formula:

$$(1 + r_j)^j = (1 + r_i)^i (1 + f_{i,j})^{ji} \tag{31}$$

2.3 Duration of coupon-bearing bond

2.3.1.a Duration for 1% bond

The duration of the 1% 7-year bonds is computed with the following formula and the results are represented in Table 4.

$$\text{Duration} = \frac{1 \times PV\left(C_1\right)}{PV} + \frac{2 \times PV\left(C_2\right)}{PV} + \frac{3 \times PV\left(C_3\right)}{PV} + \cdots + \frac{T \times PV\left(C_T\right)}{PV} \tag{32}$$

Year	1	2	3	4	5	6	7	Total
Payment	\$10	\$10	\$10	\$10	\$10	\$10	\$1010	\$1070.00
PV $(\mathbf{C}_t)$ at 5%	\$9.52	\$9.07	\$8.64	\$8.23	\$7.84	\$7.46	\$717.79	PV = \$768.55
Fraction of PV $\left[\frac{PV(C_t)}{PV}\right]$	0.0124	0.0118	0.0112	0.0107	0.0102	0.0097	0.9340	**1**
Year × Fraction of PV $\left[t \times \frac{PV(C_t)}{PV}\right]$	0.0124	0.0236	0.0337	0.0428	0.0510	0.0583	6.5377	**Duration = 6.76 years**

Table 4: Calculating the duration of the 1% 7-year bonds

2.3.1.b Modified Duration

The modified duration for the 1% bonds in the above table is computed using the following formula:

$$\text{Modified duration} = \text{volatility } (\%) = \frac{1}{P(\lambda_0)} \frac{dP(\lambda)}{d\lambda} = \frac{\text{Duration}}{1 + \frac{\lambda}{n}} \tag{33}$$

where λ represents the yield and n is the number of compounding periods per year, in this case $n = 1$. Given a yield to maturity of 5% ($\lambda = 0.05$) and the duration calculated in part (a), the modified duration is:

$$\text{Modified duration} = \frac{6.76}{1 + 0.05} = 6.44\% \tag{34}$$

As seen from Equation (33), for yearly compounding $n = 1$, the $Modified\ Duration = \frac{\text{Duration}}{1+\text{yield}}$, while for continuous compounding $n = \infty$, the two durations are exactly equivalent: $Modified\ Duration = \text{Duration}$.

2.3.1.c Advantages of duration against unexpected changes in interests rates (volatility)

A long-term investment such as a pension plan is characterized by a bigger duration than a shorter investment, thus it is more prone to changes in interest rates. Given that duration measures the average time it takes to receive a bond's cash flows, a higher duration indicates that the bond has a longer time to maturity (or a lower compounding rate), making it more sensitive to interest rate changes. Therefore investors can used it to estimate the percentage change in a bond's price for a given change in interest rates. Using Modified Duration provides the investors with the percentage change in a bond's price for a 1% change in interest rates. It is a more accurate measure of interest rate sensitivity and directly indicates the bond's price change in response to interest rate changes. Thus by using modified duration, investors can manage the interest rate risk associated with a bond.

2.4 Capital Asset Pricing Model (CAPM) and Arbitrage Pricing Theory (APT)

2.4.1 Equally-weighted Market Returns

The daily equally-weighted market returns, R_m, can be estimated by averaging each individual company's return on a daily basis. Figure 2.1 shows the daily equally-weighted market returns of all companies taken into consideration as well as the cumulative market returns. It can be observed that the market returns present an oscillating trend with a market return average of 0.00472% per day and a standard deviation of 0.00664. It can also be observed that if a portfolio were to be constructed with equally allocated investments in all assets, the cumulative return would eventually be 0% at the end of the studied period.

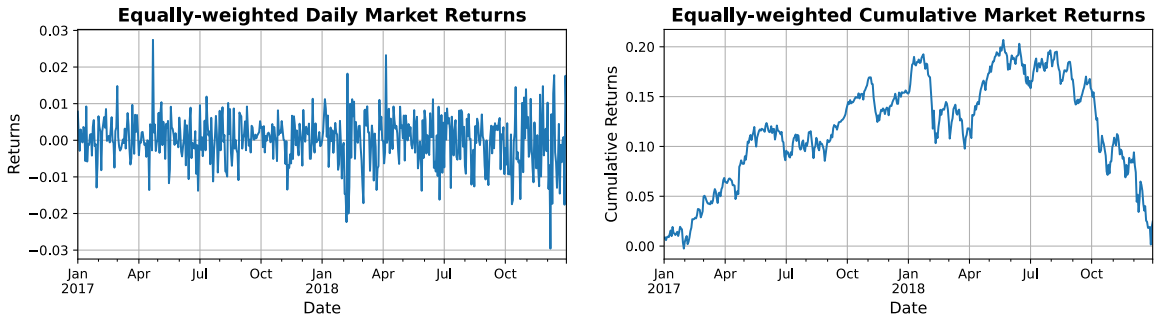


Figure 2.1: Equally-weighted market returns and cumulative market returns

2.4.2 Equally-weighted Rolling Beta

An asset's sensitivity to the market (systematic risk coefficient), β_i is defined as

$$\beta_i = \frac{cov(r_i, R_m)}{var(R_m)} \tag{35}$$

where r_i is the return of the given asset and R_m is the market return. For every company, i, the β_i systematic risk coefficient can provide the investor with a measure of the volatility of the asset i in correlation to the whole market. The volatility can be interpreted as follows:

- $\beta_i > 1$: the asset i is more volatile than the market, thus increasing the average risk in a portfolio
- $\beta_i = 1$: the asset i is as volatile than the market
- $\beta_i < 1$: the asset i is less volatile than the market, thus reducing the average risk in a portfolio
- $\beta_i = 0$: the asset i is not correlated to the market

Figure 2.2 shows a rolling (sliding) beta estimation, β_i, for every company i, with a rolling window of 22 days. It can be estimated and also observed from the plot that the average systematic risk coefficient is $\bar{\beta} = 1$ which is expected since all assets in the market are equally weighted in the portfolio. The distribution of the rolling betas shows a Gaussian-like behavior with a standard deviation of 0.708. This normal distribution of values implies that the market is composed of stocks that are diverse in terms of riskiness, it contains both: highly volatile and robust assets.

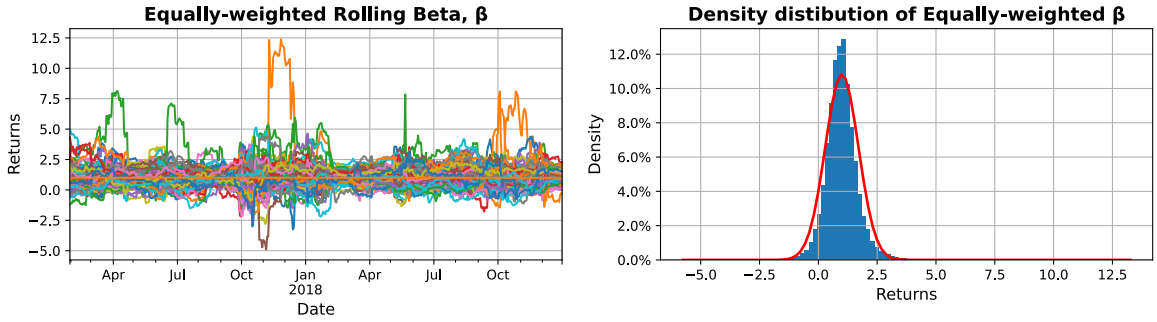


Figure 2.2: Equally-weighted rolling β and its probability density

2.4.3 Cap-weighted Market Returns

Instead of considering equally-weighted market returns, a new market portfolio is created where all assets in the market are weighted with respect to their market capitalisation, thus obtaining a weighted market return, defined as:

$$R_m = \sum_i \frac{\text{mcap}_i \times r_i}{\sum_i \text{mcap}_i} \tag{36}$$

where mcap_i represents the market cap of a company i. Figure 2.3 shows the daily weighted market returns and cumulative returns. It can be seen that the market returns also present an oscillating trend but with lower a market return average of 0.0188% per day and an std of 0.00659. Moreover, when looking at the cumulative returns, the weighted market portfolio yields a cumulative return of 9.8% over the 2 years period. These improved results show the positive impact of the weighting. Intuitively, given the robust nature of high market cap companies and the riskier nature of low market cap assets, constructing a weighted market portfolio with respect to the market cap allows to include more best-performing companies and less risky assets, thus reducing the overall risk and volatility associated with the weighted market return.

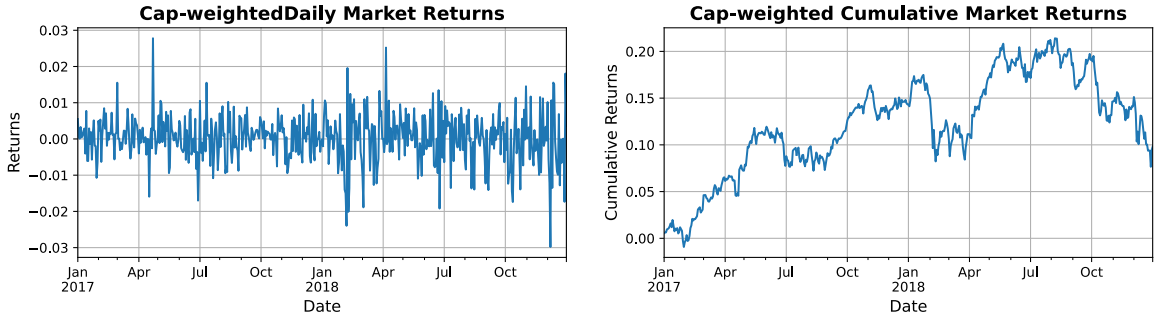


Figure 2.3: Cap-weighted market returns and cumulative market returns

2.4.4 Cap-weighted Rolling Beta

Figure 2.4 shows a rolling (sliding) beta estimation of a rolling window of 22 days applied to the cap-weighted market portfolio. The mean systematic risk $\bar{\beta}$ was found to be 0.96, implying less volatility ($\bar{\beta} < 1$), with a standard deviation of 0.694. Given the lower values obtain when compared to the equally-weighted analysis, it can be concluded that the market portfolio is composed of stocks that are less volatile and therefore less risky.

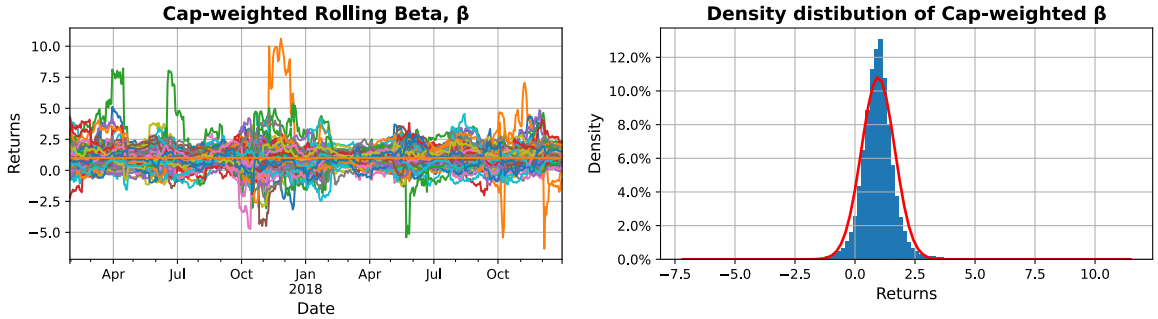


Figure 2.4: Cap-weighted rolling β and its probability density

2.4.5 Arbitrage Pricing Theory

Arbitrage Pricing Theory (APT) states that the expected return of a risky asset can be expressed as a linear combination of systematic factors, which can be mathematically expressed as the following cross-sectional regression (for two factors):

$$r_i = \alpha + \beta_{m,i} R_m + \beta_{s,i} R_s + \epsilon_i \tag{37}$$

where r_i denotes the return per company, ϵ_i represents the residual of this regression (company's specific return), α is the risk-free rate of return, R_m is the market return with a systematic risk $\beta_{m,i}$ while R_s is the return relevant to the market cap (size) of the stock with a sensitivity of $\beta_{s,i}$.

