Imperial College London

Department of Electrical and Electronic Engineering

Financial Signal Processing and Machine Learning Coursework

Author:
Baptiste PROVENDIER

Lecturer:
Prof. Danilo MANDIC

CID:
01553706

April 2022

Contents

1 Regression Methods

1.1 Processing Stock Price Data in Python

1.1.1 Natural-log transform

In Figure 1, we plot the price of the SPX Index (from 1930 to 2017) against time along with the natural logarithmic transform of the time series. Logarithmic transformation is often used for time series to stabilise the variance of the data while preserving the relative order of the values as it is a monotonic function. Trends are generally represented more clearly by compressing the range of the data as we can see in Figure 1. We observe the general upward trend more clearly and can also see regional trends, such as the crash of the index in the early 1930s, that was not visible with the normal data.

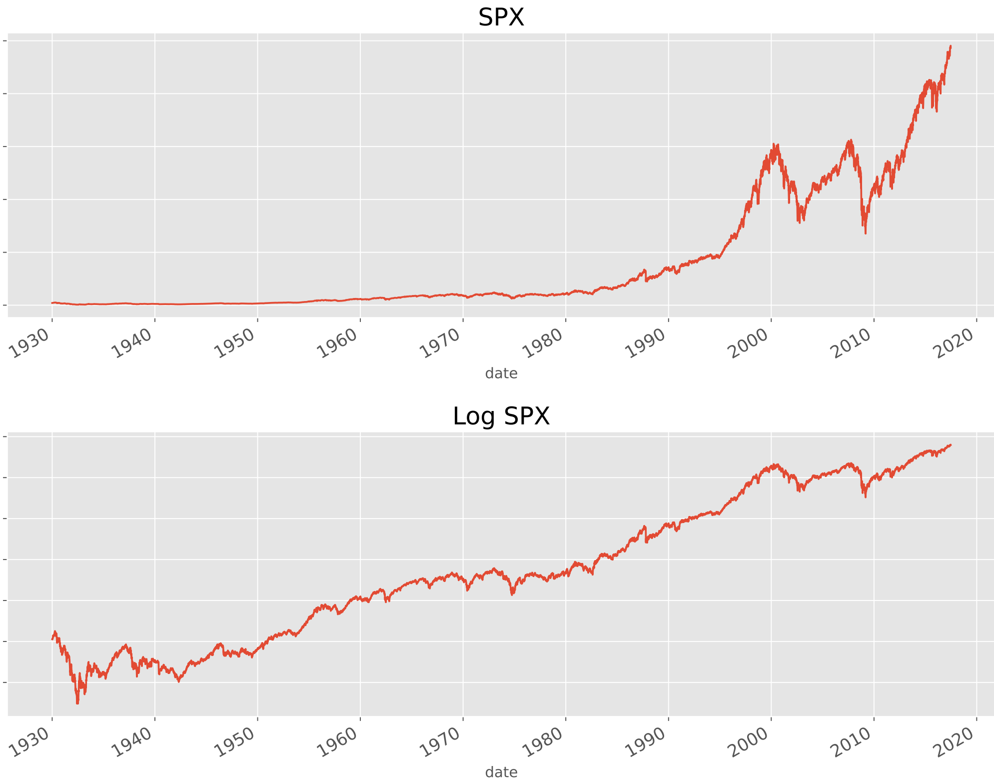


Figure 1: SPX index prices and logarithmic SPX prices over time (1930-2017).

1.1.2 Stationarity

A stochastic signal is referred to be wide-sense stationary (WSS), if the mean of the signal is time invariant and its autocorrelation function only depends on the time difference.

$$x(t) = x(t+T) \tag{1}$$

$$R_{xx}(t_1, t_2) = R_{xx}(t_2 - t_1) \tag{2}$$

Wide-sense stationarity for a signal means that the average of the signal stays the same over time and its autocorrelation only depends on the time difference. These concepts are the fundamentals behind the investment strategy based on the mean-reverting behaviour of a signal. Assuming there are 252 trading days in a year, we use a sliding window of 252 days to compute the first and second-order evolution statistics (mean and standard deviation) of the price and log-price time series. We use the sliding mean to represent the moving average of the price data which smoothen the signal. The sliding standard variation gives an idea about the volatility of the signal over the past trading year. From Figure 2, we observe that the moving average of price and log-price aren't stationary but instead upward trending. In particular, the log-price time series seem to be increasing at the constant rate over the past 70 years. We also observe the effect of the log transformation in stabilizing the variance. In the bottom right, we observe that the 252-day rolling standard deviation of the log-price data is more stationary, oscillating between 0 and 0.15 while the standard deviation of the price data is more volatile.

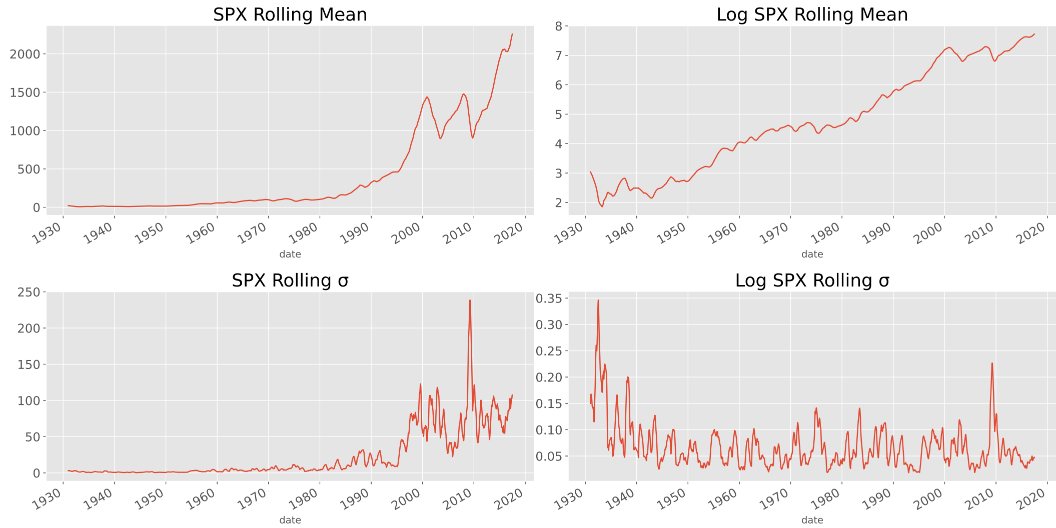


Figure 2: Rolling mean and standard deviation for SPX price data and log SPX price data.

1.1.3 Simple and log returns

We have seen in part 1.1.2 that the time series for price and log-price was not stationary. However, it is often preferable to deal with stationary signals in financial analysis. Therefore, we look into return characteristics of the signal, which are better in terms of stationarity. The simple return, is defined as the percentage increase of the price at each time step.

$$R_t = \frac{p_t - p_{t-1}}{r_{t-1}} \tag{3}$$

while the log return, is defined as the difference between the successive log values of the price.

$$r_t = ln\left(\frac{p_t}{p_{t-1}}\right) \tag{4}$$

We have in the figure below the plots of the statistics for the simple and log return time series. At first sight, the plots for simple and log returns individual statistics look very similar. Although there are slight discrepancies, the time series of the simple and log returns are almost identical. This indicated the approximate simple-log equality which suggest that simple and log returns are equal for small returns $R_t \approx r_t$.

Figure 3: Simple and log returns time data (top), rolling mean (middle) and rolling standard deviation (bottom).

Figure 3 shows the rolling window analysis applied to the returns signal. Observing these returns, the signal is much-more well-behaved, with the moving average for both returns oscillating around zero which suggests more stationarity.

1.1.4 Log returns over simple returns

As discussed in the previous sections, log transformations is suitable for time series due to its monotonicity and the fact that it compresses the range of the data which allows to see trends more clearly. In section 1.1.3, we also discussed that for small returns simple and log returns are approximately equal. In this section, we will continue the argument on why log-returns are more convenient to analyse than simple returns. In quantitative finance, we often expect the prices to be log-normally distributed over short periods of time, (for longer periods, the distribution is more skewed since the market is generally upward trending). Since prices are considered normally distributed over short periods of time,we can assume log returns to be normally distributed. Because many signal processing techniques assume normal distribution, we often prefer using log returns for analysis. See the histogram plots in Figure 4 to visualise the distribution of the returns. Another reason why we might prefer using logarithmic returns is that they provide time additivity, which is convenient when calculating compounded returns. That is, if we have 1% return one day and -1% return the next day, we end up with the same value; this is not the case with simple returns. Due to their Gaussian distribution, log returns are symmetric as seen in Figure 4. Furthermore, the sum of Gaussian variables is also Gaussian which makes it convenient for signal processing operations. We also note the fact that logarithms and exponents allow easier manipulation with calculus. Logarithms also offer numerical stability, meaning adding small numbers don't have significant effects (however, multiplying with small numbers can have large effects).

There are multiple tests available that make it is possible to test a signal's Gaussianity. One of

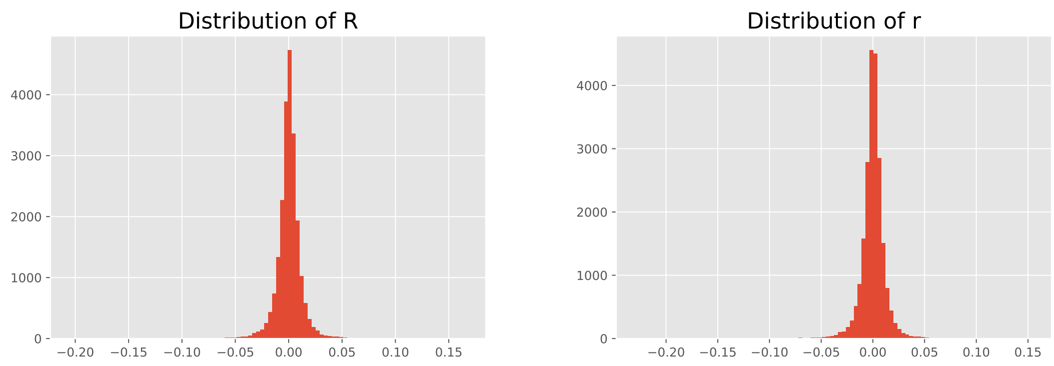


Figure 4: Distribution of SPX simple and log returns

them is the "Jarque-Bera" test, which checks whether the sample data has the skewness and kurtosis matching a normal distribution and is available under `scipy.stats` module in Python. A normal distribution has skewness equal to 0 and kurtosis equal to 3. Therefore, the closer the JB statistic is to 0, the closer the distribution of the sample is to a Gaussian.

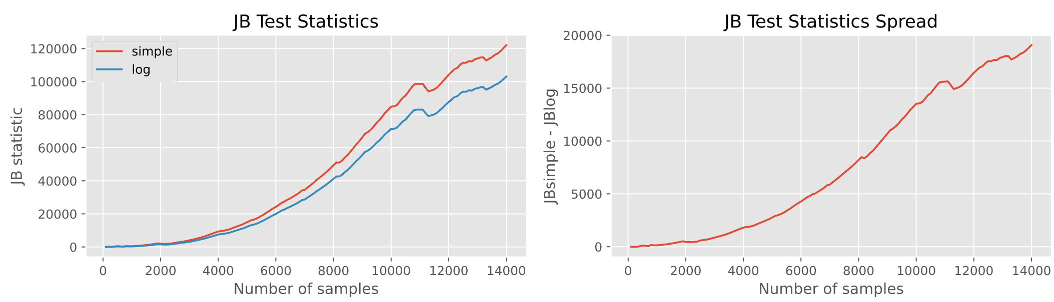


Figure 5: Jarque Bera test statistics for simple and log returns.

We plot the JB statistic of simple and log returns for different number of samples on the left side in Figure 5, observing that the values increase with increasing number of samples. The statistic for simple returns increases faster than that of log returns, and we plot this spread on the right side. This indicates the fact that log returns exhibit more Gaussianity than do the simple returns. In addition, both returns exhibit Gaussianity for the short-term (e.g. p-values for the JB Test of quarterly data are over 0.4), but that is not the case for the long-term data where the p value converges to 0.

1.1.5 Simple and log return example

If I purchase a stock at £1 and its value goes up to £2, the simple return would be 100% and the log return would be 69%. The following day the stock goes back to £1, my daily simple return is now -50% and my log return for the day is -69%. We observe that the sum of the daily logarithmic returns is zero when the value of the stock stays constant, which implies the time additivity property of log returns. We can conclude, that arithmetic sum of log returns is more suitable than simple returns to describe the value gained by an asset over time as it provides computational advantages in the analysis of financial data.

1.1.6 When not to use log returns

We mentioned in section 1.1.4 that we assume the prices to be log-normally distributed over short periods of time. However, this assumption does not hold over long time scales: log-normal distributions are positively skewed, but often financial data are negatively skewed in long time scales due to financial crashes. Therefore, we should not use logarithmic returns when doing long-term analysis. Furthermore, when calculating the overall return of a portfolio, it is more convenient to use simple returns because they are linearly additive across assets, whereas logarithmic returns are not.

1.2 ARMA vs. ARIMA Models for Financial Applications

1.2.1 Suitability of ARMA and ARIMA Models

In Figure 6, we illustrate the historical close prices for the S&P500 together with its rolling mean and standard deviation. We can easily see that there is an upward trend in the prices, which suggests that the signal is not stationary. As mentioned before, ARMA models assume stationarity in the signal hence it would not be appropriate to use them to model the S&P500 close prices. Instead, we explore the use of ARIMA models which account for nonstationarity of signal by applying a differencing operation.

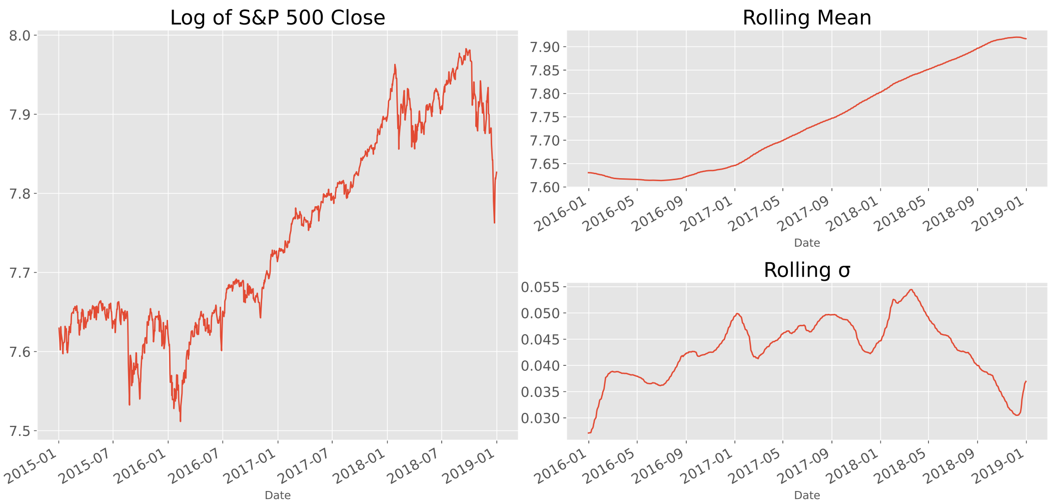


Figure 6: Historical S&P500 closing prices with rolling mean and standard deviation (for close prices)

1.2.2 Fitting an ARMA model

For this section, an ARIMA(1,0,0) model is used to fit the S&P 500 data, meaning that it is a first order auto-regressive model. This model is suitable for the financial time series because it is often assumed to be a martingale process, where the conditional expectation of next value given all prior values is equal to the present value only.

The AR(1) model fits the function quite well, with the figure below showing the true vs predicted values and the residual of last 100 days, with an mean absolute residual of 0.005974. Overall, the prediction is a lagged version of the real signal. It is also interesting to see that the residuals are smaller for well-behaved part of the signal. This is also expected since the model does not have a moving average part that models the shocks of exogenous events which can cause higher volatility.
Furthermore, because the given AR(1) model assumes the signal to be purely martingale without accounting for the shock effects in the market, it diverges from the reality. The prediction can be basically viewed as a one time step lagged version of the true signal, which is not very useful in practice. Finally, this model does not account for the non-stationarity in the signal, and we will look into this in the next section.