2.4.5.a

Equation (37) can be expressed in matrix form as:

$$\mathbf{r} = \mathbf{A}\mathbf{x} + \epsilon \tag{38}$$

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_i \end{bmatrix} = \begin{bmatrix} 1 & \beta_{m,1} & \beta_{s,1} \\ 1 & \beta_{m,2} & \beta_{s,2} \\ \vdots & \vdots & \vdots \\ 1 & \beta_{m,i} & \beta_{s,i} \end{bmatrix} \begin{bmatrix} \alpha \\ R_m \\ R_s \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_i \end{bmatrix} \tag{39}$$

Therefore it is possible to regress values for α, R_m, R_s for every day of the time series using the Ordinary Least Squares (OLS) method, which corresponds to minimising the following objective function:

$$\min_{\mathbf{x}} ||\epsilon||^2 = ||\mathbf{r} - \mathbf{A}\mathbf{x}||^2 \tag{40}$$

where the optimal solution computed is:

$$\mathbf{x}^* = \begin{bmatrix} \alpha \\ R_m \\ R_s \end{bmatrix} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{r} \tag{41}$$

2.4.5.b

From Figure 2.5 and Table 5 it is possible to notice that α has a larger magnitude, thus implying that has the biggest influence on the return, and a higher variance, suggesting higher volatility than the other parameters. Furthermore, it is possible to observe that R_s has the lowest magnitude and variance of almost 0, thus implying it has the least influence on the returns and lower volatility.

	Mean	**Std**	**Variance**
R_m	-0.000286	0.007913	0.000063
R_s	0.000192	0.001732	0.000003
α	-0.004216	0.041191	0.001697

Table 5: descriptive statistics of α, R_m, R_s

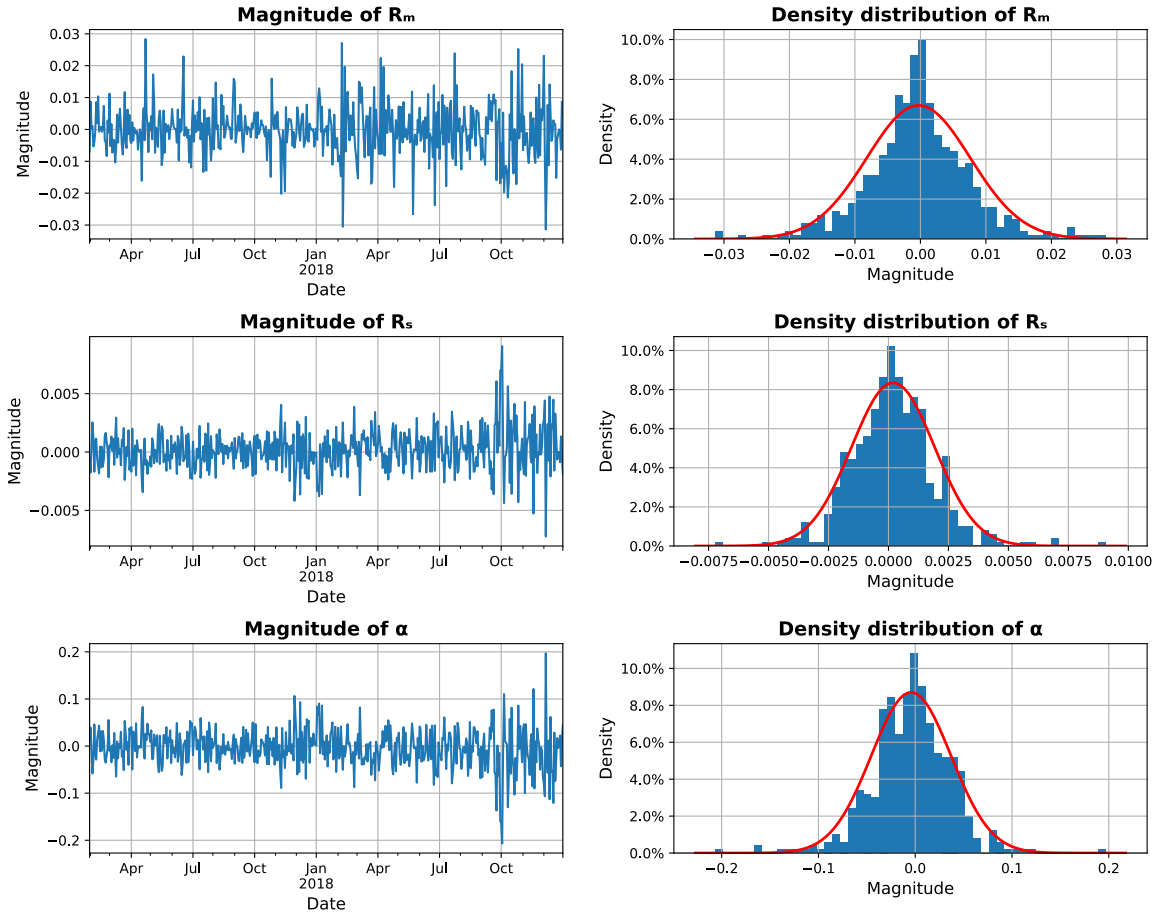


Figure 2.5: Magnitude and variance analysis

2.4.5.c

The specific return represents the difference between the actual return and the one obtained through the regression of α, R_m, R_s, essentially it is the error term of the OLS regression. Comparing the returns and specific returns in a scatter and histogram plot in Figure 2.6 reveals that ϵ_i is highly positively correlated to the returns given the line sloping upwards, and a mean correlation of 0.811. Therefore, a high correlation suggests that this two-factor model is unable to model the actual returns observed, thus more factors are needed for an adequate Arbitrage Pricing Theory model.

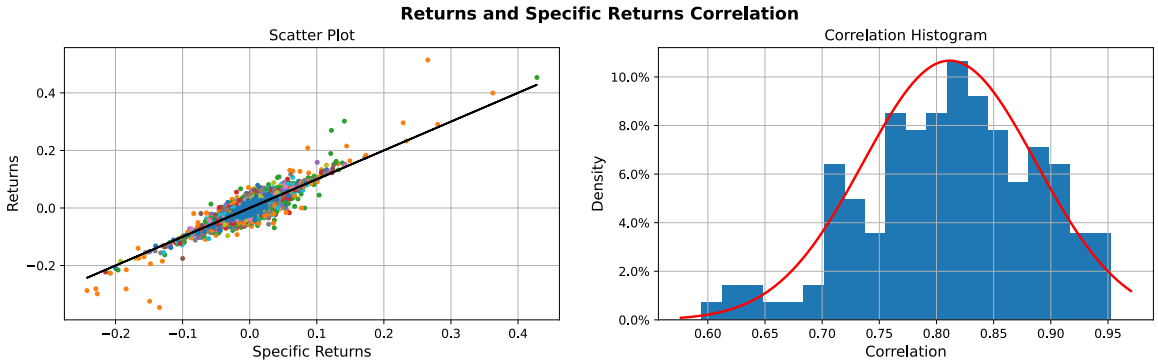


Figure 2.6: Correlation between the returns

2.4.5.d

The vectors, R_m, R_s can be combined into a matrix $\mathbf{R}$

$$\mathbf{R} = \begin{bmatrix} R_{m,1} & R_{s,1} \\ \vdots & \vdots \\ R_{m,500} & R_{s,500} \end{bmatrix} \tag{42}$$

The covariance matrix, $cov(\mathbf{R})$ is calculated using a rolling window of 22 days, and its magnitude and percentage change are shown in Figure 2.7. The plot shows that $cov(\mathbf{R})$ is very small in magnitude, close to 0, implying that there is no formal correlation between R_m and R_s. Moreover, the highly oscillatory behavior of the magnitude plot, alongside the large spikes observed from their daily percentage change, are signs of significant instability.

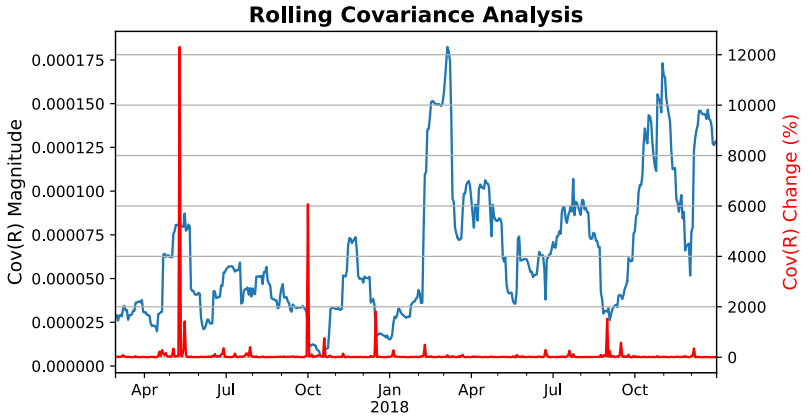


Figure 2.7: Magnitude and stability of covariance matrix

2.4.5.e

The specific return of a company i for every day t is defined as $\epsilon_{i,t}$, which in matrix form for all companies and all days is:

$$\mathbf{E} = \begin{bmatrix} \varepsilon_{1,t=0} & \cdots & \varepsilon_{157,t=0} \\ \varepsilon_{1,t=1} & \cdots & \varepsilon_{157,t=1} \\ \vdots & \ddots & \vdots \\ \varepsilon_{1,t=500} & \cdots & \varepsilon_{157,t=500} \end{bmatrix} \tag{43}$$

The covariance matrix $cov(\mathbf{E})$ is calculated and Principal Component Analysis (PCA) is performed. Through PCA, the eigenvalues of the covariance matrix are obtained and are used to determine the proportion of variance explained by each principal component. It was found that the percentage of the variance explained by the first principal component is 7.37%. Meaning that only 7.37% of information on the specific returns is represented by the first component. Moreover, from Figure 2.8, it can be observed that 76 principal components are needed out of 141 in order to explain 90.22% of the information of the original data. This shows that modeling prices accurately with the 2-factor APT model requires high dimensionality, which implies high computational expansiveness. This is not ideal since PCA aims to reduce dimensionality significantly.

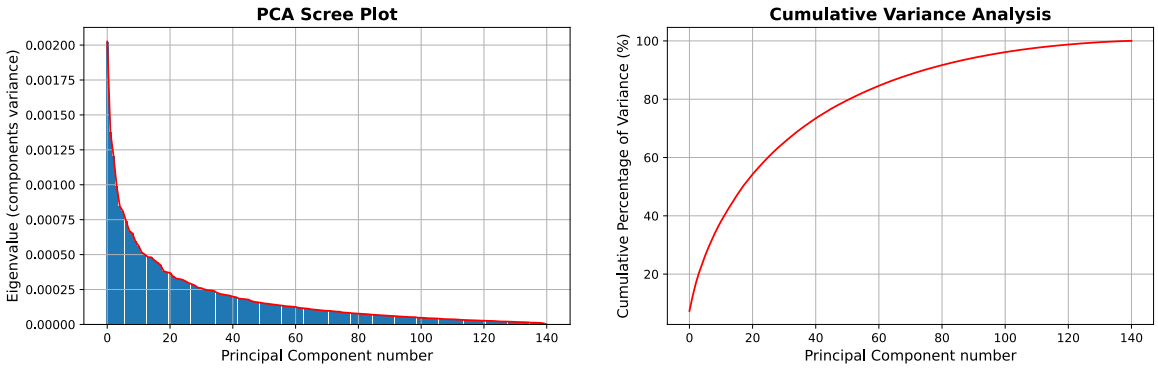


Figure 2.8: PCA analysis: principal components contribution (left) and cumulative variance analysis (right)

3 Portfolio Optimisation

3.1 Adaptive minimum-variance portfolio optimization

3.1.1 Optimal minimum variance portfolio

The optimal weights to construct the minimum variance portfolio can be obtained by solving the following optimization problem:

$$\begin{aligned} \min_{\mathbf{w}} \quad & J(\mathbf{w}, \mathbf{C}) = \frac{1}{2}\mathbf{w}^\top \mathbf{C}\mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^\top \mathbf{1} = 1 \end{aligned} \tag{44}$$

where $\mathbf{w}$ is the weights vector, $\mathbf{C}$ is the covariance matrix of all assets. It is important to mention that $\mathbf{C}$ is symmetric and invertible and $\mathbf{C}^{-1}$ is also symmetric. The expected return and variance of the portfolio are defined as:

$$\begin{aligned} \bar{\mu} &= \mathbf{w}^\top \mu \\ \bar{\sigma}^2 &= \mathbf{w}^\top \mathbf{C} \mathbf{w} \end{aligned} \tag{45}$$

The above constrained optimization problem is equivalent to optimizing the Lagrangian:

$$\min_{\mathbf{w},\lambda} \quad L(\mathbf{w}, \mathbf{C}, \lambda) = \frac{1}{2} \mathbf{w}^\top \mathbf{C} \mathbf{w} - \lambda(\mathbf{w}^\top \mathbf{1} - 1) \tag{46}$$

In order to solve the minimisation problem, the Lagrangian is differentiated with respect to $\mathbf{w}$ and λ and then both derivatives are set to zero to obtain the optimal parameters as follows:

$$\begin{aligned} 0 = \nabla_{\mathbf{w}} L = \mathbf{C}\mathbf{w} - \lambda \mathbf{1} \quad &\Rightarrow \quad \mathbf{w} = \lambda \mathbf{C}^{-1} \mathbf{1} \\ 0 = \nabla_{\lambda} L = \mathbf{w}^\top \mathbf{1} - 1 \quad &\Rightarrow \quad \mathbf{w}^\top \mathbf{1} = 1 \end{aligned} \tag{47}$$

Therefore substituting the first equation into the second one yields:

$$\begin{aligned} \left(\lambda \mathbf{C}^{-1} \mathbf{1}\right)^\top \mathbf{1} &= 1 \\ \lambda &= \frac{1}{\mathbf{1}^\top \mathbf{C}^{-1} \mathbf{1}} \end{aligned} \tag{48}$$

Hence the optimal weights are found to be:

$$\mathbf{w}^* = \lambda \mathbf{C}^{-1} \mathbf{1} = \frac{\mathbf{C}^{-1} \mathbf{1}}{\mathbf{1}^\top \mathbf{C}^{-1} \mathbf{1}} \tag{49}$$

The optimal expected return and variance of the portfolio are defined in Equation (45) can be computed using the optimal weights:

$$\begin{aligned} \bar{\mu}^* = \mathbf{w}^{*\top} \mu &= \frac{\mathbf{1}^\top \mathbf{C}^{-1}}{\mathbf{1}^\top \mathbf{C}^{-1} \mathbf{1}} \\ (\bar{\sigma}^2)^* = \mathbf{w}^{*\top} \mathbf{C} \mathbf{w}^* &= \frac{1}{(\mathbf{1}^\top \mathbf{C}^{-1} \mathbf{1})^2} \left(\mathbf{1}^\top \mathbf{C}^{-1} \ \mathbf{C} \ \mathbf{C}^{-1} \mathbf{1}\right) \\ &= \frac{\mathbf{1}^\top \mathbf{C}^{-1} \mathbf{1}}{(\mathbf{1}^\top \mathbf{C}^{-1} \mathbf{1})^2} = \frac{1}{\mathbf{1}^\top \mathbf{C}^{-1} \mathbf{1}} \end{aligned} \tag{50}$$

3.1.2 Minimum-variance portfolio analysis

Considering a dataset of 10 stocks, the aim is to compare the minimum-variance against the equally weighted portfolio strategy. The portfolio weights are trained on the first chronological half of the data (2017-2018), and then its performance tested on the other half of the set (2018-2019). Figure 3.1 shows the returns and cumulative returns (defined as $R[T] = \sum_{t=1}^{T} \bar{r}[t]$) for both the equally weighted and the minimum-variance portfolio on the training and testing sets. It can be observed that the cumulative returns on the train set are steadily increasing for both portfolios with minimal difference, while on the test set, the minimum variance portfolio is performing better

than the equally-weighted one, however presenting a down-trend. As seen from both Table 6 and Table 7 both portfolios are unable to achieve a positive cumulative return on the testing set. It can be suggested that increasing the training data can result in a better generalisation to the test set, so a different train-to-test ratio to the 50:50 split such as 80:20, could yield better results.

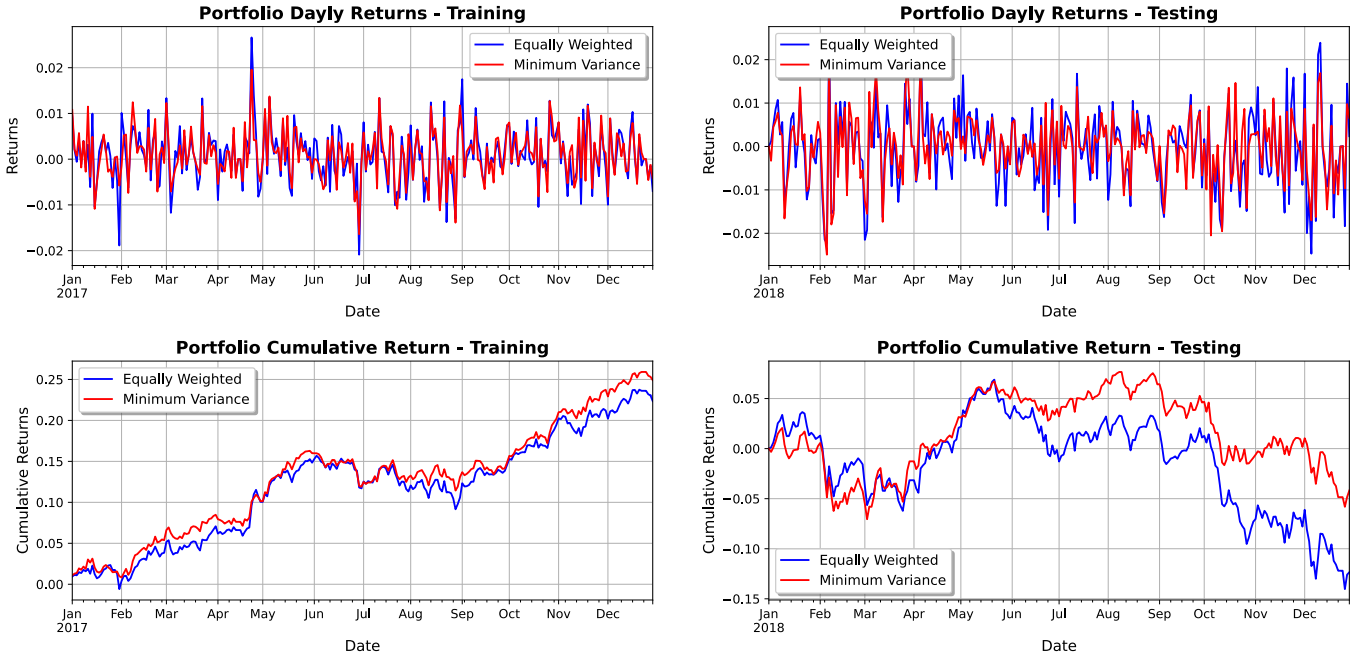


Figure 3.1: Portfolio returns and cumulative returns on training and testing set

Equally Weighted	Mean Return	Cumulative Return	Variance of Returns	Theoretical Variance	Sharpe Ration
Training	0.00086	0.223484	3.7495e-05	/	0.140374
Testing	-0.000473	-0.123503	7.93339e-05	/	-0.053126

Table 6: Equally weighted portfolio statistics

Minimum Variance	Mean Return	Cumulative Return	Variance of Returns	Theoretical Variance	Sharpe Ration
Training	0.000958	0.249062	2.86163e-05	2.86163e-05	0.179072
Testing	-0.00016	-0.041823	5.84487e-05	5.84487e-05	-0.02096

Table 7: Minimum-varaince portfolio statistics

As seen in Table 7 for the minimum-variance portfolio, the variance of returns obtained exactly matches the theoretical values computed using Equation (50) for both training and testing data.

3.1.3 Adaptive time-varying minimum variance portfolio analysis

Figure 3.2 displays the the cumulative returns for various rolling window sizes while Table 8 reports the portfolios statistics for the respective window sizes. It can be observed that as the window size increases, the timeframe considered when estimating the covariance matrix increase, and the performance of the portfolio decreases. As observed from the plots, shorter time windows, such as 22 or 44 days, are able to detected short-period patterns in the data and therefore perform better. The window size of 22 was the best performing amongst the ones analysed, achieving a positive trend and a final cumulutaive return of 0.0075. As seen in Figure 3.3, there is an improvement of performance for adaptive minimum variance portfolios compared to static minimum variance and equally weighted portfolios (for the testing data in the previous section). However, it is important to mention that in terms of variance of returns (volatility), it performs slightly worse (Table 8). This can be observed from the spikes in the plots that suggest higher volatility and therefore higher risk.

Adaptive MV window size (days)	**Mean Return**	**Cumulative Return**	**Variance of Returns**	**Sharpe Ration**
22	0.000286	0.074774	0.0001082415	0.027537
44	0.000233	0.06086	7.586e-05	0.026772
88	-0.00003	-0.007747	7.1112e-05	-0.00352
132	-0.000244	-0.063607	6.98857e-05	-0.029152
176	-0.000311	-0.081049	6.7946e-05	-0.037672
264	-0.000444	-0.115937	6.66552e-05	-0.054408

Table 8: Adaptive minimum-variance portfolio statistics

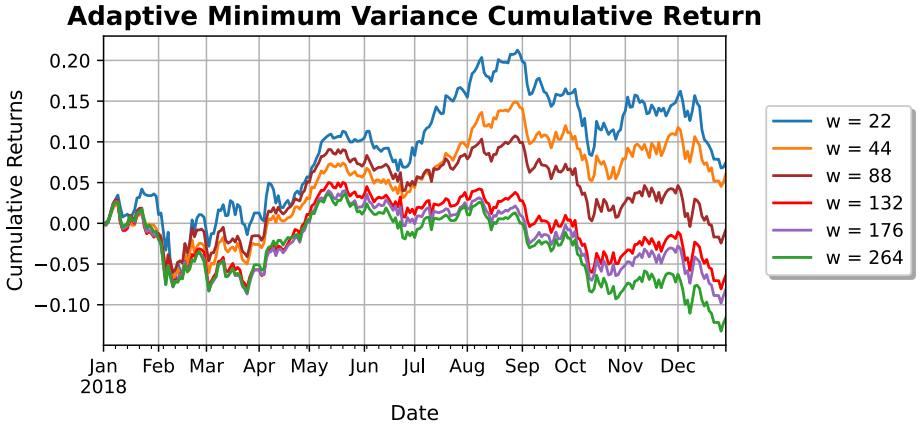


Figure 3.2: Cumulative returns for adaptive minimum-variance portfolios with different window sizes

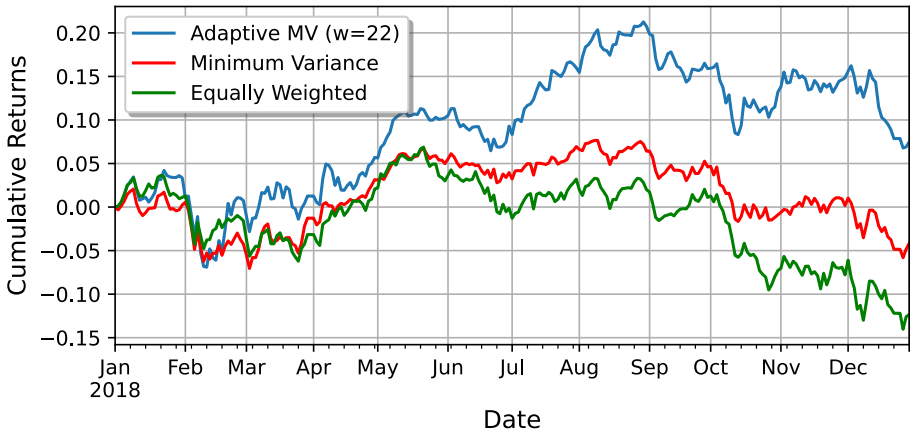


Figure 3.3: Cumulative returns of adaptive minimum-variance, minimum variance and equally weighted portfolios

The recursive update of the variables allows the model to respond to changing market conditions and maintain its minimum variance objective due to the shorter time periods considered. By continuously adjusting the portfolio weights based on the updated information about asset returns and covariances, the strategy can achieve better risk-adjusted performance compared to a static allocation approach.

A different method to compute the sample covariance matrixcould be implemented to improve performance. One method is to apply exponential weight decay by assigning higher weights to more recent observations and exponentially decreasing weights to older observations. This approach allows for a more dynamic estimation of the covariance matrix, as it gives more importance to recent data points, making it more responsive to recent market changes.