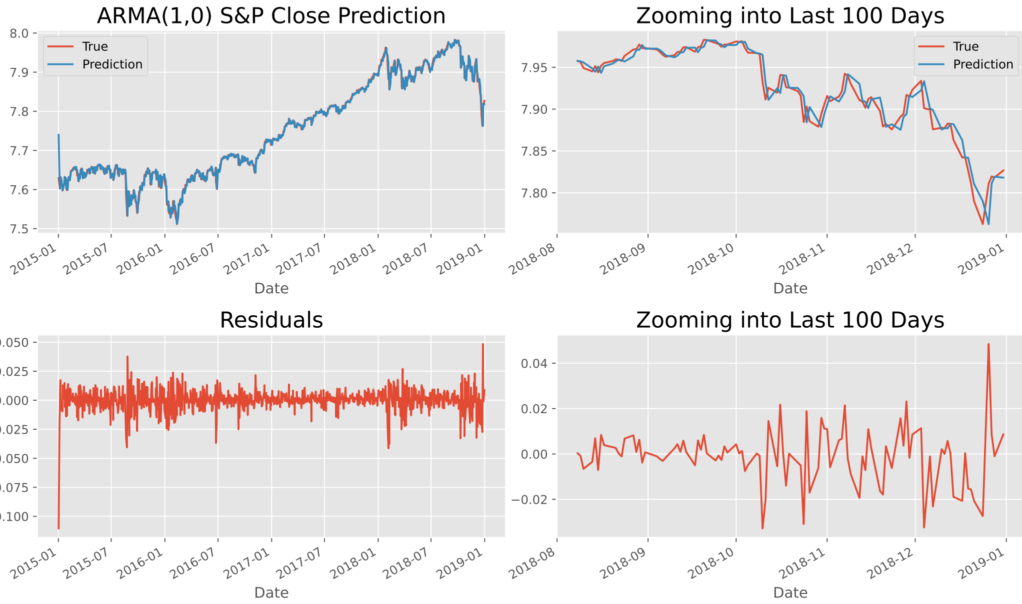


Figure 7: ARMA(1,0) S&P500 close price prediction and its residuals

1.2.3 Fitting an ARIMA model

In the previous section, we mentioned that ARMA models assume stationarity and are not suitable for non-stationary signals. In this section, we use an ARIMA(1,1,0) model which is really similar to the ARMA previously used with the exception that it applies an initial differencing on the time-series to remove elements of non-stationarity. As we had concluded that the S&P 500 close price data was showing an upward trend making the signal non-stationary, it is more appropriate to use an ARIMA model instead of an ARMA.

The plots are very similar to those of the AR(1) model in section 1.2.2, though we observe a slight improvement of 2% in the mean absolute error. The AR coefficient, -0.0088, is very close to zero, indicating there is little correlation between successive returns and that the model is not very suitable for practical purposes. Furthermore, similar to AR(1), this model also doesn't account for shock effects due to the lack of an MA part.

This ARIMA model, with an integrating order of 1, takes the difference between successive log-prices and applies an ARMA to these differences. Because these values represent log-returns, they are normally distributed and are stationary, as we discussed in previous sections. Because applying an ARMA model to a stationary signal is more appropriate, this analysis is more meaningful.

1.2.4 Necessity of taking the log for ARIMA

An ARIMA model applies a one time time series difference to an ARMA model and it uses log-return data to perform log-price prediction. Because of the time additivity property of log returns, the ARIMA model is fitting an auto regressive model on successive log returns. In addition, the log-return data is a stationary signal which makes it suitable to use on an ARMA model and improves the quality of modelling.

1.3 Vector Autoregressive (VAR) Models

The vector auto-regressive (VAR) model is an extension of the AR model where the predicted values are dependent on the past values of multiple variables. It is used to model variables that

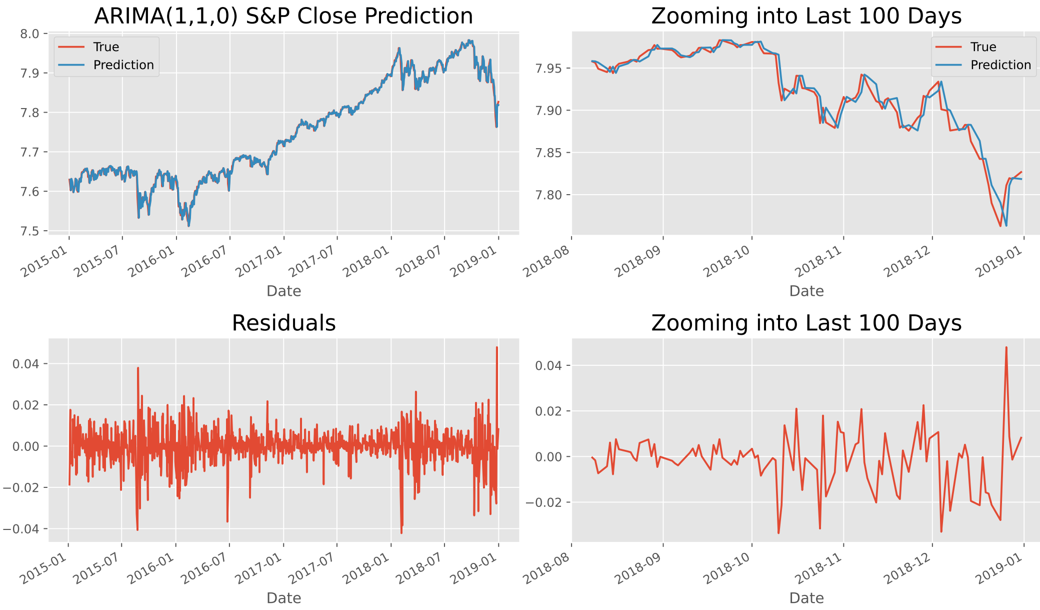


Figure 8: ARIMA(1,1,0) S&P500 close price prediction and its residuals

exhibit correlation with their own past values as well as past values from other variables. A VAR(p) process is given by:

$$\mathbf{y}_t = \mathbf{c} + \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + ... + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{e}_t \tag{5}$$

or, in expanded matrix notation:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{k,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} + \begin{bmatrix} a^1_{1,1} & a^1_{1,2} & \cdots & a^1_{1,k} \\ a^1_{2,1} & a^1_{2,2} & \cdots & a^1_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ a^1_{k,1} & a^1_{k,2} & \cdots & a^1_{k,k} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \vdots \\ y_{k,t-1} \end{bmatrix} + \cdots + \begin{bmatrix} a^p_{1,1} & \cdots & a^p_{1,k} \\ a^p_{2,1} & \cdots & a^p_{2,k} \\ \vdots & \ddots & \vdots \\ a^p_{k,1} & \cdots & a^p_{k,k} \end{bmatrix} \begin{bmatrix} y_{1,t-p} \\ y_{2,t-p} \\ \vdots \\ y_{k,t-p} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \\ \vdots \\ e_{k,t} \end{bmatrix} \tag{6}$$

1.3.1 Concise VAR matrix form

Now, taking the bases view of the matrix, we can rewrite the VAR model concisely with the following equation

$$\mathbf{Y} = \mathbf{BZ} + \mathbf{U} \tag{7}$$

where:

$$\mathbf{B} = \begin{bmatrix} \mathbf{c} & \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_p \end{bmatrix} \tag{8}$$

$$\mathbf{Z} = \begin{bmatrix} 1 \\ \mathbf{y}^T_{t-1} \\ \mathbf{y}^T_{t-2} \\ \vdots \\ \mathbf{y}^T_{t-p} \end{bmatrix} \tag{9}$$

$$\mathbf{U} = \mathbf{e}_t \tag{10}$$

$$\mathbf{Y} = \mathbf{y}_t \tag{11}$$

However, we can generalise even further by extending the equation (7) of the VAR model to the mutli-period case, where **T** time-steps are modelled. In this case, we have:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_t & \mathbf{y}_{t+1} & \mathbf{y}_{t+2} & \cdots & \mathbf{y}_{t+T} \end{bmatrix} \tag{12}$$

$$\mathbf{Z} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \mathbf{y}_{t-1} & \mathbf{y}_t & \cdots & \mathbf{y}_{t-1+T} \\ \mathbf{y}_{t-2} & \mathbf{y}_{t-1} & \cdots & \mathbf{y}_{t-2+T} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{t-p} & \mathbf{y}_{t-p+1} & \cdots & \mathbf{y}_{t-p+T} \end{bmatrix} \tag{13}$$

$$\mathbf{U} = \begin{bmatrix} \mathbf{e}_t & \mathbf{e}_{t+1} & \mathbf{e}_{t+2} & \cdots & \mathbf{e}_{t+T} \end{bmatrix} \tag{14}$$

1.3.2 Optimal VAR coefficients

In order to find the optimal set of coefficients $\mathbf{B_{opt}}$, we aim to minimise $\mathbf{U} = \mathbf{Y} - \mathbf{BZ}$, which represents the error of our VAR model. We use the least-squares method to find the optimal set of coefficients with the following objective function $\mathbf{J}(\mathbf{B})$:

$$\mathbf{J}(\mathbf{B}) = \mathbf{UU^T} = (\mathbf{Y} - \mathbf{BZ})(\mathbf{Y} - \mathbf{BZ})^\mathbf{T} \tag{15}$$

$$\mathbf{J}(\mathbf{B}) = \mathbf{YY^T} - \mathbf{BZY^T} - \mathbf{YZ^TB^T} + \mathbf{BZZ^TB^T} \tag{16}$$

Then, setting the derivate of the objective function to 0, we solve for $\mathbf{B_{opt}}$ that minimises the following equation:

$$\frac{\partial \mathbf{J}(\mathbf{B})}{\partial \mathbf{B}} = \mathbf{0} \Rightarrow \mathbf{2BZZ^T} - \mathbf{2YZ^T} = \mathbf{0} \tag{17}$$

$$\mathbf{BZZ^T} = \mathbf{YZ^T} \tag{18}$$

$$\mathbf{B_{opt}} = \mathbf{YZ^T}(\mathbf{ZZ^T})^{-1} \tag{19}$$

1.3.3 Stability of VAR

Let's consider a VAR(1) process, ie:

$$\mathbf{y_t} = \mathbf{Ay_{t-1}} + \mathbf{e_t} \tag{20}$$

$$\mathbf{y_{t-1}} = \mathbf{Ay_{t-2}} + \mathbf{e_{t-1}} \tag{21}$$

Recursively, we can rewrite the equation $\mathbf{y_t}$ from the equations above as:

$$\begin{aligned} \mathbf{y_t} &= \mathbf{Ay_{t-1}} + \mathbf{e_t} && (22) \\ &= \mathbf{A}(\mathbf{Ay_{t-2}} + \mathbf{e_{t-1}}) + \mathbf{e_t} && (23) \\ &\vdots && (24) \\ \mathbf{y_t} &= \mathbf{A^p y_{t-p}} + \sum_{i=0}^{p-1} (\mathbf{A^i e_{t-i}}) && (25) \end{aligned}$$

Assuming matrix **A** is diagonalisable, that is assuming **A** has **k** independent eigenvectors with corresponding eigenvalues, we can write:

$$
\begin{aligned}
\mathbf{A} &= \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} && (26)\\
\mathbf{A^2} &= \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} = \mathbf{V}\mathbf{\Lambda^2}\mathbf{V^{-1}} && (27)\\
\mathbf{A^n} &= \mathbf{V}\mathbf{\Lambda^n}\mathbf{V}^{-1} && (28)\\
& && (29)
\end{aligned}
$$

where $\mathbf{V}$ is a matrix composed of $\mathbf{k}$ independent eigenvectors and $\mathbf{\Lambda}$ is a diagonal matrix with the corresponding eigenvalues on the diagonals.
If we replace (28) into (25), we see that if one of the eigenvalue in the diagonal of $\mathbf{\Lambda}$ is greater than 1, the value of $\mathbf{y_t}$ diverges to infinity. For this reason, we need all the eigenvalues in $\mathbf{\Lambda}$, and by extension in $\mathbf{A}$, to be less than 1 in absolute value for the stability of the model.

1.3.4 VAR Portfolio Analysis I

In this sub-section, we will analyse the construction of a portfolio using VAR models. We will first observe the close price time-series of five stocks, with ticker symbols 'CAG' 'MAR' 'LIN' 'HCP' 'MAT', and detrend the corresponding signals using the 66 days moving average (66 corresponding to 3x22 i.e. one quarter) to obtain zero-mean stationary time-series. We start by plotting the close price time series and detrended time series in Figure 9.

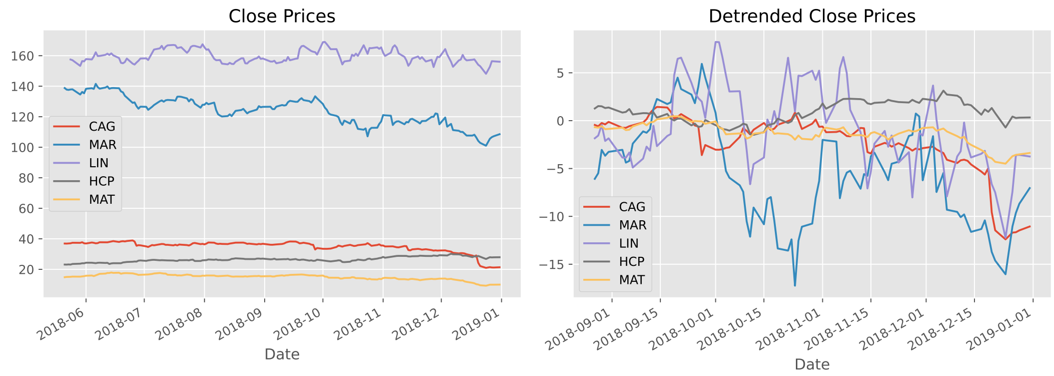


Figure 9: Close prices (left) with their detrended time-series (right) for five selected stocks

We then fit a VAR(1) model to the detrended time series and compute the eigenvalues of the regression matrix $\mathbf{A}$ as:

	CAG	MAR	LIN	HCP	MAT
L1.CAG	0.872786	0.113179	-0.281265	0.011912	0.058776
L1.MAR	-0.063745	0.895820	-0.184820	-0.005004	0.022917
L1.LIN	0.000134	-0.111678	0.704023	0.004982	-0.025557
L1.HCP	-0.084776	-0.083831	-0.401417	0.931708	-0.046406
L1.MAT	0.643072	0.094931	2.033036	-0.012884	0.802974

Table 1: Regression matrix $\mathbf{A}$ contained in a table

the eigenvalues of that matrix are: $\lambda_0 = 0.714 + 0.129j, \lambda_1 = 0.714 - 0.129j, \lambda_2 = 1.006, \lambda_3 = 0.861, \lambda_4 = 0.911$.
We can observe that the diagonal of the above matrix contains the highest values, which is expected as this means that the time-series values are autocorrelated with their own past values. However, as

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

Table 2: Sectors of activity for each of the five companies

these companies operate in different sectors (see Table 2), we see that the correlations of the assets between each other is very low and sometimes negative. This suggests that it is a good thing to build a portfolio containing these stocks, as having uncorrelated assets creates a diversified portfolio which lowers the risk.

1.3.5 VAR Portfolio Analysis II

We reproduce the process applied in section 1.3.4, this time applying a VAR(1) model to entire sectors and we obtain the eigenvalues of the regression matrix $\mathbf{A}$ for each sector.
The overall expected return of a portfolio composed of n assets is: $\overline{r_p} = \sum_{i=1}^{n} w_i \overline{r_i}$ and its variance is $\sigma^2 = \sum_{i,j=1}^{n} w_i \sigma_{ij} w_j$ where $\overline{r_i}$ is the expected return for a particular asset i, w_i is the weight assigned to that asset and σ_{ij} is the covariance between two assets i and j. For a given target return, investors aim to minimise the risk associated to it, represented by the variance of the portfolio.
From the results in Table 3, we see that stocks from the same sectors are highly correlated between each other, which means the term σ_{ij} is going to be high. Hence, in order to minimise risk it is advisable to construct a portfolio using assets from different sectors that are uncorrelated and will have a lower covariance.

	min eigenvalue	mean eigenvalue	max eigenvalue
Industrials	0.371246	0.763932	0.991721
Health Care	0.092157	0.621440	0.994153
Information Technology	0.374081	0.809351	0.992738
Communication Services	0.752488	0.926293	0.982263
Consumer Discretionary	0.447563	0.810433	0.990650
Utilities	0.042115	0.599877	0.985648
Financials	0.152575	0.631218	1.004340
Materials	0.137838	0.621833	0.991744
Real Estate	0.763563	0.919369	0.982785
Consumer Staples	0.546458	0.852121	0.991508
Energy	0.825707	0.930601	0.985577

Table 3: Eigenvalue statistics for each of the studied sector.

2 Bond Pricing

2.1 Examples of bond pricing

2.1.1 Nominal rates

Let's consider an investment of USD 1,000 today that will return USD 1,100 in one year. This means that this investment returns an efective annual interest rate r_{eff} of 10%. We will now calculate the equivalent nominal interest rate for different compounding periods. We first introduce the following formula where r is the nominal interest rate and m is the number of compounding periods in a year, the the yearly effective interest rate becomes:

$$1 + r_{eff} = \left(1 + \frac{r}{m}\right)^m \tag{30}$$

we use equation (30) to get the nominal interest rate as:

$$r = m((1 + r_{eff})^{\frac{1}{m}} - 1) \tag{31}$$

We use equation (31) to calculate the nominal interest rates for different compounding periods given the effective interest rate $r_{eff} = 10\%$, then:
a) Annual compounding: m=1, r=10%
b) Semi-annual compounding: m=2, r=9.76%
c) Monthly compounding: m=12, r=9.57%
d) Continuous compounding: In here, we take equation (30) to reformulate the problem for t years. We obtain the following equation:

$$1 + r_{eff} = \left(1 + \frac{r}{m}\right)^{mt} \tag{32}$$

As we are dealing with continuous compounding, we study the result for m going to infinity:

$$\lim_{m\to\infty} \left(1 + \frac{r}{m}\right)^{mt} = e^{rt} \tag{33}$$

Hence the continuous compounding rate after 1 year ($t = 1$) is $r = ln(1 + r_{eff})$ which in our case is $r = 9.53\%$.