4 Robust Statistics and Non-Linear Methods

4.1 Data Imports and Exploratory Data Analysis

4.1.1 Key Statistics of AAPL, IBM, JPM, DJI

AAPL	Open	High	Low	Close	Adj Close	Volume
Mean	187.6867	189.5618	185.8237	187.7120	186.1743	32704750.199
Median	186.2900	187.4000	184.9400	186.1200	184.3518	29184000.000
StdDev	22.1456	22.2816	22.0088	22.1607	21.9047	14179721.593
MAD	15.8900	15.6100	15.9200	15.9400	15.4760	7573900.000
IQR	36.0000	36.3400	36.0600	36.7550	35.6854	16311700.000
Skew	0.2599	0.3004	0.2205	0.2638	0.2908	1.743
Kurtosis	-0.9126	-0.9246	-0.9176	-0.9324	-0.9280	4.353

IBM	Open	High	Low	Close	Adj Close	Volume
Mean	138.4544	139.4921	137.3292	138.3631	134.9028	5198937.450
Median	142.8100	143.9900	142.0600	142.7100	138.5664	4237900.000
StdDev	12.1143	11.9131	12.2046	12.0281	10.6716	3328955.530
MAD	5.2700	5.3100	5.1900	5.2300	4.4935	920700.000
IQR	15.3800	14.7200	16.3400	15.5050	14.1039	1952950.000
Skew	-0.6760	-0.6227	-0.7134	-0.6822	-0.8112	3.193
Kurtosis	-0.5853	-0.6236	-0.5620	-0.5840	-0.4209	11.797

JPM	Open	High	Low	Close	Adj Close	Volume
Mean	108.7077	109.6521	107.6830	108.6066	107.2626	14700689.243
Median	109.1800	110.5300	107.7900	109.0200	107.2193	13633000.000
StdDev	5.3591	5.2029	5.4325	5.3005	4.8333	5349770.564
MAD	4.4700	4.3100	4.2400	4.3500	3.4502	3035400.000
IQR	8.8100	8.8450	8.8450	8.8350	7.2224	6233600.000
Skew	-0.4208	-0.3762	-0.3775	-0.3749	-0.3445	1.693
Kurtosis	-0.3225	-0.5442	-0.2657	-0.3966	-0.1054	4.430

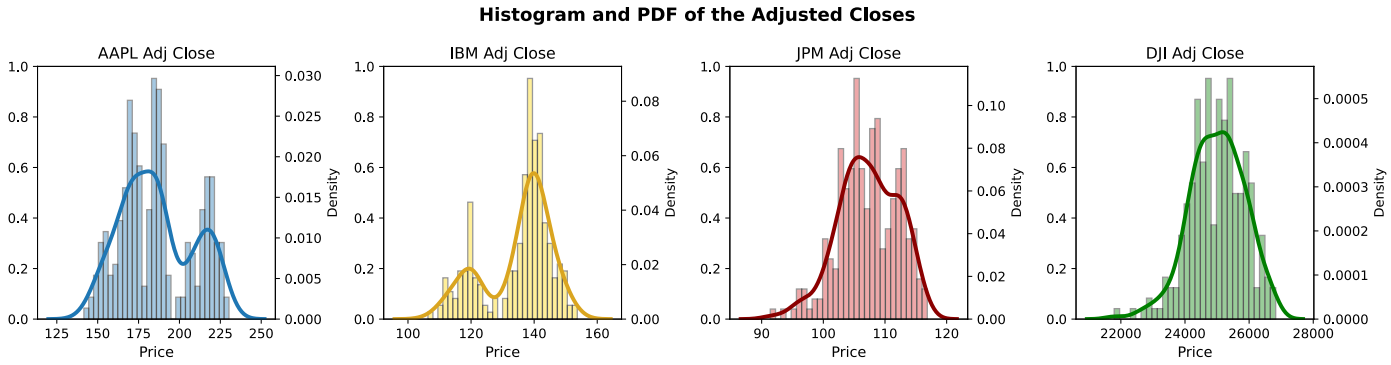
DJI	Open	High	Low	Close	Adj Close	Volume
Mean	25001.2573	25142.0420	24846.0022	24999.1536	24999.1536	332889442.230
Median	25025.5801	25124.0996	24883.0391	25044.2891	25044.2891	313790000.000
StdDev	858.8347	815.2040	903.3022	859.1321	859.1321	94078038.141
MAD	543.5410	537.6191	601.5684	590.7207	590.7207	50460000.000
IQR	1109.4346	1077.8164	1204.4189	1158.1553	1158.1553	108930000.000
Skew	-0.3721	-0.2394	-0.4564	-0.3801	-0.3801	1.740
Kurtosis	0.4857	0.1182	0.5576	0.4007	0.4007	5.858

Table 9: Key descriptive Statistics for the stocks AAPL, IBM, JPM and the index DJI

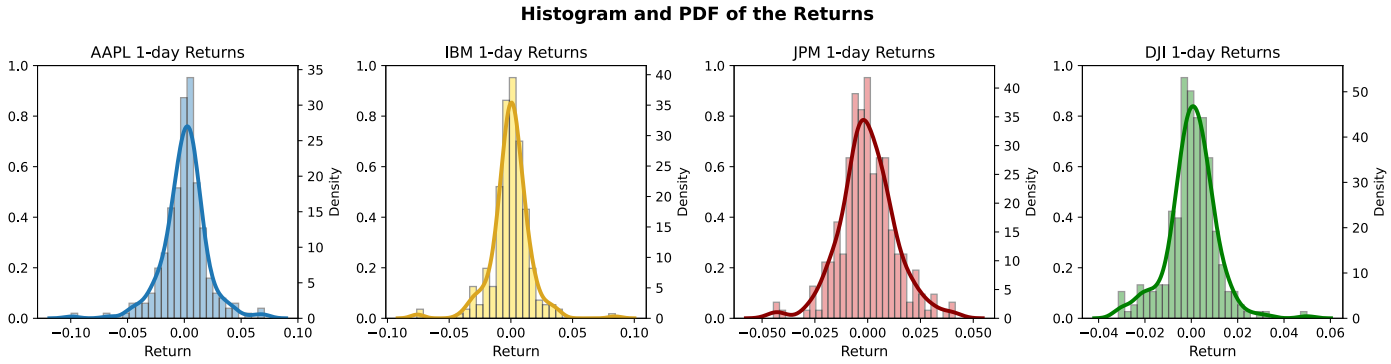
Three stocks: AAPL, IBM, JPM and one index: DJI are considered. For each asset the key descriptive statistics that summarize the distribution of the dataset are generated and reported in Table 9. The key statistics reported are: mean, median, which give an insight of the data tendencies, standard deviation, median absolute deviation (MAD) and interquartile range (IQR), which quantify data dispersion, skewness to measure the asymmetry and kurtosis to measure the difference of the tail to a Normal distribution.

4.1.2 Histograms and probability density functions

In Figure 4.1, the histogram and the corresponding fitted probability density functions (PDF) of the adjusted closing prices and the 1-day returns are plotted for each asset. Figure 4.1a shows that APPL and IBM are the most volatile stocks as they require a multi-gaussian (two gaussians) PDF to accurately fit the price evolution. On the other hand, JPM presents less volatility as the price distribution approaches a normal gaussian-like behaviour. This is expected as the technology sector presents an higher volatility than the financials one. Finally the DJI Index adjusted price follows a gaussian-like behaviour. As it was previously discussed in section 1.1.3, the returns tend to follow a Gaussian distribution (Figure 4.1b). It can be noticed that IBM and AAPL tend to have a higher variance than JPM thus further supporting the reasoning presented before. Additionally DJI returns curve shows stronger signs of asymmetry, as it presents a bigger tail in the negative side, thus meaning that lower returns are more likely than higher ones.



(a) Adj-close prices



(b) 1-day returns

Figure 4.1: probability density function for AAPL, IBM, JPM, DJI

4.1.3 Rolling Mean and Median Analysis

Figures 4.3 and 4.4 show two Z-score based methods to detect outliers of the adjusted closing prices of each asset. The first approach employs the rolling mean and standard deviations of the assets to classify prices as outliers that lie outside the range defined by $\pm 1.5 \times$ standard deviations relative to the rolling mean. The second method uses the median and median absolute deviation (MAD) to create the range $\pm 1.5 \times$MAD relative to the median. These regions for the two methods are represented by the faded yellow surface in the plots. It can be noticed that the rolling mean and median plot have a very similar behaviour. However the range of the rolling deviations related to the mean is wider than the median method, which leads to less outliers being detected. This is due to the mean and standard deviation being less robust to sudden price changes than the descriptive statistics used in the median method. This can be further observed in Table 10 which confirms that the mean method detected less outliers points outside that deviation range as it is less constrained.

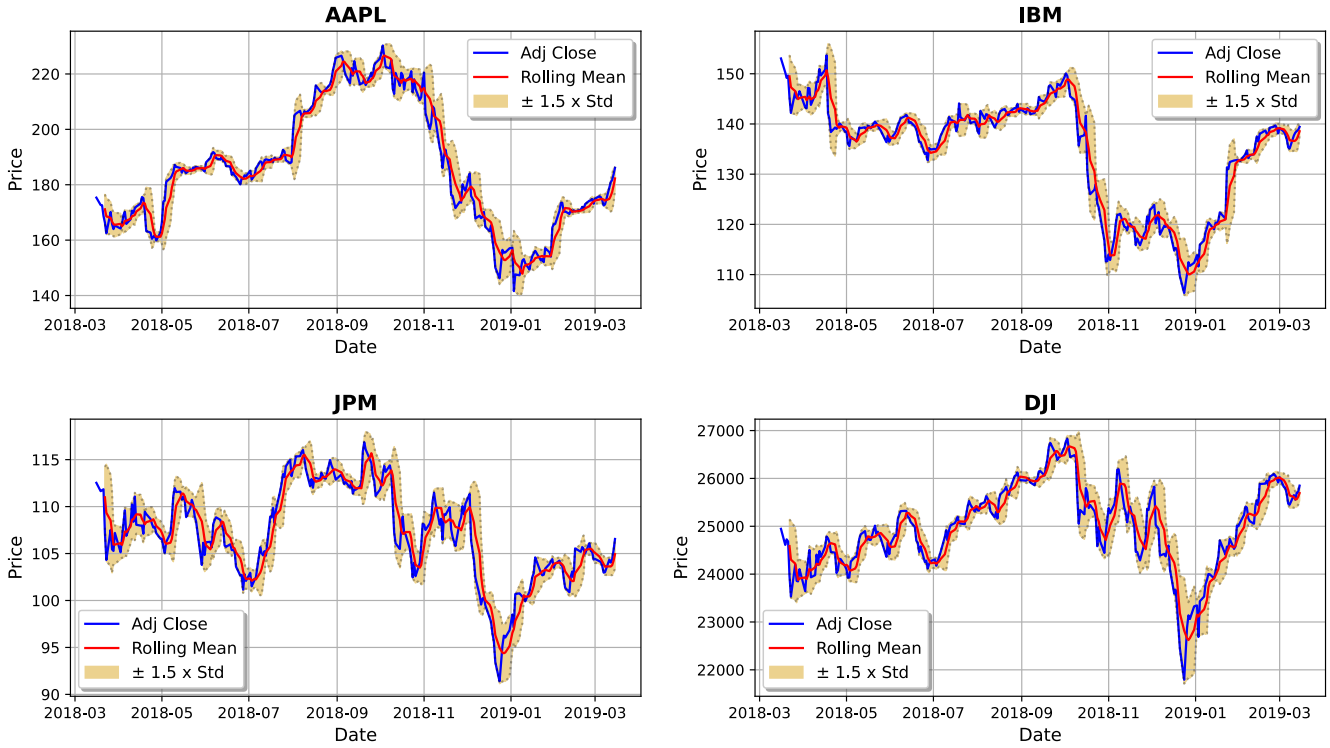


Figure 4.2: Rolling Mean (using a 5-day window) with $\pm 1.5 \times$ Sandard Deviation