2.1.2 Equivalent rates

In this section, we will see what interest with continuous compounding is equivalent to 15% annual rate with monthly compounding. Using equation (30) and (33) we can derive the following equation:

$$e^r = \left(1 + \frac{0.15}{12}\right)^{12} \tag{34}$$

Solving this equation gives us $r = 14.91\%$.

2.1.3 Interest calculation

This section will evaluate the interest paid on a 10,000 USD deposit that pays 12% per year with continuous compounding but only receives interest each quarter.
As $r = 12\%$, we know that the effective rate is $r_{eff} = e^r - 1 = 12.75\%$. Since the interest are paid quarterly, we can use the same equation to get the quarterly effective rate as $r_{eff} = e^{0.12/4} - 1 = 3.045\%$. We can then have the detail of the interests paid quarterly on an initial investment of 10,000 dollars as follow:
First quarter: \$304.5
Second quarter: \$313.77
Third quarter: \$323.33
Forth quarter: \$333.17
Note: these numbers are approximations to the second digit.

2.2 Forward Rates

2.2.1 Forward rates example

In this section, we discuss an investment with a one-year interest rate r_1 of 5% and a two-year interest rate r_2 of 7%. Investing in the two year investment brings an additional $1.07^2/1.05 - 1 = 9\%$ additional return.
a) The one year investment allows more flexibility to the investor to get their money back with interest and use it for other investments. Locking the money up for a longer period of time rewards the investor with a higher interest rate. For no arbitrage to exist, these two strategies should be equivalent as the two year investment should already include the expected 1 year spot rate for next year into these assets. Therefore, in theory, the investor should be neutral on whether to invest one year with rate r_1 or two years with rate r_2.
b) As discussed above, the strategy should not be based on the rates offered for different investment periods as the expected forward rates should already be included in these rates, but the strategy should focus on opportunity cost of investing and need of liquidity. The implied forward rate in r_2 of 9% should be close to the forward market rate, although they can slightly differ due to market's imperfections.
c) The forward rate $f_{1,2}$ is higher than the current rates r_1 and r_2 as investors take the risk of future interest rate changes which can impact their investment either way. If interest rates go up, then the investor has potentially missed out on more income while if they go down, then he should have bought the assets with the implied forward interest rate of 9%. These are often used as hedges against interest rate change but can go either way for the investor.
d) Forward rates are calculated using the forecasted spot rates. Given the spot rates s_i and s_j for the i^{th} and j^{th} years, the implied forward rate $f_{i,j}$ can be calculated as follow:

$$f_{i,j} = \left[\frac{(1+s_j)^j}{(1+s_i)^i}\right]^{1/(j-i)} - 1 \tag{35}$$

Note that market forward rates are usually different from the implied forward rates due to inefficient market imperfections.

2.3 Duration of a Coupon-Bearing Bond

2.3.1 Macaulay duration

The duration, or Macaulay duration is the weighted average of the times to each of the cash payments. The weight for each year is the present value of the cash flow received at the time divided by the total value of the bond:

$$Duration = \frac{1 \times PV(C_1) + 2 \times PV(C_2) + \cdots T \times PV(C_T)}{PV} \tag{36}$$

We compute the duration for the 1% seven-year treasury bond assuming annual payments in Table 4. We first compute the present value of the payment cash flow stream as well as the final face value. We obtain a total present value of \$ 768.55 . We then, compute the proportion of each present value to the total present value and multiply that by the year to obtain the duration of each cash flow. Adding these up, we obtain a duration of 6.76.

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Total
Payment	10	10	10	10	10	10	1010	1070
PV(Ct)	9.52	9.07	8.64	8.23	7.84	7.46	717.79	768.55
PV(Ct)/PV	0.012	0.012	0.011	0.011	0.01	0.01	0.934	1
t * PV(Ct)/PV	0.012	0.024	0.033	0.044	0.05	0.06	6.538	**6.76**

Table 4: Calculating duration of a 1% 7-year bond with a 5% yearly yield to maturity

2.3.2 Modified duration

Investors and financial managers track duration because it measures how bond prices change when interest rates change. For this purpose it is best to use modified duration or volatility which is the percentage change in the bond's price for a change in the yield λ:

$$D_M = -\frac{1}{P(\lambda_0)}\frac{dP(\lambda)}{d\lambda} = \frac{D}{1+\lambda/m} \tag{37}$$

The bond in the table has a yield λ = 5% and we take annual compounding hence m = 1. The modified duration is hence:

$$D_M = \frac{6.76}{1+0.05} = 6.44 \tag{38}$$

The modified duration is lower than the Macaulay duration and would have been equal for continuous compounding (that is $m \to \infty$ and hence $D_M = D$).

2.3.3 Advantages of duration against volatility

A bond with a longer time to maturity has a price very sensitive to he yield and is more impacted by unexpected changes in interest rates. Duration gives investors a quantitative measure of that interest rate sensitivity and the risk level of a bond. For example, in the previous section, if the yield increases by 1%, the price of the bond would fall 6.44%. Using duration, we can however hedge that risk using "Immunisation", which is a strategy matching the price and duration of the assets and liabilities in a portfolio to protect against interest rate changes. This can useful for pension funds for example to protect to minimise the risks.

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

In this section, we will study the daily returns of 157 european companies and study the Capital Asset pricing model (CAPM) as well as the Arbitrage pricing theory (APT). CAPM is is a model that estimates the retrun of an individual asset in the presence of a risk-free asset assuming certain market conditions. Using CAPM, the retrun for an asset i, R_i is:

$$r_i = r_f + \beta_i(r_M - r_f) + \epsilon_i \tag{39}$$

where r_f is the risk free return, r_M is the market return, and ϵ_i is the company specific risk uncorrelated to the market. The asset sensitivity to the market β_i is given by:

$$\beta_i = \frac{cov(r_i, r_M)}{var(r_M)} \tag{40}$$

We can then define the expected return of an asset i as:

$$\bar{r_i} = r_f + \beta_i(\bar{r_M} - r_f) \tag{41}$$

2.4.1 Non-weighted market return

In this section, we estimate the daily market return r_M averaging the daily returns of each of the 157 companies (note that the results only show the results for 141 companies as some of the data points are missing for 16 companies). We plot the non-weighted results in Figure 10 and observe an average return of 0.005% with an average standard deviation of 0.0067. We obtain a Sharpe ratio of 141 for the non-weighted market portfolio.

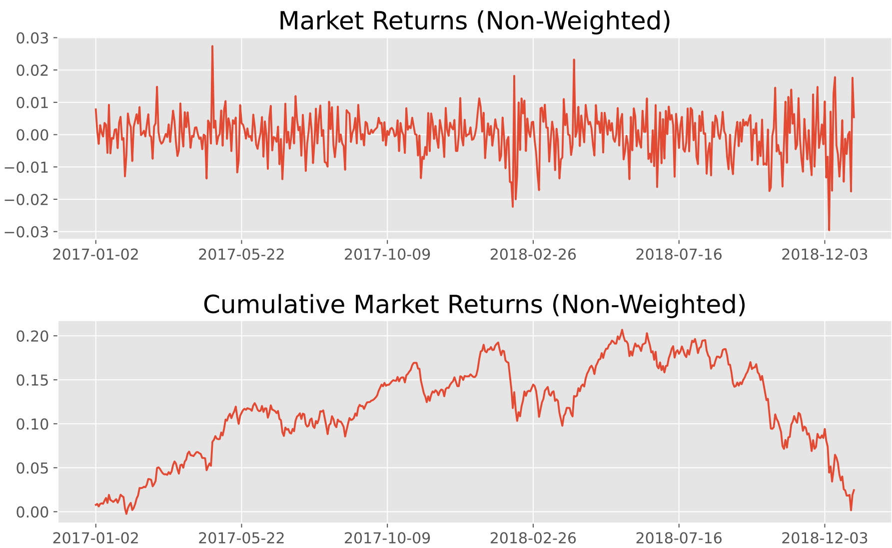


Figure 10: Non-weighted market returns and cumulative returns

2.4.2 Non-weighted rolling beta

In this section, we calculate the rolling β for each company for a window of 22 days (one trading month). We use equation (40) hence we need to calculate the covariance of each asset with the market and the variance of the market both for a 22 day rolling window. We plot the rolling βs in the first plot and then present the mean and standard variation of these in Figure 11. The average β value for a company is just under 1 while the standard variation is 0.55.

An average β of 1 signifies that, on average, a company moves together with the market and this tends not to be very volatile. However, we also see that individually companies can have very volatile βs which means that certain companies tend to be less consistent with the market than others.

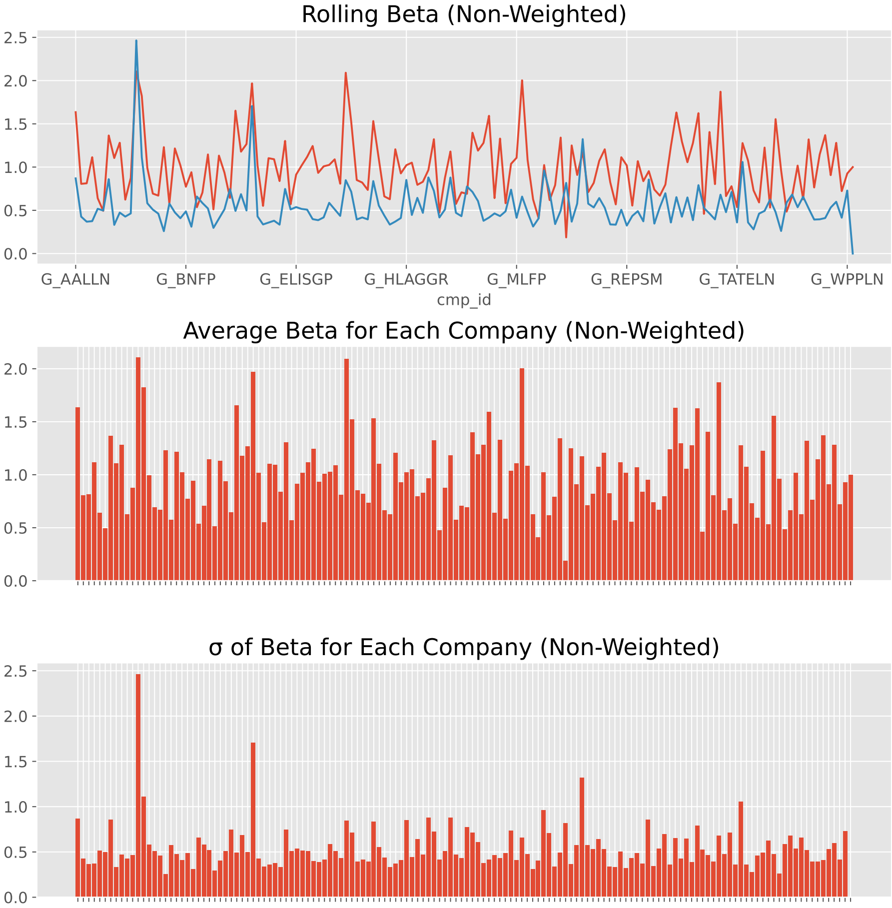


Figure 11: Non-weighted rolling beta (top), average beta for each company (middle) and standard deviation of the beta for each company (bottom)

2.4.3 Mcap-weighted market return

In this section, we will evaluate market returns by weighting the companies based on their market capitalisation. The cap-weighted market return R_m is given by:

$$R_m = \sum_i = \frac{mcap_i \times ret_i}{\sum_i mcap_i} \tag{42}$$

The marketcap weighted market return gives us an average return of 0.019% with an average standard deviation of 0.0066. Here the Sharpe ratio is 35, much smaller than the one of the non-weighted market return as the small-cap stocks represent a much greater risk and are more volatile but given their small proportion in the marketcap weighted market return, that risk is vastly dissipated.

2.4.4 Mcap-weighted rolling beta

In this section, we take a similar approach as in section 2.4.2 estimating rolling βs but this time using mcap weighted market return. The plots in Figure 13 are similar to the ones in section 2.4.2. The

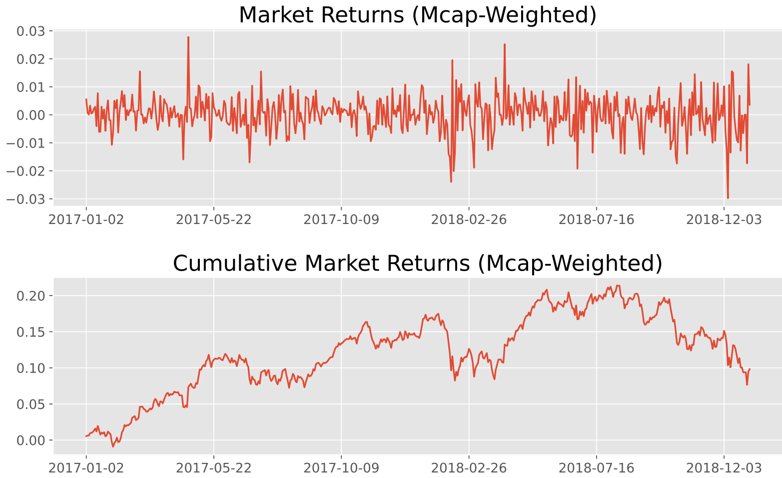


Figure 12: Market capitalisation weighted market returns and cumulative returns

average β is 0.96 (lower than the non-weighted method) and the standard deviation is 0.56 (slightly higher than the non-weighted method).

2.4.5 Arbitrage pricing theory (APT)

a) In this section, we assume that the Arbitrage pricing theory holds for a two-factor model and hence we can write:

$$r_i = a + b_{mi} R_m + b_{si} R_s + \epsilon_i \tag{43}$$

where b_{mi} is the exposure to the market return R_m and b_{si} is the exposure to R_s the return relevant to size. We assume that $b_{si} = ln(size)$ and hence we can rewrite equation (43) as:

$$\mathbf{r} = \mathbf{B}\mathbf{x} + \mathbf{e} \tag{44}$$

where

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 1 & b_{m1} & bs1 \\ 1 & b_{m2} & bs2 \\ \vdots & \vdots & \vdots \\ 1 & b_{mn} & bsn \end{bmatrix}, \mathbf{x} = \begin{bmatrix} a \\ R_m \\ R_s \end{bmatrix}, \mathbf{e} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{45}$$

Since the values of $\mathbf{r}$ and $\mathbf{B}$ are known, we can use Ordinary Least Squares method to determine a, R_m and R_s as:

$$\mathbf{x} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{r} \tag{46}$$

b) We plot the daily values for the parameters a, R_m and R_s in Figure 14 as well as their mean and standard variation in Table 5.
We see that the magnitude of a is the biggest on average, which we expect as it reflects the upward trend in the market. We also notice that the market factor is bigger than the size factor, however the average exposure to the market factor is about 1 while the exposure to the size is around 20. We could also note that R_m has the worst Sharpe ratio among the parameters.
c) In this section, we study the correlation between the returns r_i and the specific returns ϵ_i which are plotted on the left of Figure 15. We see a positive linear trend which indicates a strong correlation

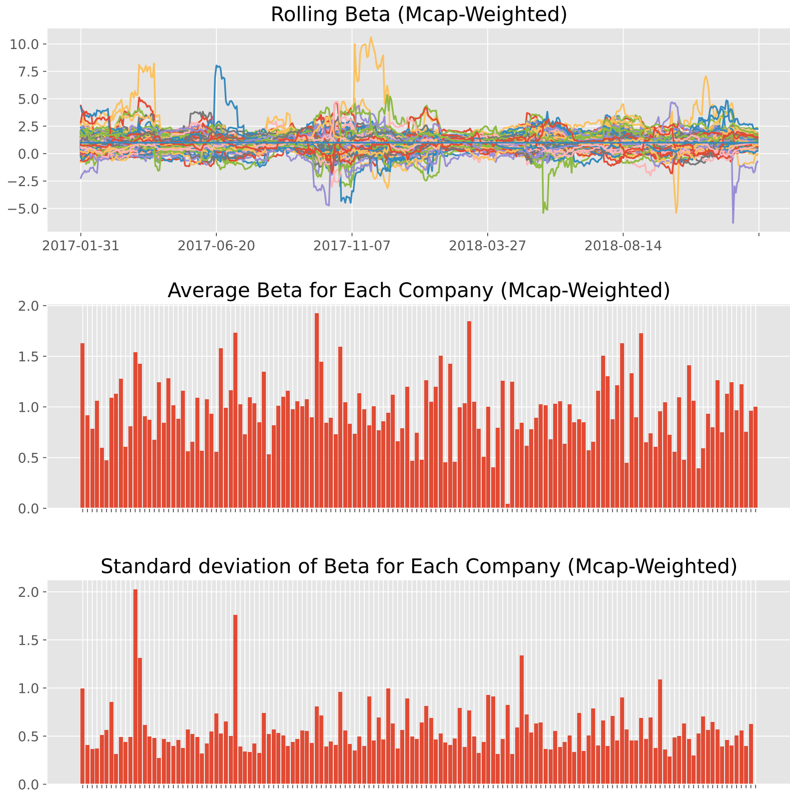


Figure 13: Market capitalisation weighted rolling beta (top), average beta for each company (middle) and standard deviation of beta for each company (bottom)

	a	Rm	Rs
mean	-0.004216	-0.000286	0.000192
std	0.041191	0.007913	0.001732
$\|$**mean/std**$\|$	9.771005	27.670946	9.004126

Table 5: Estimated factors statistics

between these values. The histogram on the right confirms that theory as all the values are above 0.60 and the average correlation is 0.81, which is very high. Since ϵ_i is the difference between the real returns and the returns explainable by the factors, the high correlation between the two factors indicates that they are not sufficient to to explain the given returns (otherwise correlation would be 0 but as it is high, this means that they explain the same part of the returns).

d) In this section, we explore the magnitude and stability of the covariance matrix $cov(R)$ using a rolling window of 22 days (one trading month) and is defined as follows:

$$R = \begin{bmatrix} R_{m1} & Rs1 \\ \vdots & \vdots \\ R_{m500} & Rs500 \end{bmatrix} \tag{47}$$

We plot the covariance matrix magnitude below and we see that the signal is highly fluctuating suggesting instability in Figure 16. From the magnitude percentage change plot we also observe that the values can drastically change from one day to another, re-enforcing the sentiment of instability. We concludes that the past estimations of covariance values tend to be obsolete to determine the

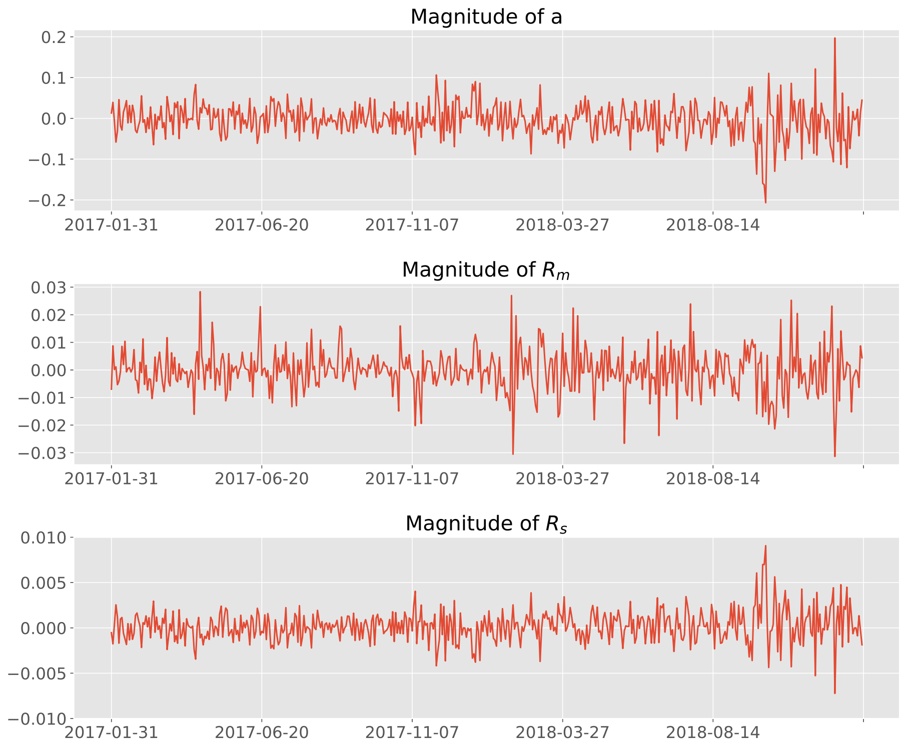


Figure 14: Estimated magnitude of a (top), R_m (middle) and R_s (bottom)

future ones.

e) From Part 2.4.5 a) we have for every company i, and for every day t, the specific return ϵ_i with the following matrix:

$$\mathbf{E} = \begin{bmatrix} \epsilon_{1,t=0} & \cdots & \epsilon_{157,t=0} \\ \epsilon_{1,t=1} & \cdots & \epsilon_{157,t=1} \\ \vdots & \ddots & \vdots \\ \epsilon_{1,t=5000} & \cdots & \epsilon_{157,t=500} \end{bmatrix}_{500\times 157} \tag{48}$$

We first compute the covariance matrix and perform PCA on it.
PCA is a dimensionality reduction method which takes the idea that the largest eigenvalues and their corresponding eigenvectors carries the principal information about the behaviour of the matrix. In our case, that would mean that they represent the main variation direction.
We plot the eigenvalues in descending order in black in the figure below as well as the percentage of explained variance by the eigenvalues in red in Figure 17. We can observe that the larger eigenvalues capture most the variance and simply by taking the first principal component (largest eigenvalue and its eigenvector) we can explain over 7% of the total variance.

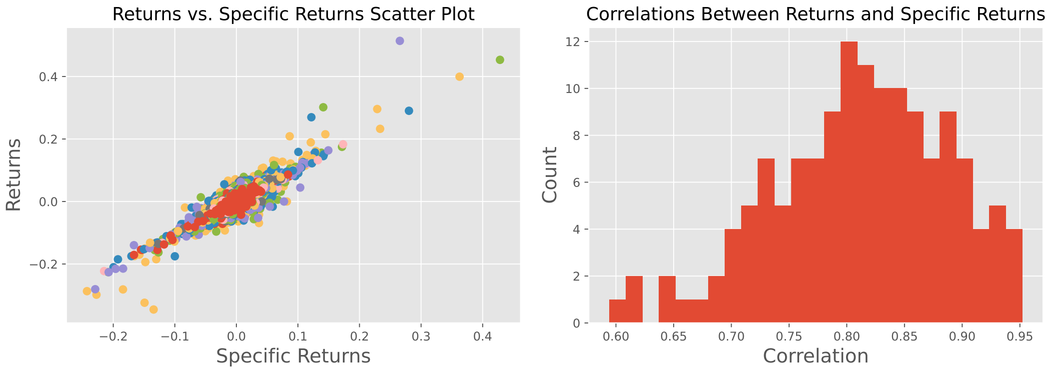


Figure 15: Returns and specific returns scatter plot (left) and correlation bar chart (right)

3 Portfolio Optimisation

3.1 Adaptive Minimum-Variance Portfolio Optimisation

3.1.1 Optimal weights of the minimum-variance portfolio

The minimum variance portfolio is found by deriving the optimal set of weights $\mathbf{w}$, which minimises the portfolio variance C_p^2 subject to the budget constraint, i.e. the portfolio weights must sum up to 1 $\mathbf{w}^T\mathbf{1} = 1$. We can express our optimisation problem in the following way:

$$\text{minimise} \qquad \frac{1}{2}\mathbf{W}^T\mathbf{C}\mathbf{w}, \tag{49}$$

$$\text{subject to} \qquad \mathbf{w}^T\mathbf{1} = 1 \tag{50}$$

equivalently, we can define the Lagrangian optimisation problem:

$$\min_{\mathbf{w},\lambda} J(\mathbf{w}, \lambda, \mathbf{C}) = \frac{1}{2}\mathbf{w}^T\mathbf{C}\mathbf{w} - \lambda(\mathbf{w}^T\mathbf{1} - 1) \tag{51}$$

From this Lagrangian we obtain the following two optimality conditions:

$$\frac{\partial J}{\partial \mathbf{w}} = \mathbf{0} \Rightarrow \mathbf{C}\mathbf{w} - \lambda\mathbf{1} = \mathbf{0} \tag{52}$$

$$\frac{\partial J}{\partial \lambda} = 0 \Rightarrow \mathbf{w}^T\mathbf{1} = 1 \tag{53}$$

We can rewrite the optimality conditions can be rewritten in matrix form:

$$\begin{bmatrix} \mathbf{C} & -\mathbf{1} \\ -\mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ -1 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{C} & -\mathbf{1} \\ -\mathbf{1}^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ -1 \end{bmatrix} \tag{54}$$

Now, we can get the value of $\mathbf{w}$ from equation (52) such as:

$$\mathbf{w} = \lambda\mathbf{C}^{-1}\mathbf{1} \tag{55}$$

We use this value of $\mathbf{w}$ into equation (55) to get:

$$\lambda\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1} = 1 \Rightarrow \lambda = \frac{1}{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1}} \tag{56}$$

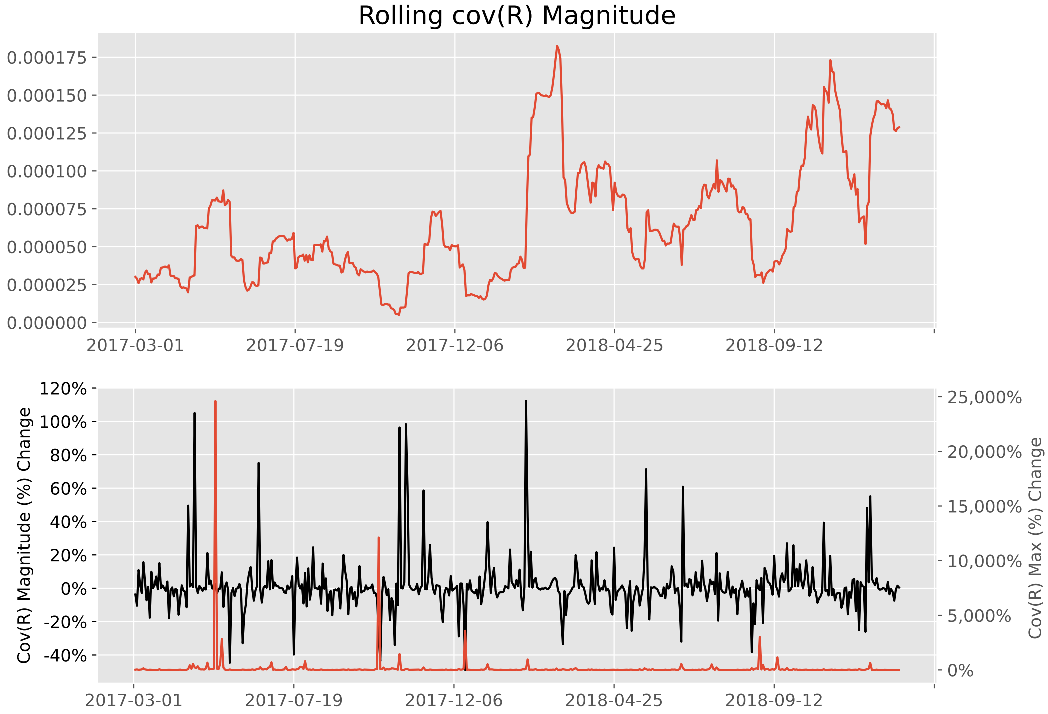


Figure 16: Magnitude changes of the rolling covariance matrix $cov(R)$

We can replace the value of λ of equation (56) into the formula for $\mathbf{w}$ in equation (55) to compute the optimal set of weights $\mathbf{w}_{opt}$:

$$\mathbf{w}_{opt} = \frac{\mathbf{C}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1}} \tag{57}$$

If we were to apply the minimum variance estimator, then the theoretical variance of the portfolio returns would be:

$$\begin{aligned}
\sigma_p^2 &= \mathbf{w}_{opt}^T\mathbf{C}\mathbf{w}_{opt} && (58)\\
&= \frac{\mathbf{1}^T(\mathbf{C}^{-1})^T\mathbf{C}\mathbf{C}^{-1}\mathbf{1}}{(\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1})^2} && (59)\\
&= \frac{\mathbf{1}^T(\mathbf{C}^{-1})^T\mathbf{1}}{(\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1})^2} && (60)\\
&= \frac{1}{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1}} && (61)
\end{aligned}$$

with $\mathbf{C}$ being a positive semi-definite matrix and $\sigma_{opt}^2 = \lambda$.
We can note two things about this optimisation problem. First, we aimed to create the portfolio with the minimum variance (risk) and haven't put a constraint on the portfolio target return. Secondly, there is no constraints regarding the signs of the weights of the assets, effectively allowing short selling which might always be available.

3.1.2 Minimum-variance portfolio vs. equally weighted portfolio

In this section, we investigate the daily returns from January 2017 to June 2018 of 10 stocks in order to construct the minimum variance portfolio. We split the data equally in two parts, half for training

Figure 17: Using eigenvalues to explain variance of the covariance matrix of the specific returns $cov(E)$

and the other half for testing. Then, we compute the covariance matrix for the training returns and evaluate the optimal set of weights as we described in the previous section.
We apply those weights to both the training and testing data to evaluate the performance of the minimum variance portfolio over a year and plot the daily and cumulative returns in the figures below and get some performance statistics for each model. We reproduce the same thing for the equally weighted portfolio where the weight for each of the 10 stocks is equal to 0.1.

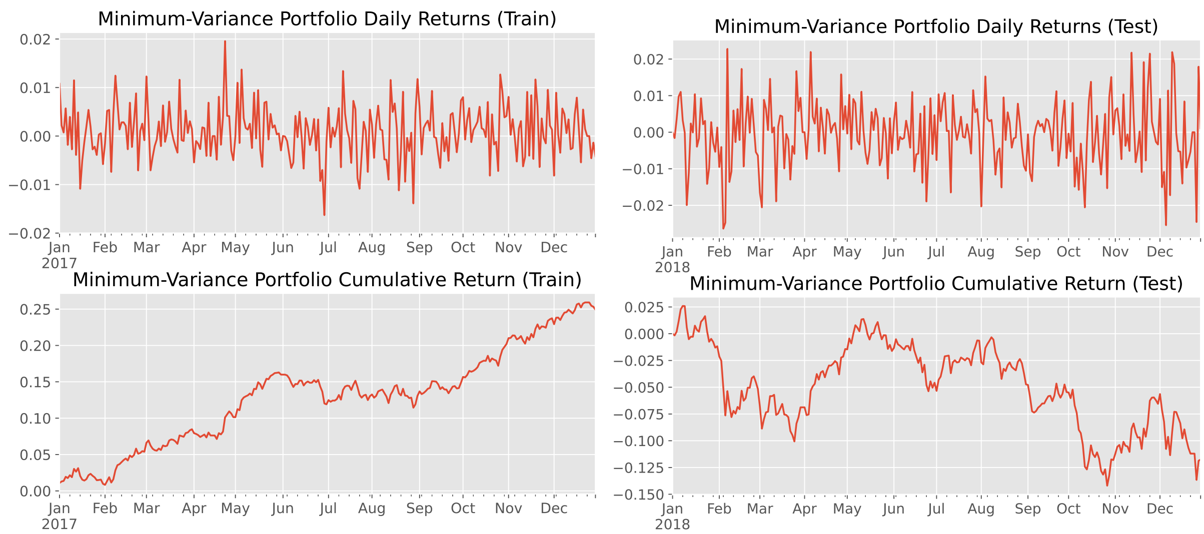


Figure 18: Minimum-variance portfolio returns.

We can see from the table that the returns of the returns from the training period of the minimum variance portfolio has the smallest variance at only 2.9×10^{-5} which corresponds to the theoretical minimal variance calculated from equation (22) in the section above. However, when we apply test data, the minimum variance portfolio performs worse than the equal weight portfolio in terms of volatility. This can be explained by the unsteady nature of financial markets and highlights the need for a more sophisticated model.
We can also note that during the training periods, both models showed upward tending cumulative returns, while these cumulative returns were down-trending during the test periods. This could highlight

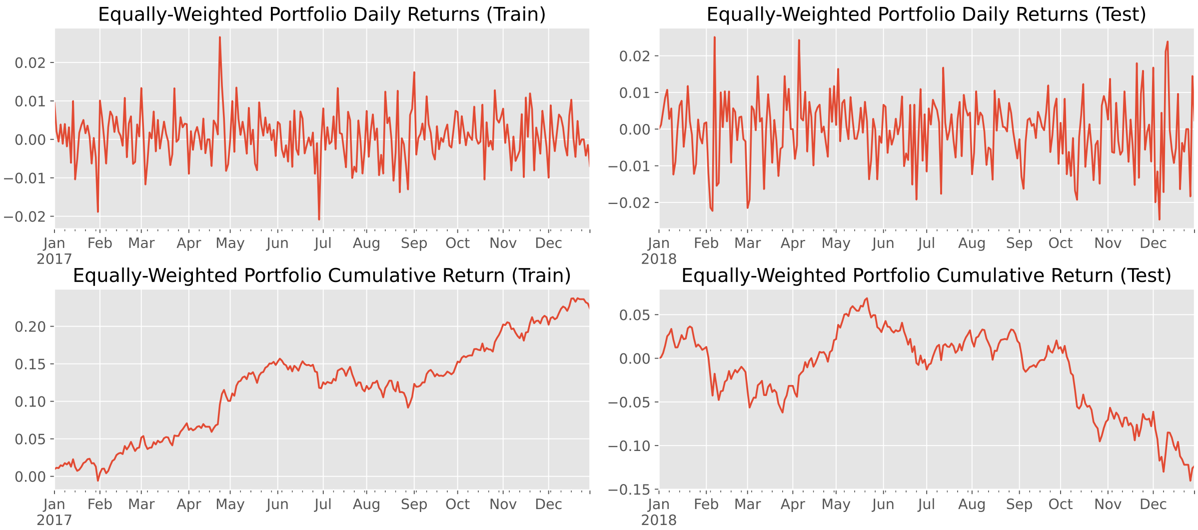


Figure 19: Equal weighting portfolio returns.