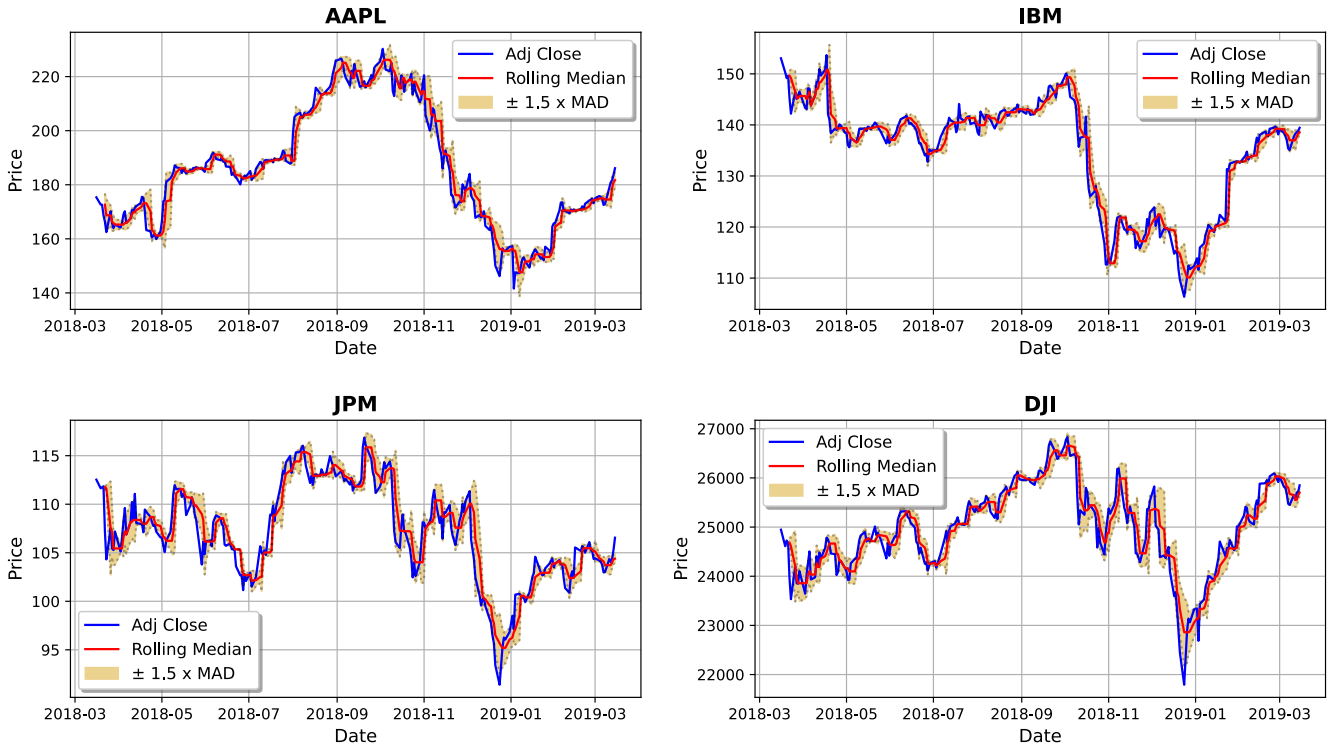


Figure 4.3: Rolling Median (using a 5-day window) with ±1.5×MAD

4.1.4 Rolling Mean and Median Analysis with outliers

Four outlier points for the adjusted closing prices were introduced in the the dates {14-05-2018, 14-09-2018, 14-12-2018 and 14-01-2019} with a value equal to 1.2× the maximum value of the column. It can be observed from the plots in Figures 4.4 and 4.5 that the deviation range of the mean method is highly affected whereas the median z-score range is largely unchanged by the introduction of large artificial outliers. It can also be concluded that the rolling median method with with the median absolute deviation outperformed the rolling mean with standard deviation method on highly noisy/corrupted data. Thus confirming the vulnerability to abrupt changes of the mean method. The numbers of outliers detected for the two methods are reported in Table 10.

	Mean Method	**Median Method**	**Mean Method with Outliers**	**Median Method with Outliers**
AAPL	30	103	32	102
IBM	31	94	31	93
JPM	33	105	33	101
DJI	30	97	29	96

Table 10: Number of outlier points for each asset using the two methods with and without added outliers

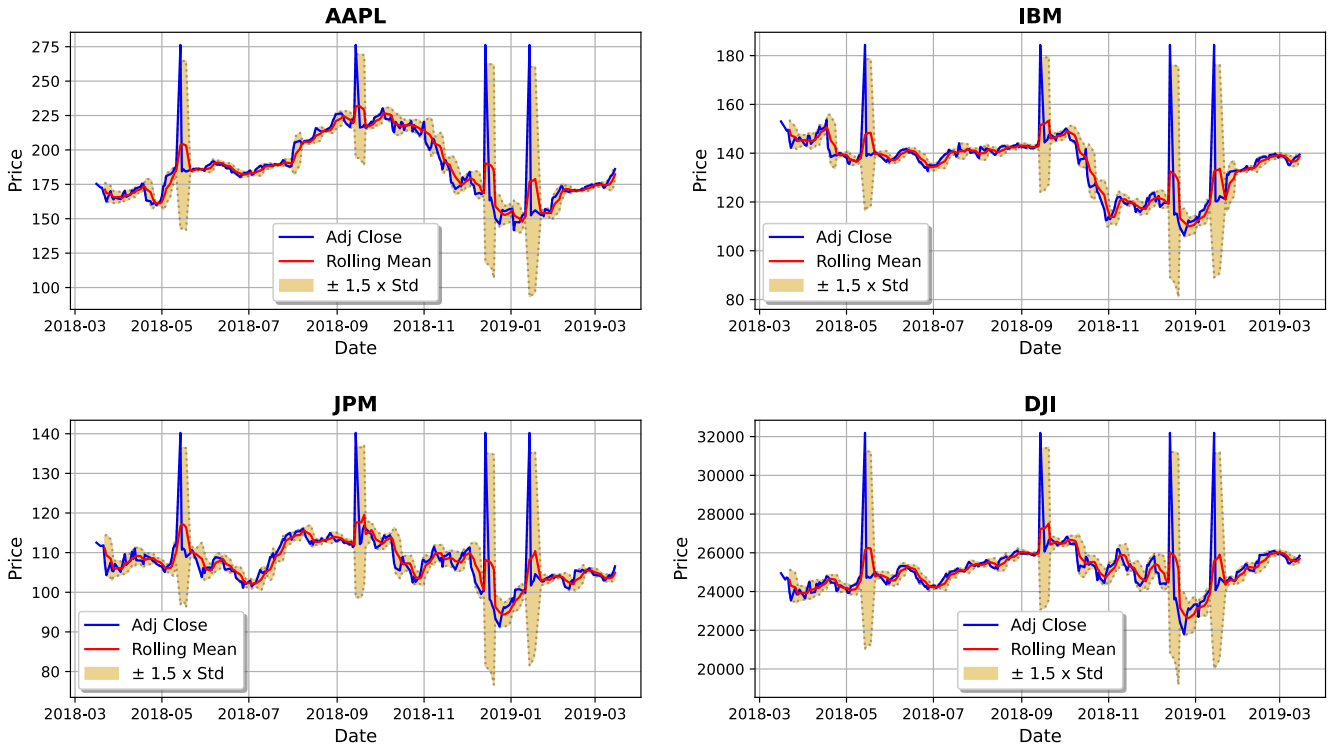


Figure 4.4: Rolling Mean outlier detection (using a 5-day window) with ±1.5 × Sandard Deviation

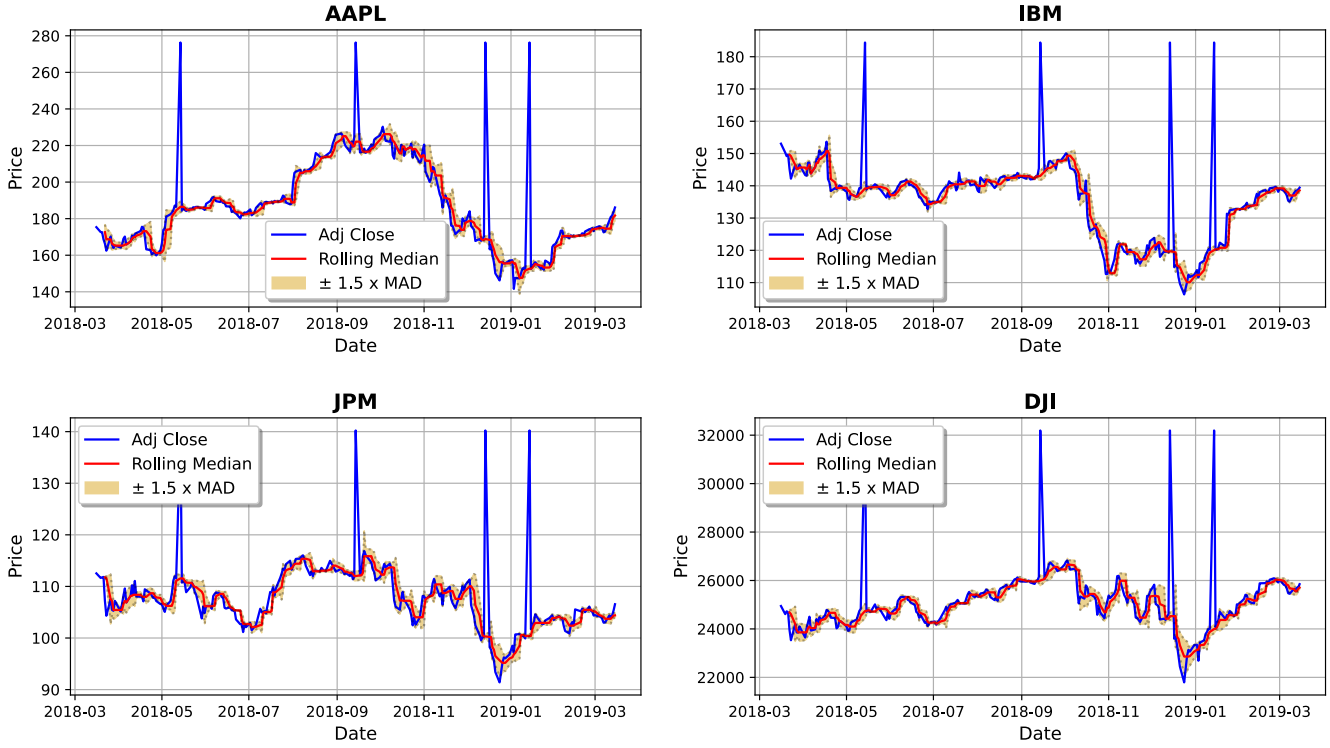


Figure 4.5: Rolling Median outlier detection (using a 5-day window) with ±1.5×MAD

4.1.5 Box Plots Analysis

A box plot is a method for graphically illustrating the descriptive statistics and demonstrating the locality, spread, and skewness of the data. Looking at any plot in Figure 4.6 the black vertical line in the box represents the median while the left and right boundaries of the box represent the first and third quartile respectively, thus the blue shaded area denotes the interquartile range (IQR). The left and right vertical lines away from the box indicate the minimum and maximum respectively, whereas the red dots outside of the plot represent the outliers.

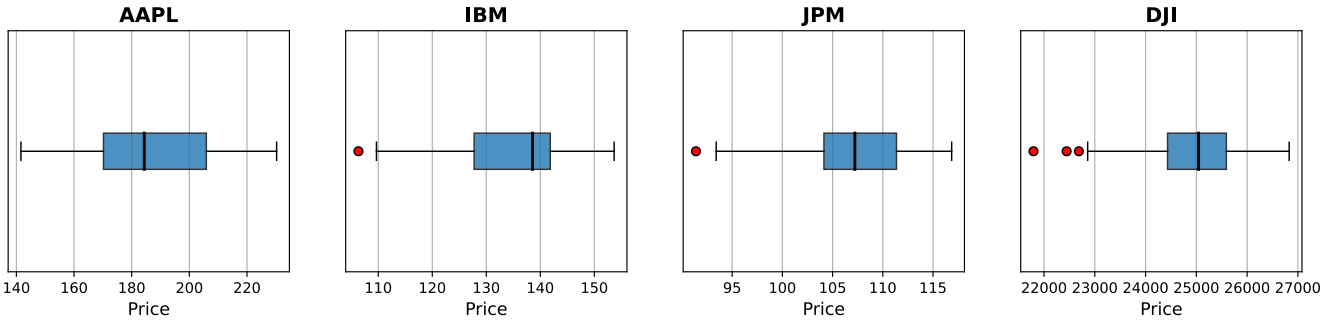


Figure 4.6: Box Plots for each asset

By examining the box plots we can compare the range and distribution of the prices for each asset. It can be observed that there is a greater variability for AAPL and IBM prices as they display the largest IQRs, thus implyaing high volatility of the assets as well as larger outliers. It can also be observed that none of the assets are normally distributed as the median is not in the centre of the box, except for DJI that presents Gaussian-like behaviour as previously mention in Section 4.1.2. Moreover the asymmetric division of the boxes denotes the asymmetry in the two tails of the distributions thus skewness. It can concluded that all assets present some level of skewness with IBM's being the largest and most negative, thus suggesting lower prices than the median.