	Minimum variance portfolio		Equal weighting portfolio	
	Train	Test	Train	Test
Mean Return	0.000958	-0.000451	0.000860	-0.000473
Cumulative Return	0.249062	-0.117582	0.223484	-0.123503
Variance of Returns	0.000029	0.000082	0.000037	0.000079
Sharpe Ratio	0.179072	-0.049774	0.140374	-0.053126

Table 6: MV and EW portfolio train and test performance statistics

the mean reversing nature of the stocks.

3.1.3 Adaptive minimum-variance portfolio

In this section, we implement an adaptive time-varying minimum-variance portfolio using a rolling window of length N days. In Figure 20, we plot the performance of the adaptive portfolios with varying windows lengths.
As we increase the window size, we see a clear downward trend in performance, suggesting that the more recent data performs better terms of cumulative returns to compute the optimal weights of the portfolio. However, we can see an improvement in terms of the variance of the portfolio as we increase the window length (although it stays flat beyond a 2 month window size).
In Figure 21, we also plot the daily returns, cumulative returns and value of the rolling weights for different window sizes, namely 22 days (one month), 66 days (one quarter) and 252 days (one year) and summarise the findings in Table 7.

Overall, we can see an improvement of performance for adaptive minimum variance portfolios compared to static minimum variance portfolios or equal weighting portfolios (in the testing periods). Only the adaptive MV with a rolling window of 22 days (one month) performs worse in terms of volatility (variance of returns).
In terms of the effect of the recursive update of the variables, we observe that even though the volatility of returns decreases when increasing the window size, the returns decrease and turn negative. This is due to the fact that our optimisation problem has been defined to minimise the variance without any constraint on target returns and hence they are not taken into account. Therefore, we observe

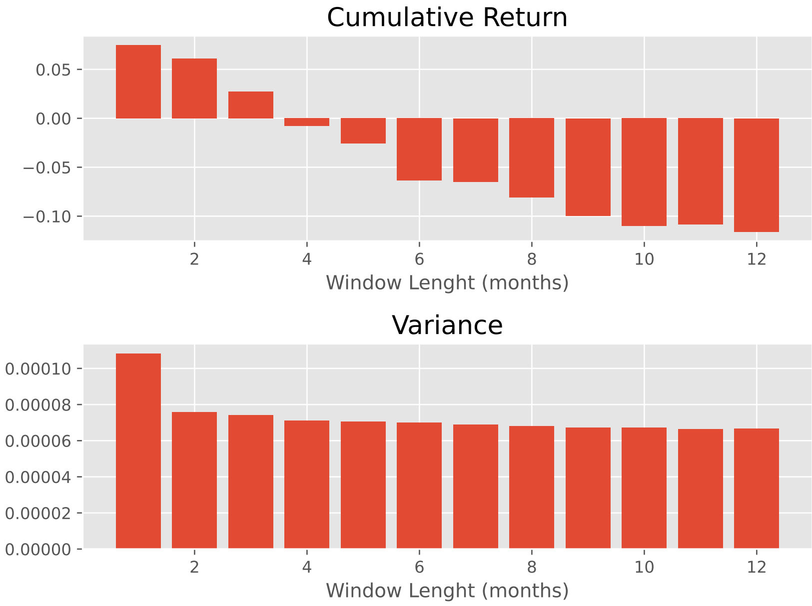


Figure 20: Cumulative return and variance for different window lengths.

poor performance in terms of returns.

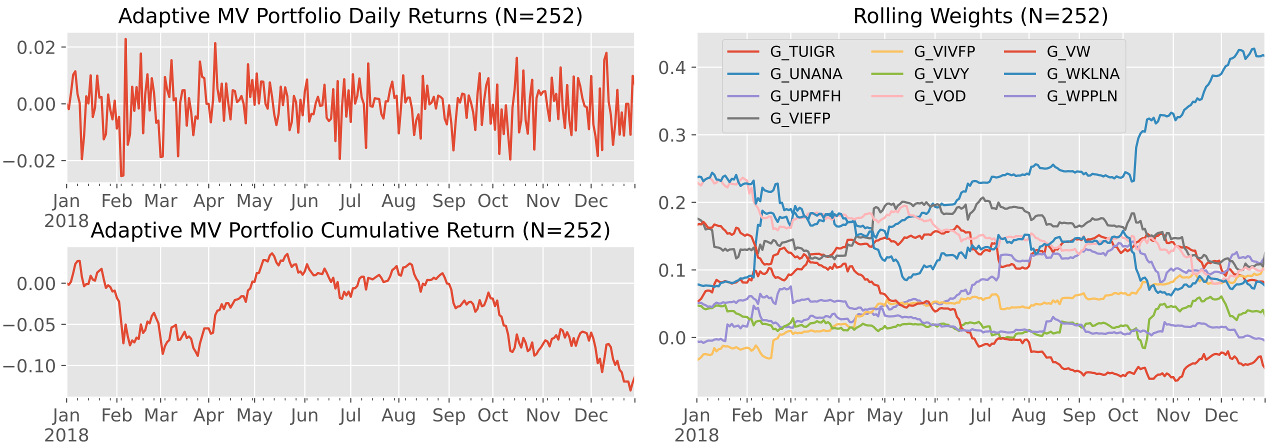
	Adaptive minimum variance portfolio		
	N=252	N=66	N=22
Mean Return	-0.000436	0.000104	0.000286
Cumulative Return	-0.113804	0.027190	0.074774
Variance of Returns	0.000066	0.000074	0.000108
Sharpe Ratio	-0.053534	0.012113	0.027537

Table 7: Adaptive MV portfolio performance statistics for three chosen window lengths.

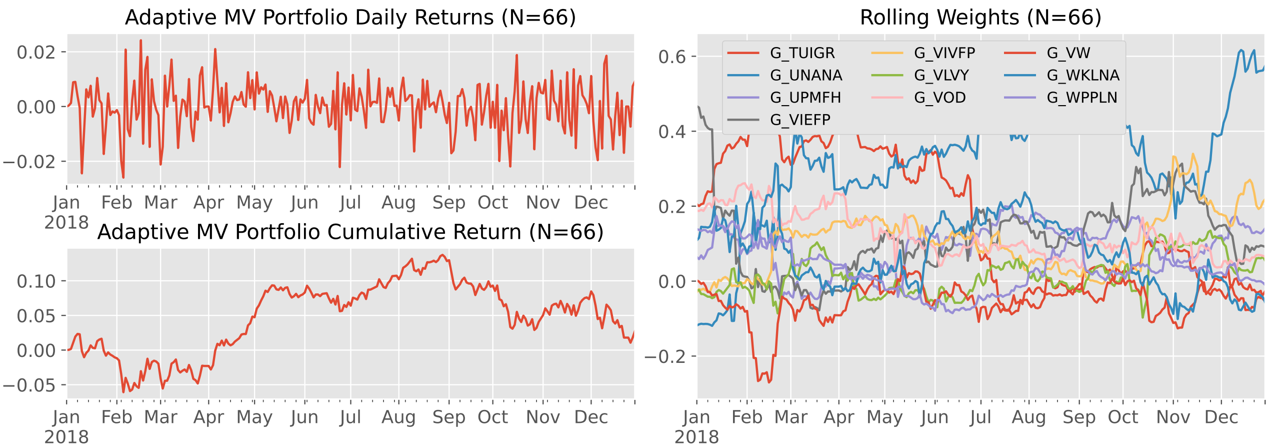
Finally, it is possible to calculate the covariance matrix of the returns using other methods. One of them could be to use the most recent N returns of each company using the following formula:

$$\hat{\mathbf{\Sigma}}[t] = \frac{1}{N} \sum_{r=t-N}^{t-1} (\mathbf{r}[\tau] - \hat{\mu}[t])(\mathbf{r}[\tau] - \hat{\mu}[t])^T \tag{62}$$

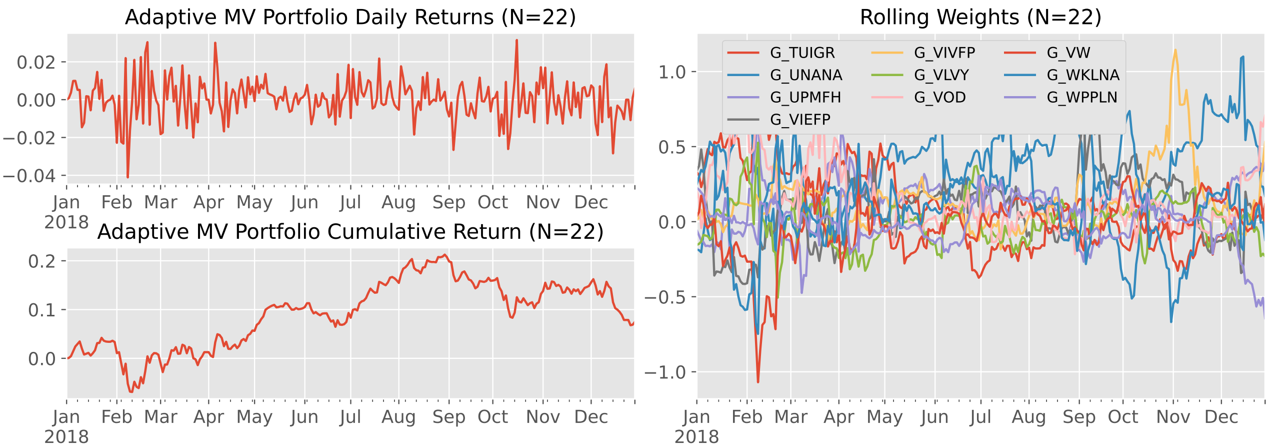
We can also change the calculation by assigning different weights to different days. That way, we could allocate the weights in an exponentially weighted manner to put more emphasis on more recent returns. This could further enhance the model however, it would also introduce new tunable parameters (exponential weight decay).



(a) Window length of 252 days (one year)



(b) Window length of 66 days (one quarter)



(c) Window length of 22 days (one month)

Figure 21: Adaptive MV portfolio returns and rolling weights

4 Robust Statistics and Non-Linear Methods

In this section, we will study different robust statistical techniques based on the data of three selected stocks (AAPL, IBM and JPM) as well as one index (DJI). We will use the following data from the period 16-03-2018 to 11-03-2019: open, close, high, low, adj. close.

4.1 Data Import and Exploratory Data Analysis

4.1.1 Descriptive statistics of stocks

AAPL	Open	High	Low	Close	Adj Close	Volume
Mean	187.69	189.56	185.82	187.71	186.17	32704750
Median	186.29	187.40	184.94	186.12	184.35	29184000
Stddev	22.15	22.28	22.01	22.16	21.90	14179721
MAD	15.89	15.61	15.92	15.94	15.48	7573900
IQR	36.00	36.34	36.06	36.76	35.69	16311700
Skewness	0.26	0.30	0.22	0.26	0.29	1.74
Kurtosis	-0.91	-0.92	-0.92	-0.93	-0.93	4.35

IBM	Open	High	Low	Close	Adj Close	Volume
Mean	138.45	139.49	137.33	138.36	134.90	5198937
Median	142.81	143.99	142.06	142.71	138.57	4237900
Stddev	12.11	11.91	12.20	12.03	10.67	3328955
MAD	5.27	5.31	5.19	5.23	4.49	920700
IQR	15.38	14.72	16.34	15.50	14.10	1952950
Skewness	-0.68	-0.62	-0.71	-0.68	-0.81	3.19
Kurtosis	-0.59	-0.62	-0.56	-0.58	-0.42	11.80

JPM	Open	High	Low	Close	Adj Close	Volume
Mean	108.71	109.65	107.68	108.61	107.26	14700689
Median	109.18	110.53	107.79	109.02	107.22	13633000
Stddev	5.36	5.20	5.43	5.30	4.83	5349770
MAD	4.47	4.31	4.24	4.35	3.45	3035400
IQR	8.81	8.85	8.85	8.83	7.22	6233600
Skewness	-0.42	-0.38	-0.38	-0.37	-0.34	1.69
Kurtosis	-0.32	-0.54	-0.27	-0.40	-0.11	4.43

DJI	Open	High	Low	Close	Adj Close	Volume
Mean	25001.26	25142.04	24846.00	24999.15	24999.15	332889442
Median	25025.58	25124.10	24883.04	25044.29	25044.29	313790000
Stddev	858.83	815.20	903.30	859.13	859.13	94078038
MAD	543.54	537.62	601.57	590.72	590.72	50460000
IQR	1109.43	1077.82	1204.42	1158.16	1158.16	108930000
Skewness	-0.37	-0.24	-0.46	-0.38	-0.38	1.74
Kurtosis	0.49	0.12	0.56	0.40	0.40	5.86

Table 8: Key descriptive statistics of the four assets.

For each of the stock and each column we generate key statistics to summarise the distribution of the data which are: mean, median, standard deviation, median absolute deviation (MAD), interquartile range (IQR), skewness (to measure the asymmetry) and kurtosis (measure for the tail) that are presented in Table 8. We also add the one day returns as a new column to use later on.

We observe that the means and medians are close to each other within columns of a stock, whereas the standard deviation and median absolute deviation are more separated in terms of value. Additionally, we note that the median standard deviations are generally smaller than the standard deviations and that it is close to half of the IQR in most cases.

4.1.2 Pdf of adjusted close and returns

In Figure 22, we plot the probability density functions of the adjusted 1 day close prices returns. As we previously discussed in section 1.1.3, returns tend to follow a Gaussian distribution when prices don't and this can be observed in Figure 22. When a distribution of a signal is not Gaussian, it may not be appropriate to represent that signal with its mean and standard deviation. In this case, we can use other statistics such as the median and median absolute deviation which would give use a more robust presentation of the signal of interest.

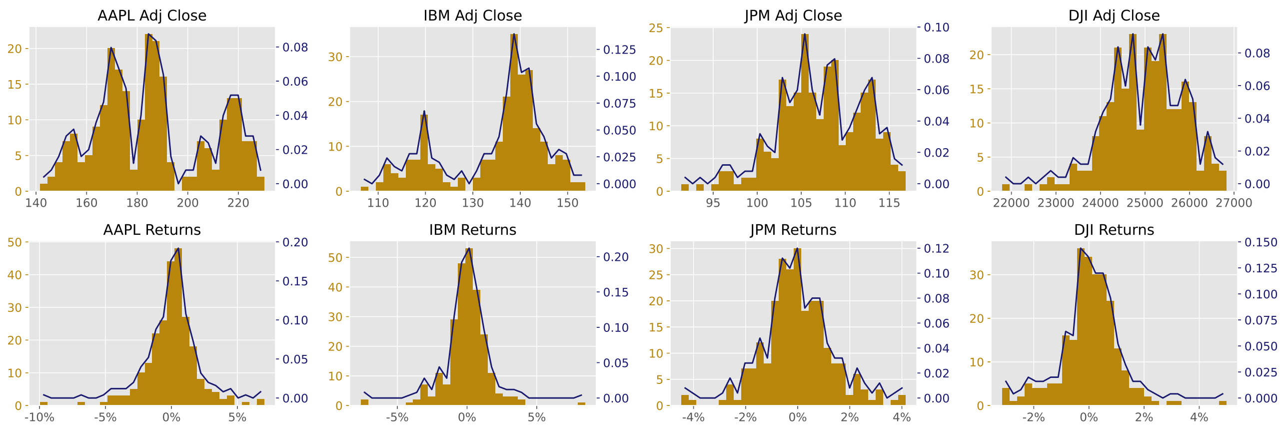


Figure 22: Probability distribution function of the 1 day close prices and returns.

4.1.3 Rolling mean and median statistics

In Figure 23,k we plot the adjusted close prices with its associated **rolling mean** (using a 5 days window) and the $\pm 1.5\times$ **standard deviations** relative to the rolling mean for each of the asset. We repeat the same operation in Figure 24 using the **rolling median** and the $\pm 1.5\times$ **standard deviations** associated to it.

The rolling mean (Figure 23) and rolling median (Figure 24) both represented in red are very similar to each other. However, we observe that the rolling deviations related to the median is much tighter and closer to the price action that the one of the mean. In Table 9 we compare the number of outlier points in each method and can confirm that indeed the median method contains more points outside that deviation area as it is more constrained. This suggests that mean estimators are more susceptible to price changes than the median estimators.

	mean method	median method
AAPL	30	103
IBM	31	94
JPM	33	105
DJI	30	97

Table 9: Number of outlier points for each asset using two methods.

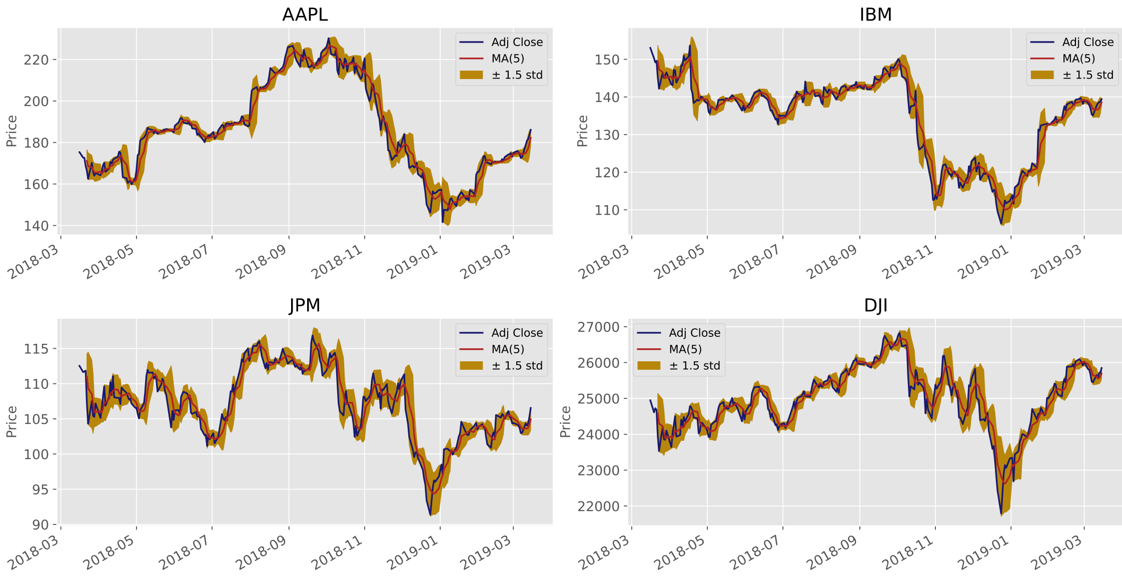


Figure 23: Adj. close prices, its rolling mean and the rolling $1.5 \times \sigma$ bound.

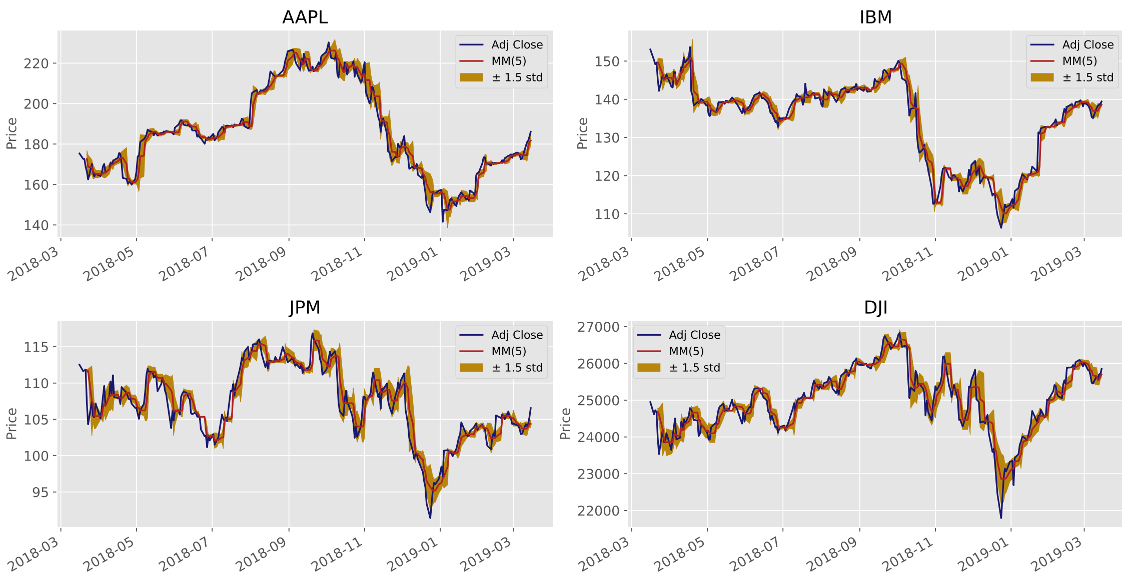


Figure 24: Adj. close prices, its rolling median and the rolling $1.5 \times \sigma$ bound.

4.1.4 Robustness to outliers

In this section, we introduce four outlier points for the adj. close with a value equal to 1.2× the maximum value of the column at 4 dates. We want to observe the impact of those points on the graphs using the statistics of part 4.1.3.

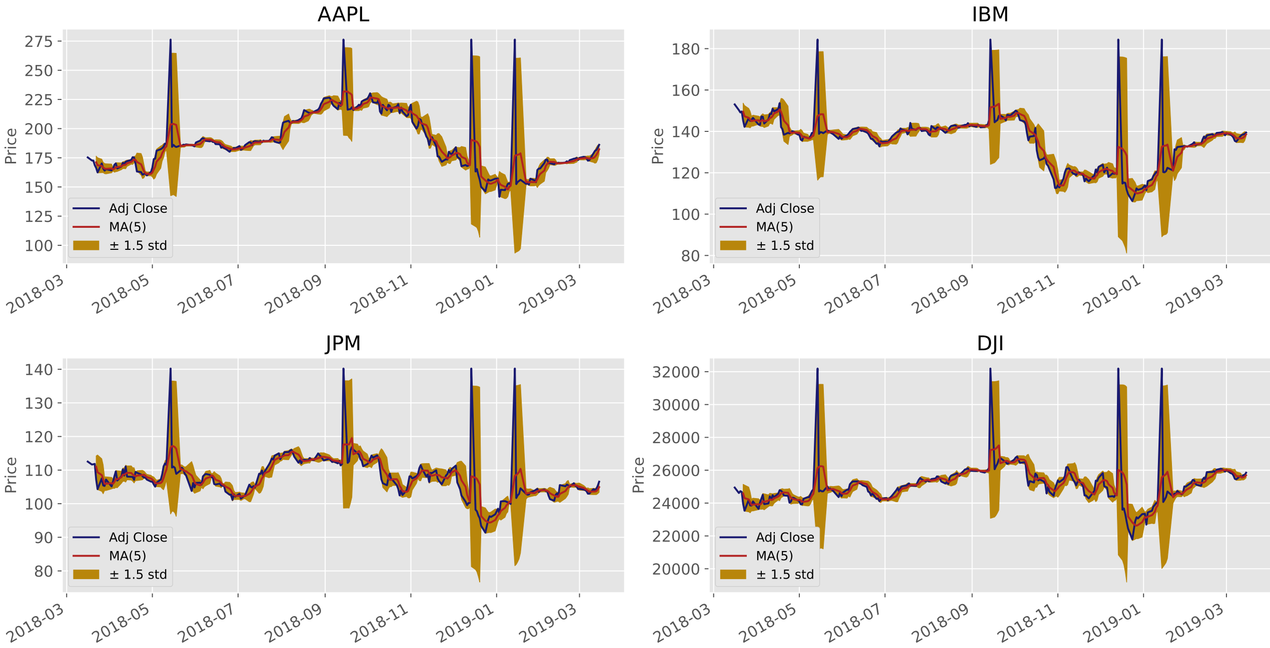


Figure 25: Adj. close prices, its rolling mean and the rolling $1.5 \times \sigma$ bound (with outliers).

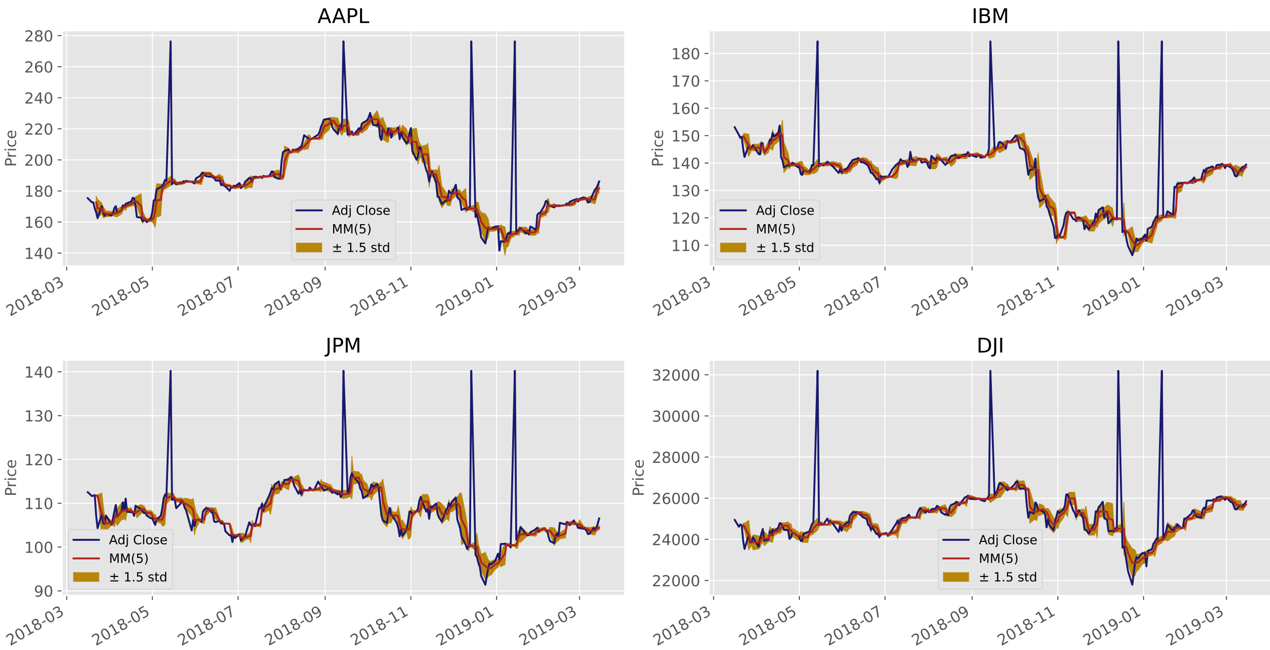


Figure 26: Adj. close prices, its rolling mean and the rolling $1.5 \times \sigma$ bound (with outliers).

We can observe from Figure 25 that the plot is largely affected by outlier points with a big standard deviation while the one of the median in Figure 26 seems unaffected in its rolling standard deviation. The latter seems more robust in sudden changes of price. We also observe that the rolling mean reacts more abruptly to these price changes that the rolling median.

These observations suggest that the median estimators are more robust against noisy data that contains temporary outlier points, while mean estimators are more affected and susceptible to these changes.

4.1.5 Box plots of prices

In order to better visualise distribution, dispersion and the key statistics of the data we can use another tool called box plot. We use it to graphically visualise the dispersion of our adj. close prices in Figure 27.

The blue line at the center of the box denotes the median of the data. The rectangular box represents the interquartile range (from Q1 where 25% of the data can be found to Q3 where 75% of the data can be found) while the whiskers extend to represent the total range of the price data. From these are excluded the outliers which are represented as points and are defined to be outside of the 1.5× the IQR.

Overall, we see that the adj. close price data tends to be skewed with asymmetric quartiles and multiple outlier points. This confirm that the classic Gaussian models are unfit to describe price distributions.

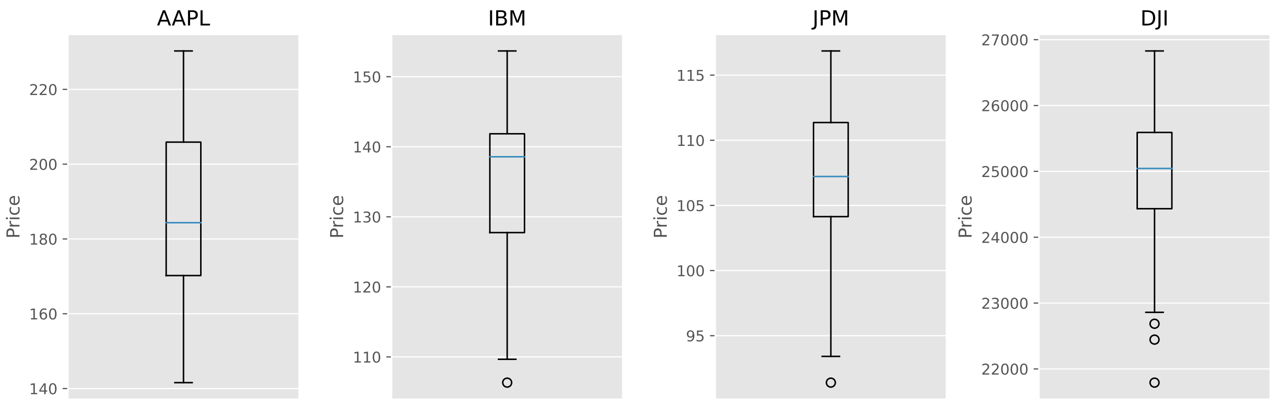


Figure 27: Box plots of close prices for each of the four assets.

4.2 Robust Estimators

In this section, we are going to implement, analyser and assess the following estimators:

- Robust location estimator: median.
- Robust scale estimator: IQR (interquartile range) and MAD (median absolute deviation)

4.2.1 Python Implementation

We first start by creating Python functions for the estimators:

Listing 1: Python example

```
def Custom_Median (input):
    tmp = input.sort_values() # sorting
    median = tmp[round(len(tmp)/2)] # find middle values
    return median

def Custom_IQR (input):
    tmp = input.sort_values() # sorting
    q1 = tmp[round(len(tmp)/4)] # 25th percentile value
    q3 = tmp[round(len(tmp)/4*3)] # 75th percentile value
    iqr = q3-q1
    return iqr

def Custom_MAD (input):
    median = Custom_Median (input)
    deviations = abs(input - median)
    mad = Custom_Median(deviations) # compute the median of deviations
    return mad
```

The logic for each of the functions is the following:

- Median: we first sort the input series and take the middle value as the median of that series. We round up the index to cover both cases when the length of the input is odd and even.
- IQR: we sort the input series and find the 25th and 75th percentile values (corresponding to the the values where 25% and 75% of the data can be found respectively). The interquartile range is the absolute difference between these two values.
- MAD: in order to find the median absolute deviation, we first find the median value. We then create a new series called `deviations` which is calculated from the absolute difference between the median and each element of the original input series. The MAD is the median of the newly created series `deviations`

4.2.2 Complexity Analysis

We will now study the computational complexity of the estimators we discussed:

- Median: we first perform a sorting operation which takes $\mathcal{O}(nlog(n))$. Then we find the middle value of the sorted series, which is a $\mathcal{O}(1)$ operation. Hence, the overall complexity of the median estimator is $\mathcal{O}(nlog(n))$.

- IQR: sorting the series takes a complexity of $\mathcal{O}(nlog(n))$. We then find two individual value in the 25th and 75th percentile value and perform a subtraction which are all $\mathcal{O}(1)$ operations. The overall complexity of the IQR estimator is $\mathcal{O}(nlog(n))$.
- MAD: computing the median of a series takes a $\mathcal{O}(nlog(n))$ complexity as seen above. From there, we can compute our new `deviations` series at a $\mathcal{O}(n)$ complexity. Computing the median of that series is again $\mathcal{O}(nlog(n))$ complexity. The overall complexity of the MAD estimator is $\mathcal{O}(nlog(n))$.

We can note that the computational complexities of these estimators are much higher than those of more classic estimators such as the mean or the standard deviation which only take $\mathcal{O}(n)$.

4.2.3 Breakdown Points

A robust estimator should not be affected too much by large deviations from the model. In order to define the robustness of an estimator, we use its breakdown point which is the maximum fraction of outliers that an estimator can tolerate and it has a value between 0 and 0.5. The breakdown point is the percentage of data (expressed in decimal terms) that can be contaminated without deteriorating the estimators results. The higher the breakdown point, the more robust our estimator is. We observe the results for our estimators and the more traditional ones in Table 10.

Estimator	Mean	Std	Median	MAD	IQR
Breakdown Point	0	0	0.5	0.5	0.5

Table 10: Breakdown Point for different estimators.