4.2 Robust Estimators

4.2.1 Estimators implementations

```
# Robust location estimator:
def median_estimator(data):
    # Sort the data in ascending order
    sorted_data = data.sort_values()
    n = len(sorted_data)
    # If the dataset has an even number of elements,
    # calculate the median as the average of the two middle elements
    if n % 2 == 0:
        median = (sorted_data.iloc[n//2-1]
                  + sorted_data.iloc[n//2])/2
    # If length of the data is am odd number,
    # the median is the middle element
    else:
```

```
        median = sorted_data.iloc[n//2]
    return median

# Robust scale estimators:
def iqr_estimator(data):
    sorted_data = data.sort_values()
    n = len(sorted_data)
    # If the length of the data is divisible by 4
    # calculate the first (Q1) and third (Q3) quartiles
    # using exact quartile position
    if n % 4 == 0:
        q1 = sorted_data.iloc[n//4]
        q3 = sorted_data.iloc[(3*n)//4]
    # Else, calculate the first (Q1) and third (Q1) quartiles
    # using the average of two closest quartile positions
    else:
        q1 = (sorted_data.iloc[n//4-1] + sorted_data.iloc[n//4])/2
        q3 = (sorted_data.iloc[(3*n)//4-1] +
              sorted_data.iloc[(3*n)//4])/2
    # Compute the IQR as the difference between Q3 and Q1
    iqr = q3 - q1
    return iqr

def mad_estimator(data):
    # compute the magnitude of the data deviation from the median
    mad = median_estimator(abs(data - median_estimator(data)))
    return mad
```

4.2.2 Computational Efficiency Analysis

The computational complexity of `median_estimator` can be determined by analysizing its individual components. Sorting the data using the Pandas built-in function `sort_values` that applies a quicksort algorithm, has a complexity of $\mathcal{O}(n\ log(n))$, where n is the length of the `pandas.Series`. The other operations have complexity $\mathcal{O}(1)$ because accessing specific elements in a `pandas.Series`, performing arithmetic (modulus) and the built-in `len` function that accesses the length of the series, are constant-time operations. Therefore the overall computational complexity is $\mathcal{O}(n\ log(n))$.

Similarly to the `median_estimator`, the `iqr_estimator` requires sorting using quicksort, uses `len`, performs arithmetic operations and accesses specific elements in a `pandas.Series`. Hence, the overall computational complexity is the same: $\mathcal{O}(n\ log(n))$.

The `mad_estimator` requires finding the median which has complexity of $\mathcal{O}(n\ log(n))$, as previously discussed. Then, subtracting the median from the data and taking the absolute values of the resulting `pandas.Series` takes $\mathcal{O}(n)$ time, as each element in the list has to be visited once. In the end the `median_estimator` is called again on the absolute deviations which has again $\mathcal{O}(n\ log(n))$ complexity. Therefore the overall complexity is $\mathcal{O}(n\ log(n)) + \mathcal{O}(n\ log(n)) + \mathcal{O}(n)$ which can which can be simplified to $\mathcal{O}(n\ log(n))$ complexity.

4.2.3 Breakdown Points

Breakdown point analysis is a statistical technique used to evaluate the robustness of an estimator in the presence of outliers. The breakdown point of an estimator refers to the percentage of outliers that the estimator can tolerate before it completely breaks down and gives incorrect results.

The breakdown point of the median is 50%, meaning that up to 50% of the data can be contaminated or replaced with extreme values before the median becomes an arbitrary or excessively large value. This high breakdown point makes the median a robust statistic that is resistant to the influence of outliers.

Similarly, the IQR needs to estimate the 25th and 75th percentiles, which is equivalent to splitting the data in half and finding the median of each half series. Therefore the breakdown point of the IQR is 25%. Hence a quarter of the data can be contaminated or replaced before the estimator breaks down.

Finally, since MAD is based on the median (calculated by finding the median of the absolute deviations of each data point from the overall median of the dataset), it inherits the median's robustness and resistance to outliers. Therefore the breakdown point is 50%.

4.3 Robust and OLS regression

The Dow Jones Industrial Average (DJI) is a market index that tracks the daily price movements of 30 large-cap stocks in the US market, among which Apple (AAPL), IBM and J.P. Morgan (JPM). Therefore in this section, the two-way relationship between the index and any of its underlying stocks will be studied by regressing the 1-day returns of each individual stock against the 1-day returns for DJI using OLS and Huber regressions.

4.3.1 Ordinary Least Squares (OLS) Regression

The regression problem solved with OLS is defined by:

$$\mathbf{r} = \mathbf{A}\mathbf{x} + \mathbf{e} \tag{51}$$

where $\mathbf{r}$ is the vector of returns of the stock, $\mathbf{A}$ is a matrix containing one column with ones and one with the DJI returns, $\mathbf{e}$ is the error and $\mathbf{x}$ is the coefficient vectors that relates the returns of the stock with the ones of DJI.
OLS regression seeks to minimize the mean squared error expressed as:

$$\min_{\mathbf{x}} ||\mathbf{e}||^2 = ||\mathbf{r} - \mathbf{A}\mathbf{x}||^2 \tag{52}$$

where the optimal solution can be computed to be

$$\mathbf{x}^* = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{r} \tag{53}$$

However, in the case of an outlying point, the error that is objective to minimization will be very large. This high sensitivity to data corruption affects significantly the performance of the underlying regression model.

The regression parameters obtained for each stock are given in Table 11 and the return predictions as well as the regression linear approximations plots are given in Figure 4.7.

	OLS α (Intercept)	**OLS β (coefficient)**	**OLS MSE**
AAPL	0.000165	1.32558	0.00018
IBM	-0.000441	0.960092	0.00014
JPM	-0.000316	0.931408	0.000076
DJI	0.0	1.0	/

Table 11: Parameters and MSE of OLS regression

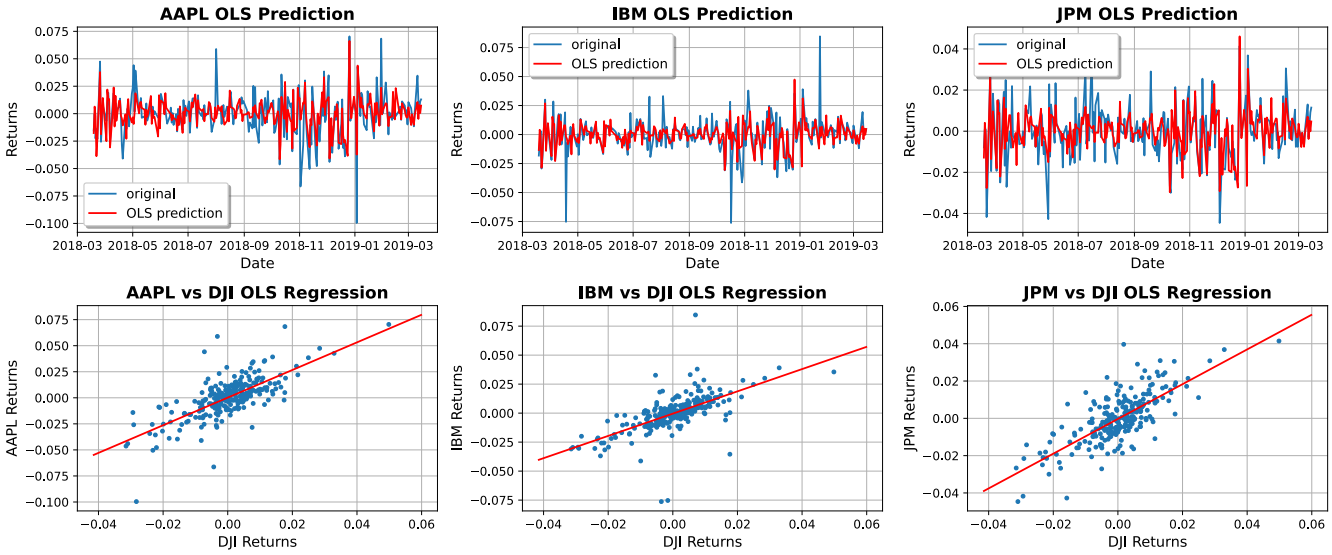


Figure 4.7: Stocks 1-day returns predictions (top) and stocks vs DJI returns regressions (bottom) using the OLS method

4.3.2 Huber (robust) Regression

Huber regression improves on the OLS method by transforming the minimization problem if the error values exceed a threshold ϵ. Thus limiting the overshoot of the error caused by an outlier.

$$\min_{\mathbf{x}} \begin{cases} ||\mathbf{e}||^2 = ||\mathbf{r} - \mathbf{A}\mathbf{x}||^2, & \text{for } \frac{||\mathbf{r}-\mathbf{A}\mathbf{x}||}{\sigma} < \epsilon \\ ||\mathbf{e}|| = ||\mathbf{r} - \mathbf{A}\mathbf{x}||, & \text{for } \frac{||\mathbf{r}-\mathbf{A}\mathbf{x}||}{\sigma} \geq \epsilon \end{cases} \tag{54}$$

The parameter ϵ represents the value after which a point is considered an outlier. Therefore by imposing a threshold ϵ and modifying the objective function, the impact of the outlier is reduced as they are on a linear scale rather than quadratic. Thus making the regression method more robust.

Figure 4.8 shows the return predictions and the regression linear approximations plots obtained with the Huber method while the model parameters obtained for each stock are given in Table 12.

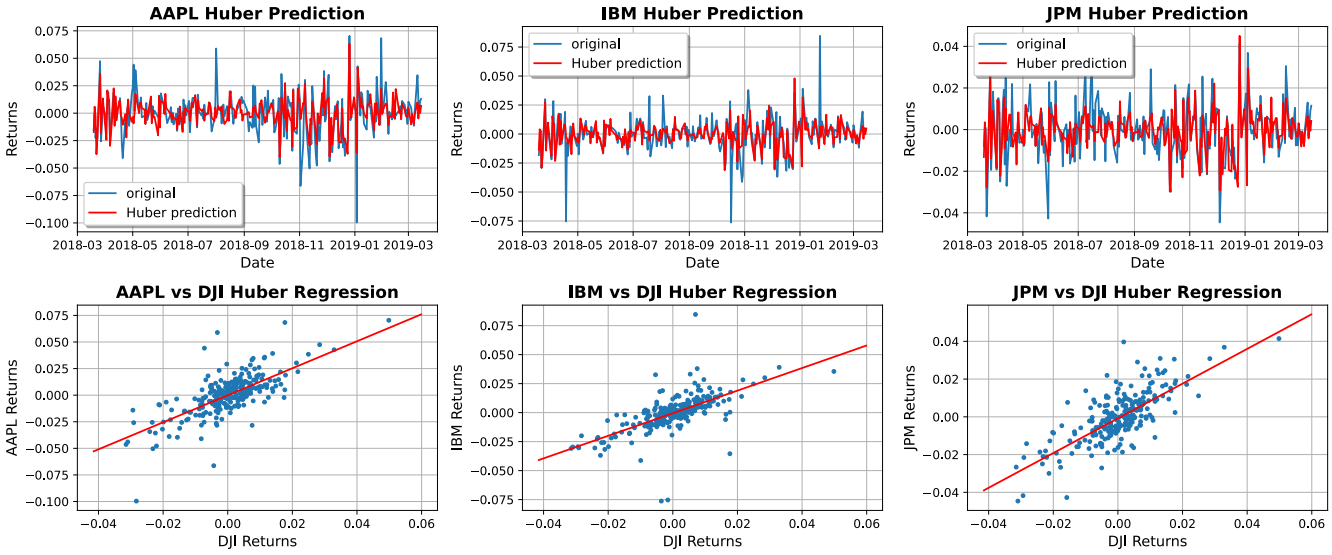


Figure 4.8: Stocks 1-day returns predictions (top) and stocks vs DJI returns regressions (bottom) using the Huber method

	Huber α (Intercept)	**Huber β (coefficient)**	**Huber MSE**
AAPL	-0.00013	1.270212	0.00018
IBM	-0.000509	0.973562	0.00014
JPM	-0.000801	0.919662	0.000076
DJI	0.0	1.0	/

Table 12: Parameters and MSE of Huber regression

4.3.3 Performance assessment of both regression methods

When looking at Figures 4.7 and 4.8, it is seen that both the OLS and Huber regression gave very similar results, with negligible differences. Therefore, to further test and compare the regression methods, the dataset is corrupted with 4 artificial outliers at certain dates with values equivalent to 1.2 times the maximum value of the returns.

Figure 4.9 shows the return predictions and the regression linear approximations plots obtained with both methods. From these plots, the predictions of both regressors with the outlier points are very similar. However, by comparing the values obtained in Table 11 and 12 for the regression parameters without outliers with the new ones in Table 13, it can be observed that the Huber regression displays a more robust behavior to the abrupt changes in returns as the model parameters remain relatively constant. There is a more significant impact on the OLS regressor coefficients.

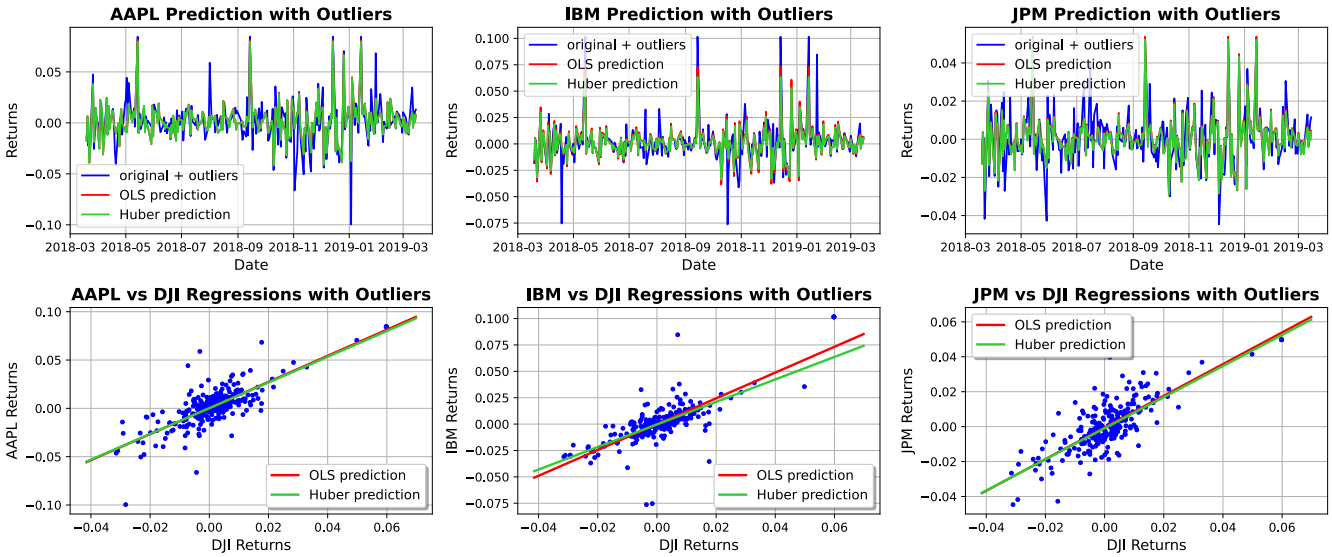


Figure 4.9: Returns predictions and regressions using OLS and Huber methods applied on data with outliers

	OLS with Outliers, α	OLS with Outliers, β	OLS with Outliers, MSE	Huber with Outliers, α	Huber with Outliers, β	Huber with Outliers, MSE
AAPL	0.00035	1.350465	0.000179	0.000122	1.330773	0.000179
IBM	-0.000063	1.22204	0.00016	-0.000339	1.066643	0.000164
JPM	-0.000467	0.906468	0.000075	-0.000933	0.890359	0.000075
DJI	0.0	1.0	/	0.0	1.0	/

Table 13: Parameters and MSE of both OLS Huber regression on data with outliers

4.4 Robust Trading Strategies

4.4.1 Moving Average Crossover

Figure 4.10 displays the rolling mean moving averages: 20-day and 50-day MA as well as the buy (green) and sell (red) regions which are colored following the crossover trading strategy. By looking at the buying and selling regions in the plot, it is clear that the presence of artificial outliers has a significant impact on the crossover strategy. This is an expected behavior since the mean is not a robust descriptive statistic and is susceptible to abrupt changes in the data. This is further confirmed by the results reported in the first column of Table 14 which shows the percentage similarities between the MA strategies for the real and the corrupted prices. It is reported that the lowest similarity among the considered assets was obtained with IBM's prices (83%), suggesting that the applied MA strategy was the most influenced by the outliers (it can also be observed by the two plots in the second row of Figure 4.10).

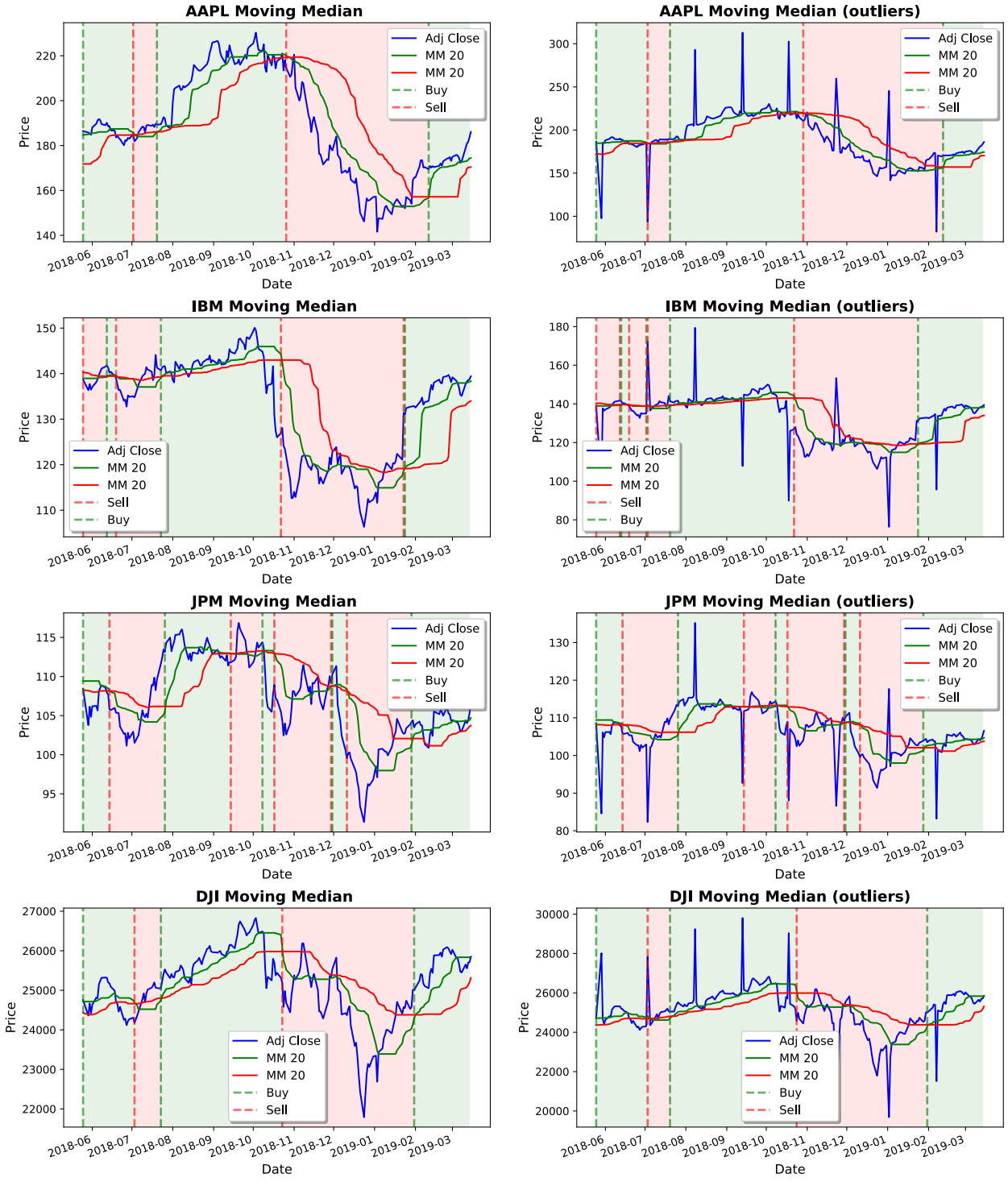


Figure 4.10: Moving Average Crossover strategy for the Adj. close values of each asset with and without added artificial outliersy

4.4.2 Moving Median Crossover

Figure 4.11 displays the crossover strategy applied to the same real and corrupted data for each asset, using the rolling median as the descriptive statistic. It can be observed that the moving median strategy applied to the corrupted data presents more

overlapping to the real data, thus providing almost identical buy and sell decisions. This is further supported by the second column in Table 14 which reports over 98% of similarity between the moving median strategy applied on real and corrupted data for all assets. Therefore it is evident that the moving median trading strategy is far more robust to outliers than the moving mean one.

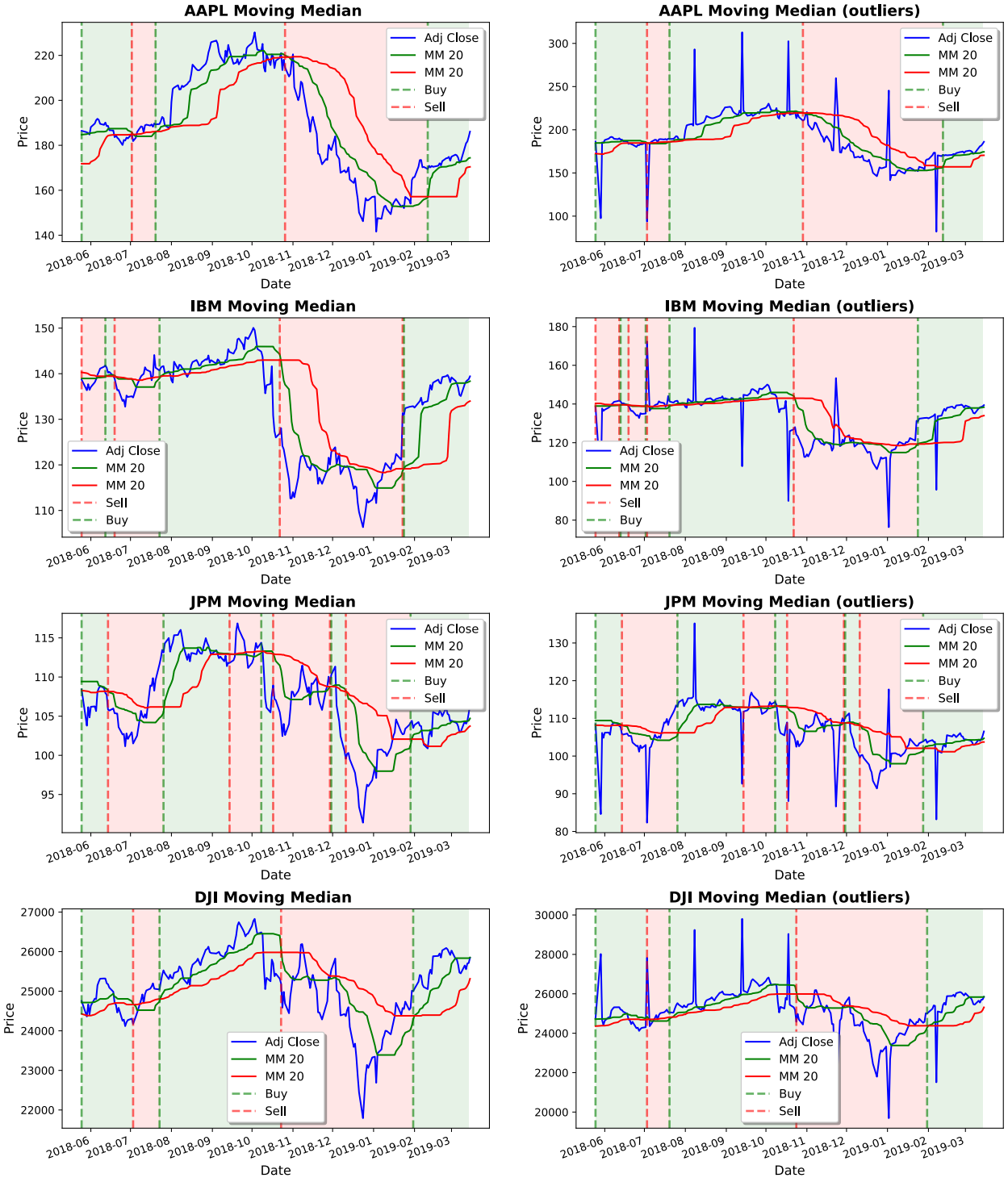


Figure 4.11: Moving Median Crossover strategy for the Adj. close values of each asset with and without added artificial outliers

	Moving Average similarity (%)	Moving Median similarity (%)
AAPL	93.56	98.51
IBM	83.66	98.02
JPM	90.10	99.50
DJI	98.51	99.01

Table 14: Similarity of crossing regions between the adjusted prices with and without outliers for each trading strategy

5 Graphs in Finance

5.1 SP 500 Stock Selection

The 10 selected stocks from the SP 500 are from the GICS sector of "Financials" and with their headquarters in the East Coast of the US. The overall motivation for selecting these stocks is their strong market position, well-established brands, and diversified product and service offerings within the financial sector. These companies represent a mix of banking, investment services, credit, insurance, and asset management firms. They benefit from a combination of factors such as economic growth, increasing consumer spending, and the global trend toward digital payments. Table 15 shows the information relative to selected companies.