The median has a breakdown point of 0.5 which means it is a very robust estimator against outliers, which we had confirm in section 4.1. Since the median is the middle value in the set, even if half of the dataset was contaminated it would still be able to estimate correctly a valid point within the dataset. This contrasts with the mean which has a breakdown point of 0 as it is an average and by definition very susceptible to outliers as even one extreme value can falsify the estimation.

The median absolute deviation (MAD) also has a breakdown point of 0.5. The argument is similar to the one of the median as even if half of the dataset had extreme values, the median would still give a correct estimation and half of the `deviations` series would still stand correct and the median of that series would stand intact too. Once again, this contrasts with the standard deviation which gets affected by a single outlier.

Finally, the interquartile range (IQR) has a breakdown point of 0.25. As it estimates the 25th and 75th percentiles (equivalent to splitting the data in half and finding the median of each half series) then it can allow a quarter of the data to be contaminated (half for either series) before breaking down.

Overall, we note a trade-off between the robustness to outliers and the computational complexity in our estimators.

4.3 Robust and OLS Regression

4.3.1 OLS regression

In this section, we have regressed each stock's 1-day returns against the 1-day returns of the DJI index using the Ordinary Least Squares (OLS) regression. We can define this regression as:

$$\mathbf{r}_{asset} = \mathbf{X}\mathbf{b} + \mathbf{e} \tag{63}$$

which can be rewritten:

$$\mathbf{r}_{asset} = \begin{bmatrix} \mathbf{1} & \mathbf{r}_{DJI} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \mathbf{e} \tag{64}$$

where $\mathbf{r}_{asset}$ is the vector representing the daily return of our considered asset (stock), $\mathbf{r}_{DJI}$ is the vector represent the daily returns of the Dow Jones, α and β are the regression parameters and finally $\mathbf{e}$ is the error vector.

The OLS method consists in minimising the squared error given by $\|\mathbf{e}\|^2 = \|\mathbf{r}_{asset} - \mathbf{X}\mathbf{b}\|^2$. The optimal solution is the one that minimises the loss function and is given by $\hat{\mathbf{b}} = (\mathbf{X}^\mathbf{T}\mathbf{X})^{-1}\mathbf{X}^\mathbf{T}\mathbf{r}_{asset}$. The regression parameters obtained for each stock are given in Table 11a and the predictions of the regressor are given in Figure 28 alongside the true value of the returns.

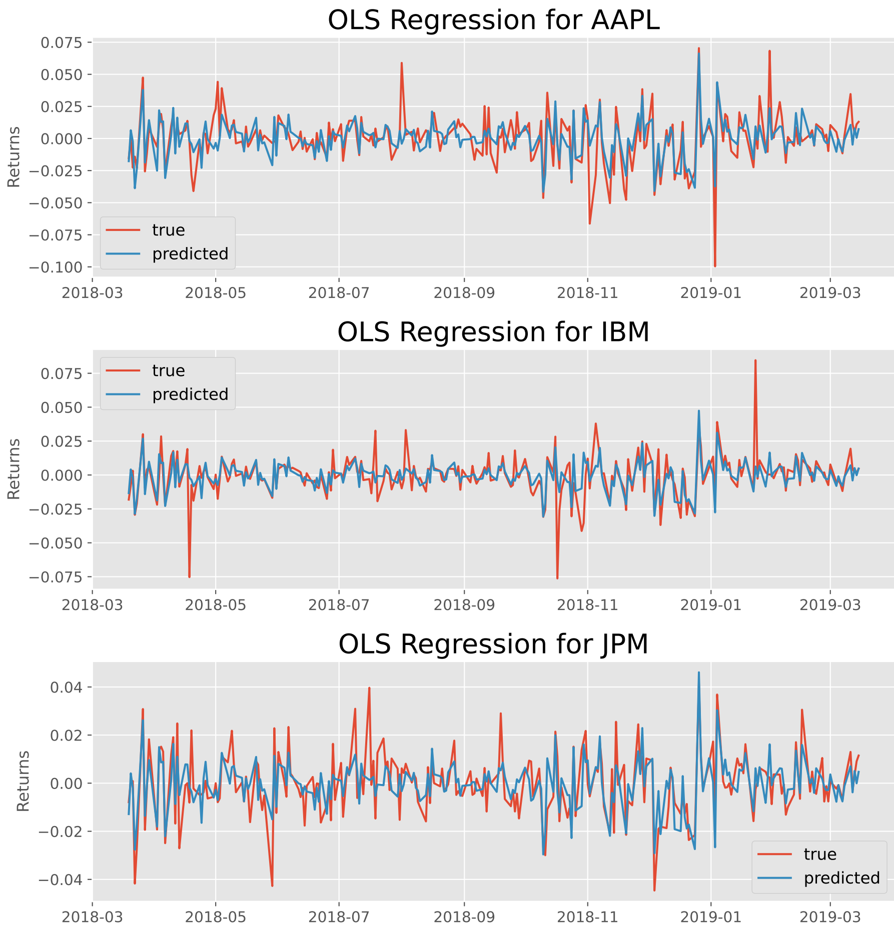


Figure 28: Predictions using the OLS regression.

4.3.2 Robust regression (Huber regression)

The Hubert regression is similar to the OLS regression except for the loss function to be minimise. This loss function J is given by:

$$J(y, \hat{y}) = \begin{cases} (y - \hat{y})^2, & \text{if } \frac{|y-\hat{y}|}{\sigma} \leq \epsilon \\ |y - \hat{y}|, & \text{otherwise} \end{cases} \tag{65}$$

The Huber regressor aims to minimise the squared loss for the samples where $\frac{|y-\hat{y}|}{\sigma} \leq \epsilon$ and the absolute error otherwise. This means that, contrarely to the OLS regression, the loss function is not heavily impacted by the outliers while still not ignoring its effects. The parameter σ is used to be able to normalise the fraction in case y get scaled up or down, then ϵ does not need to be scaled too and can be used to control the error allowed to consider a point to be an outlier.

Figure 29: Predictions using the Huber regression.

4.3.3 Huber vs. OLS regression

The coefficients obtained using the different regression methods have shown similarity in the results and so did the predictions when using both regressors. To be able to further test and compare the robustness of the two regressors, we introduce four outlier points at certain dates with values equivalent to 1.2× the maximum value of the series (similar to what we did in section 4.1.4). The obtained coefficients using both methods with the contaminated data are given in Table 11c and Table 11d. The predictions of both regressors with the outliers are shown in Figure 30.

	AAPL	IBM	JPM
α	0.00016466	-0.000440572	-0.000316331
β	1.32558	0.960092	0.931408

(a) OLS regression

	AAPL	IBM	JPM
α	-0.000130371	-0.000509435	-0.000800961
β	1.27021	0.973562	0.919662

(b) Huber regression

	AAPL	IBM	JPM
α	0.00177777	0.00127644	0.000479203
β	1.23342	0.87377	0.887547

(c) OLS regressionwith outliers

	AAPL	IBM	JPM
α	0.000324306	-0.000341806	-0.000704428
β	1.24487	0.972195	0.920989

(d) Huber regression with outliers

Table 11: Breakdown Point for different estimators.

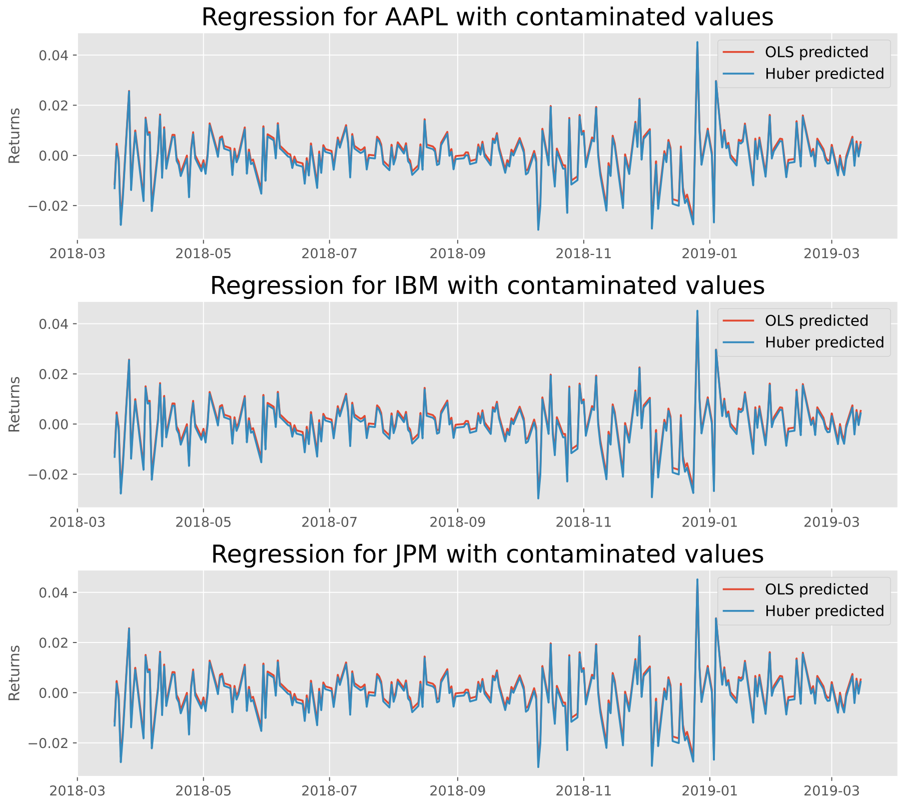


Figure 30: Box plots of close prices for each of the four assets.

We observe from Figure 30 that the predictions of both regressors with the outlier points are very similar. Since we only added 4 points, this small impact in performance is to be expected. However, if we look at Table 11 and the changes of the coefficients when adding the outlier points, we observe that the Huber regression displays a more robust behaviour. Adding these four outlier points has a significant impact on the coeffcients of the OLS regressor while the ones of the Huber regressor stay almost intact.

As mentionned in section 4.3.2, since the Huber regression limits the impact of very large errors in the estimation, it is more robust to outliers than the OLS regression. This has been demonstrated in this section as well.

4.4 Robust Trading Strategies

The Moving Average Crossover is a simple trading strategy that consists of the following rules:

- Buy X shares of a stock when its 20-day MA $>$ 50-day MA
- Sell X shares of the stock when its 20-day MA $<$ 50-day MA

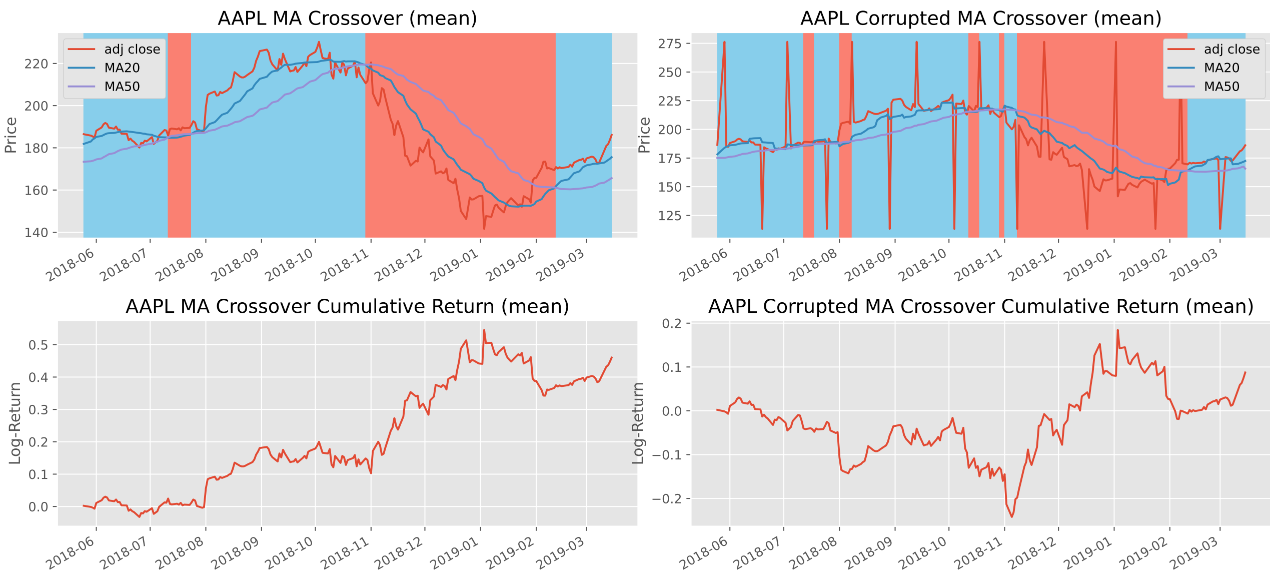
In this section, we will implement the simple and robust variation of the Moving Average Crossover Strategy.

4.4.1 Simple Moving Average Crossover

For each stock, we plot the adj. close as well as the 20-day and 50-day **rolling mean** moving averages on the left plot of Figure 31 as well as the buy (blue) and sell (red) regions which are highlighted following the crossover strategy. The performance of the strategy over time in terms of cumulative returns is also provided for each of the stock. We then repeat the same procedure but this time we include corrupted adj. close prices and we plot the corresponding results on the right plot of Figure 31. Table 12 displays the percentage of overlap between the two methods in order to quantify the robustness of the crossover strategy in the presence of outlier points.

Stocks	Region Overlaps (%)
AAPL	89.6
IBM	86.1
JPM	84.2
DJI	85.1

Table 12: Region overlaps percentages (rolling mean method).



We observe that the percentage of crossover, and hence the buy and sell decisions, are largely impacted by the presence of outliers which leads to significant differences in performance. This suggests the sensibility of rolling means to outliers and the need for a more robust strategy.

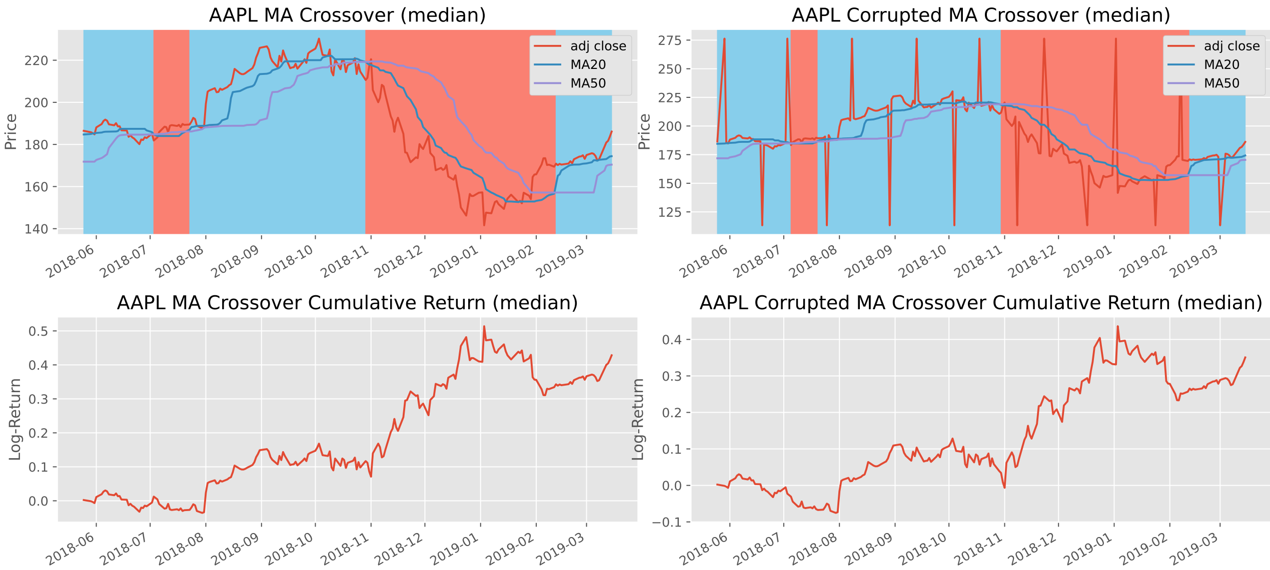
Figure 31: Mean average crossover strategy and buy and sell regions and cumulative returns with and without corrupted data for different assets.

4.4.2 Robust Moving Average Crossover

We repeat the strategy studied in section 4.4.1, this time using the **rolling median** instead of the rolling mean in order to compare the methods and show the results in Figure 32 as well as the percentage of crossover regions in Table 13.

Stocks	Region Overlaps (%)
AAPL	98.5
IBM	98.0
JPM	97.0
DJI	98.5

Table 13: Region overlaps percentages (rolling median method).



We observe from Table 13 and from the highlighting of the plots in Figure 32 that the rolling median has much more overlapping when adding the outlier points which suggests it is a much more robust method to contaminated data than the mean method. The buy and sell decisions, and consequently the cumulative returns, are almost identical when adding the outliers than when not. There, in accordance with our previous observations and conclusions, we define that using the mean to estimate the average of a signal in presence of outliers is not appropriate. Instead, the median estimator is a more robust solution to outliers and is defined to be a better choice in defining corrupted signals.