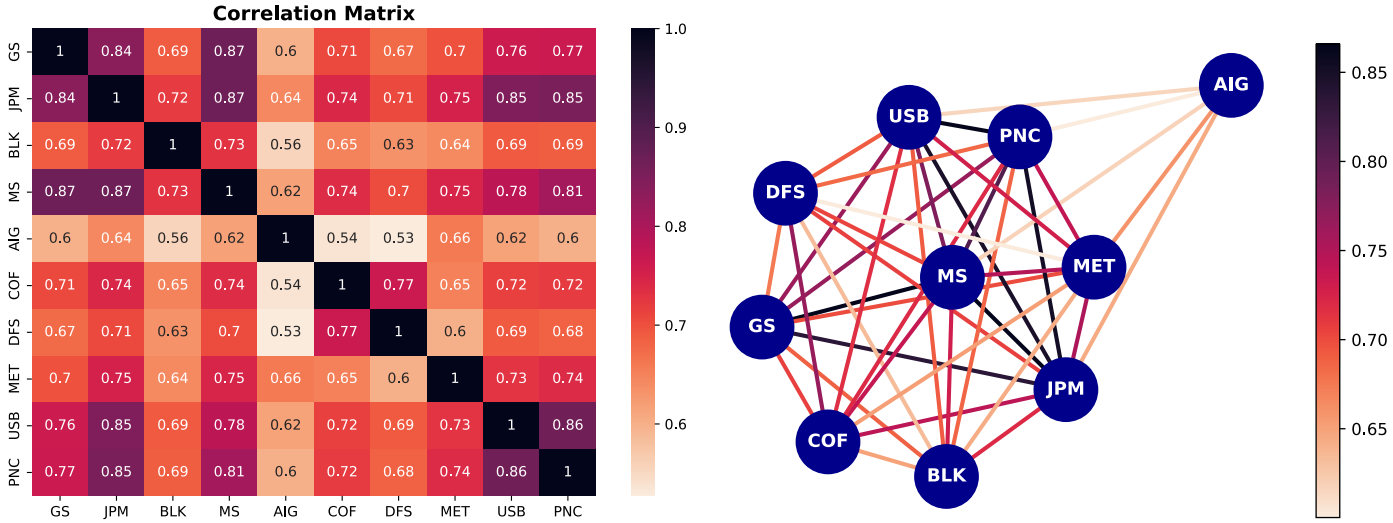
Symbol	Security	GICS Sector	GICS Sub Industry
GS	Goldman Sachs Group	Financials	Investment Banking & Brokerage
JPM	JPMorgan Chase & Co.	Financials	Diversified Banks
BLK	BlackRock	Financials	Asset Management & Custody Banks
MS	Morgan Stanley	Financials	Investment Banking & Brokerage
AIG	American International Group	Financials	Property & Casualty Insurance
COF	Capital One Financial	Financials	Consumer Finance
DFS	Discover Financial Services	Financials	Consumer Finance
MET	MetLife Inc.	Financials	Life & Health Insurance
USB	U.S. Bancorp	Financials	Diversified Banks
PNC	PNC Financial Services	Financials	Regional Banks

Table 15: Selected companies of the SP 500 in the Financials sector

5.2 Graphs based on Correlation Matrix

Figure 5.1a shows the correlation matrix for the selected 10 stocks. As expected, all assets are auto-correlated (shown on the diagonal with the value of 1), while the cross-correlation between assets varies depending on relationships between different stocks in terms of their returns movements. it can be noted that the as the correlation increases the color becomes darker.

Graphs in finance are a visualisation tools to represent the relationships between assets. Given that the correlation provides a quantitative indication of the relationships between two assets assets, it is possible to construct graph of all 10 stocks based on the correlation matrix (Figure 5.1b). It can be observed that the edges connecting the node (stocks) are colour-coded based on their correlation, with darker colors indicating stronger correlations between assets. The figure shows the edges with correlation weights more than 0.6, except for self-correlation.



(a) Correlation Matrix

(b) Network with minimum correlation of 0.6

Figure 5.1: Correlation Matrix and its corresponding Network

5.3 Correlation graph analysis

From Figure 5.1 and the graph features previously introduced, it is possible to identify the underlying relationships among the companies and therefore potential patterns and associations between certain firms. Companies such as Morgan Stanley (MS), J.P. Morgan (JPM) Goldman Sachs (GS) and PNC, hold a central position and show strong correlation amongst themselves as well as numerous connections with other companies reflecting the diversified nature of their business operations. Since they have divisions such as investment banking, asset management, brokerage, they are correlated with most companies in the Financials sector. On the other hand, companies are less linked with the others given their business activities such such as American International Group (AIG) that is related to insurance, thus it less prone to interactions with the other firms and exhibits lower correlation. Additionally, this is also an indicator that their market presence might not be as substantial as the other companies'.

Additionally, from Figure 5.2 it is possible to notice that re-ordering of the graph vertices or the re-ordering of the timer-series would not affect the structure of the graph. This is because the graph is based on the correlation matrix and correlation is independent of data reordering since it measures the similarity of the assets. Hence, re-ordering the vertices would change the positions of the stocks in the matrix but not their values, thus the graph would is unchanged.

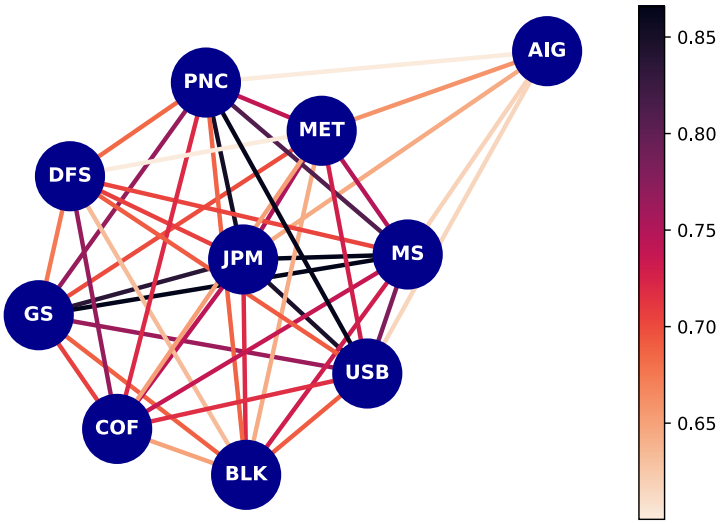
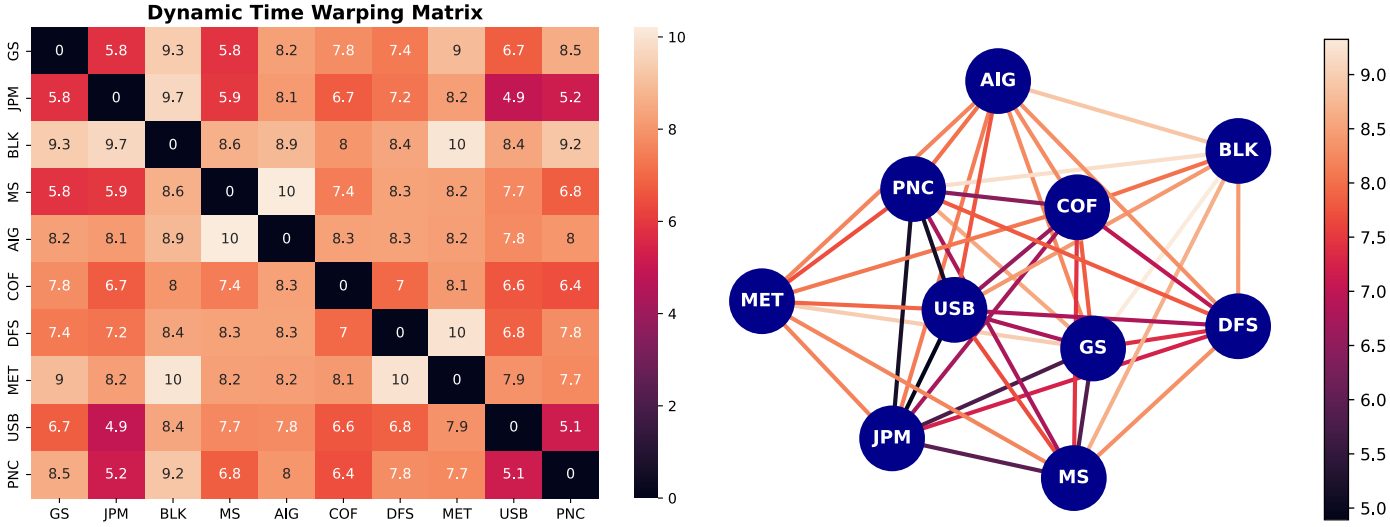


Figure 5.2: Correlation network after time-series data shuffling

5.4 Graphs based on Dynamic Time Warping Matrix

An alternative metric that can be used other than correlation to construct graphs is Dynamic Time Warping (DTW). This is a method provides the similarity between two temporal sequences, which may not be aligned in time. When using euclidean distance as a measure, DTW calculates the shortest (or least cost) path that matches the points of the two series. The flexibility of DTW allows to match sequences of different lengths and find optimal alignments even if the sequences have different speeds or are shifted in time, thus making it a suitable metric for time-series.



(a) Dynamic Time Warping Matrix

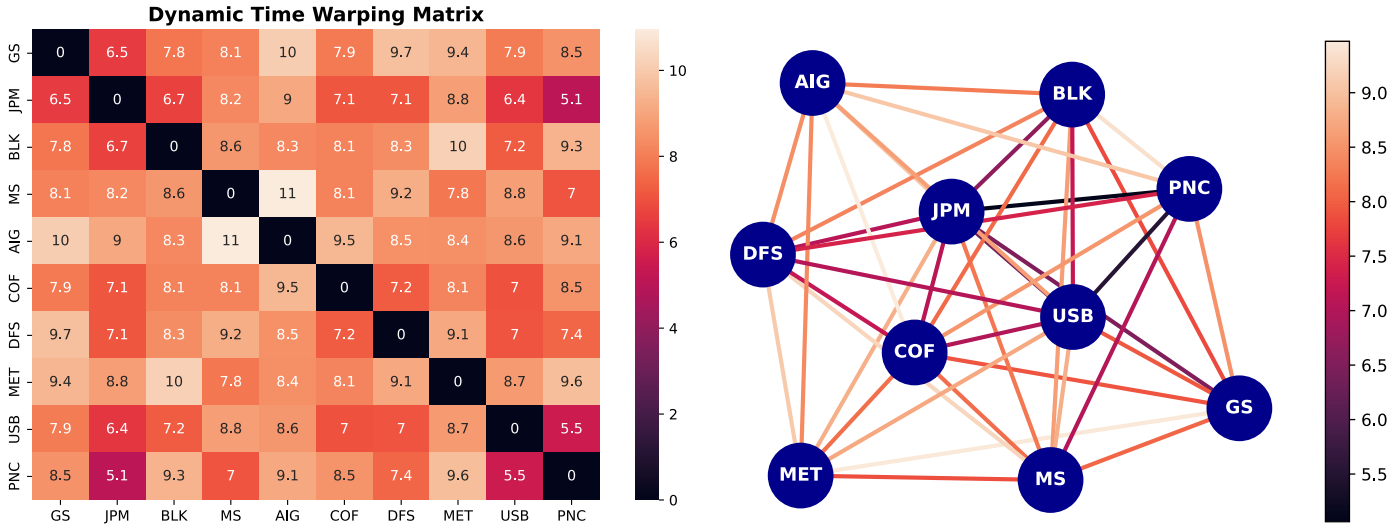
(b) Network with minimum distance of 10

Figure 5.3: Dynamic Time Warping Matrix and its corresponding Network

The DFT matrix for these companies is shown in Figure 5.3a. Notice that in this section the darker the color, the shorter the distance and the higher the similarity

between two companies. Thus maximum distance is 0 which is shown in the diagonal as the distance to the asset's self. It's graphical interpretation for assets whose distance do not surpass 9.5, is seen in Figure 5.3b, where darker edges indicate stronger similarities between assets. Once again, it is possible to observe companies like GS, JPM, MS, and PNC being more connected than AIG as in Section 5.3.

Contrary to correlation, DTW is dependent on the time-series data order, therefore reshuffling the assets returns would change the resulting graph. As seen from the DFT matrix and resulting graph in Figure 5.4 the relationships between the stocks changed drastically.



(a) Dynamic Time Warping Matrix

(b) Network with minimum distance of 10

Figure 5.4: Dynamic Time Warping network after time-series data shuffling

5.5 Raw Prices analysis

In the previous sections, log returns were used instead of raw prices. As it was discussed in Section 1, raw prices are not stationary and often contain trending components for a given asset, often upward. Therefore the graphs would represent the relationships between the upward trends of the stocks, which would have more correlated values, rather than finding patterns in price fluctuations within the sector. The results obtained would contain little information about the general patterns and would not present the intrinsic relationships and connections between the companies.