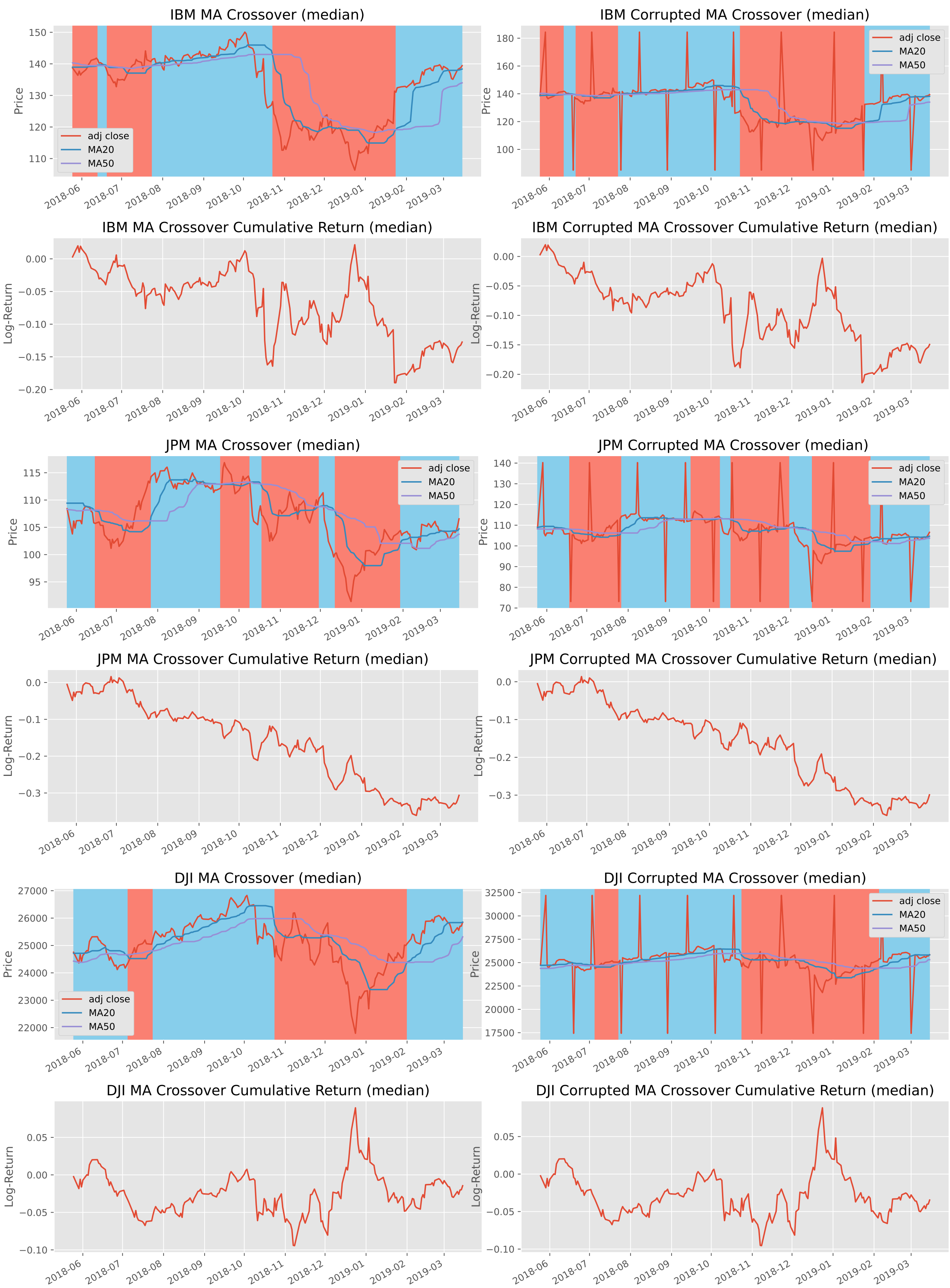


Figure 32: Median average crossover strategy and buy and sell regions and cumulative returns with and without corrupted data for different assets.

5 Graphs in Finance

5.1 S&P 500 Stock Selection

In this section, we start by selecting 10 stocks from the S&P 500. We choose to analyse stocks from the financial sector and particularly those headquartered in New York City. We selected the 10 companies matching those criteria with the largest market capitalisation. The details of the chosen companies can be found in Table 14.

The chosen financial companies are closely related by a number of different macroeconomics factors. First, since all of these institutions are located in the US and more precisely in New York, they are all affected by the same regulations and other geographical and socioeconomic factors. Their sector of activity draws obvious links between them as well. Finally, by including companies from different sub industries, we can also observe the relationships across and within different financial sub industries.

Symbol	**Security**	**GICS Sub Industry**
BLK	BlackRock	Asset Management & Custody Banks
AXP	American Express Co	Consumer Finance
JPM	JPMorgan Chase & Co.	Diversified Banks
C	Citigroup Inc.	Diversified Banks
SPGI	S&P Global, Inc.	Financial Exchanges & Data
MMC	Marsh & McLennan	Insurance Brokers
MS	Morgan Stanley	Investment Banking & Brokerage
GS	Goldman Sachs Group	Investment Banking & Brokerage
MET	MetLife Inc.	Life & Health Insurance
AIG	American International Group	Property & Casualty Insurance

Table 14: Selected assets along with their GICS Sub Industry.

5.2 Visualising Stock Correlations with Graphs

Graphs are a visualisation tools to represent the underlying relationships in a given dataset. We can construct a graph of our 10 selected stocks based on the correlation matrix $\mathbf{A}$, where each element of the matrix $\mathbf{A}_{ij}$ denotes the weight of the edge between the two stocks i and j.

The weight of an edge represents the strength of the connection / relationship between the vertices. Correlation between assets in a quantitative indication of their relationship to each other, hence we use the correlation matrix of our 10 chosen stocks, given in Table 15, and visualise these relationships using the graph in Figure 33. The correlation between two stocks i and j defines the weight of the edge between vertices i and j.

From Figure 33, we observe that the strength of an edge is represented both by its thickness and the darkness of its colour. The importance of a vertex is represented by its size and the vertices are naturally positioned in such a way that the strongly connected vertices are closer to each other. We have removed the edges where the weights were less than 0.5 and the edges for self-correlation.

	BLK	AXP	JPM	C	SPGI	MMC	MS	GS	MET	AIG
BLK	1.000	0.523	0.720	0.723	0.650	0.638	0.729	0.689	0.643	0.556
AXP	0.523	1.000	0.583	0.575	0.463	0.470	0.577	0.559	0.496	0.435
JPM	0.720	0.583	1.000	0.888	0.611	0.627	0.866	0.837	0.752	0.644
C	0.723	0.575	0.888	1.000	0.592	0.587	0.857	0.824	0.751	0.638
SPGI	0.650	0.463	0.611	0.592	1.000	0.587	0.618	0.595	0.505	0.460
MMC	0.638	0.470	0.627	0.587	0.587	1.000	0.596	0.581	0.546	0.531
MS	0.729	0.577	0.866	0.857	0.618	0.596	1.000	0.865	0.747	0.617
GS	0.689	0.559	0.837	0.824	0.595	0.581	0.865	1.000	0.699	0.598
MET	0.643	0.496	0.752	0.751	0.505	0.546	0.747	0.699	1.000	0.659
AIG	0.556	0.435	0.644	0.638	0.460	0.531	0.617	0.598	0.659	1.000

Table 15: Correlation matrix of the 10 selected stocks.

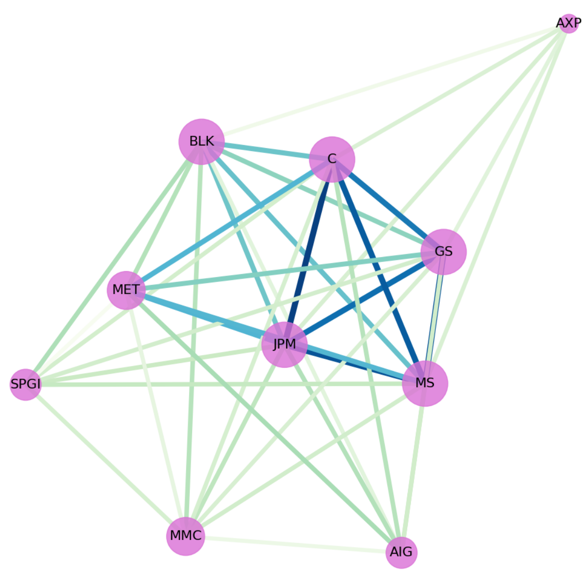


Figure 33: Graph of the 10 stocks based on the correlation matrix of their returns.

5.3 Graph Analysis

We can observe from Figure 33 that the nodes are positioned in such a way that the ones with strong connections are closer to each other, and the ones that have a weaker connection are further apart. For example, we can see that Gs, Ms, C and JPM which are highly correlated between each other are all packed in the same area of the graph. On the other hand, AXP, which exhibits the lowest correlation to the other stocks is positioned further appart, on its own. This positioning indicates that companies within the same GICS sub industry (JPM and C or GS and MS) exhibit a higher correlation than the companies in other sub categories (such as AXP).

The positioning of the nodes is done using the Fruchterman-Reingold force-directed algorithm, which simulates a force-directed representation of the network treating edges as springs holding nodes close, while treating nodes as repelling objects, and the simulation continues until the positions are close to an equilibrium. Furthermore, the edges are styled in such a way that the ones with larger weights are darker and thicker, while the others are lighter and thinner.

The degree of a node is defined by the number of edges it has and represents the importance of a node within the graph. The graph in Figure 33 illustrates this importance by changing the size of the nodes, meaning the nodes with higher degrees are larger than the others. For example, banks with diversified operations such as JP Morgan Chase (JPM) or Citigroup (C) hold central positions with larger nodes reflecting the diversified nature of their business. Since they operate in a variety of sectors, such as asset management, brokerage, retail and investment banking, they have high correlations with most of the companies in the financial sector. On the other hand, a company like American Express (AXP), which main business is centered around debit and credit cards and does deal with retail and investment banking activities, they exhibit lower correlation with the other companies and therefore are placed further away in a smaller node.

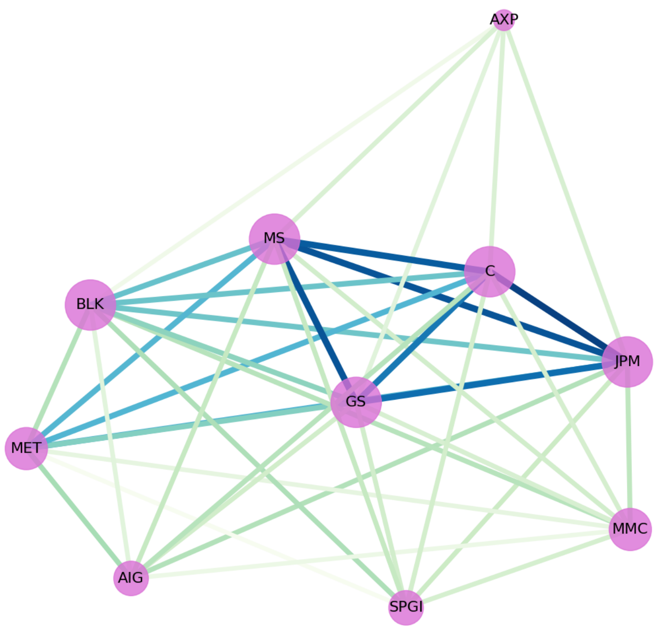


Figure 34: Graph of the 10 stocks based on the correlation matrix of their returns (re-ordered).

Finally, we note that the re-ordering of the graph vertices or the re-ordering of the timer-series would not affect the structure of the graph. As previously mentioned, a graph is defined by its adjacency matrix, where each column or row represents a certain vertex, and each element in the adjacency matrix represents the weight of the edge between vertices. Therefore, the re-ordering of the vertices would change the position of the elements in this matrix but not their values and would not affect the resulting graph. Furthermore, because the re-ordering of the time series data would not have changed the corresponding correlation matrix, this would not affect the resulting graph either. To prove this point in practice, we re-order the graph vertices and the times-series data and generate the graph given in Figure 34, which exhibits a similar topology with exact same connections.

5.4 Spectral Similarity Graph

In the previous sections, we constructed the graphs using the correlations between the return of the assets. In this section, we will follow a similar process, but this time using the spectral similarity of the return series of the chosen assets in order to construct the graph. We plot the spectra of each asset in Figure 35 and the graph defined by the correlation matrix of the spectras is given in Figure 36. Note that the edges are only drawn for pair with correlation higher greater 0.25, and self-correlations are also omitted.

The motivation for using this measure is the fact that companies undergo certain business and economic cycles which can affect their returns. It is possible to visualize the similarity of the firms in this sense, that is which firms are affected by similar economic cycles, with a graph.

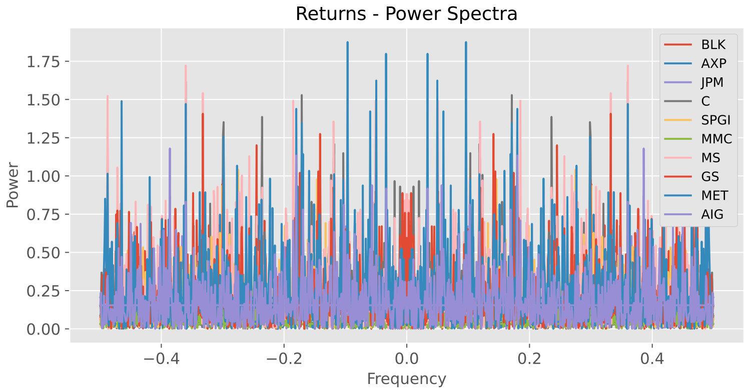


Figure 35: Spectra of the returns of the 10 selected stocks.

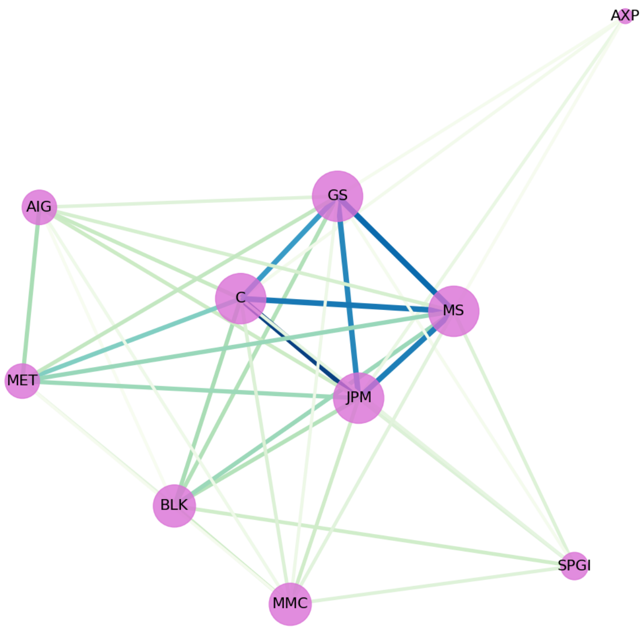


Figure 36: Spectral correlation graph of the 10 selected stocks.

Firstly, we can observe the position and weakness of the links of American Express (AXP) which highlights the previous observations about the nature of their business discussed in the previous section. More interestingly, this graph highlights the relationships between different business cycles. For instance, we observe JPM, C, GS, MS meaning they are all affected by the similar business cycles. The same thing can be observed with AIG and MET, both insurance companies and hence affected by their own particular business cycles.

In the previous section, we mentioned that the re-ordering of the time-series doesn't have an affect on the construction of the graph. This is not the case for the spectral graph constructed in this section because when the time-series is re-ordered, its spectra changes as it can be seen in Figure 37. Then, the values in the correlation matrix calculated for these spectra also changes, which as a result affects the weights of the graph to be constructed. We have the graph for the re-ordered series in Figure 38, which is similar to the graph we previously had, but it does have some discrepancies. Note that the re-ordering of the vertices still won't have any affects on the graph properties.

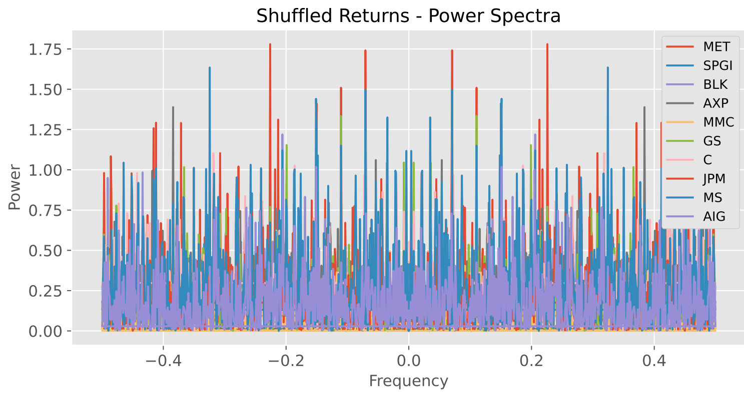


Figure 37: Spectra of the re-ordered returns of the 10 selected stocks.

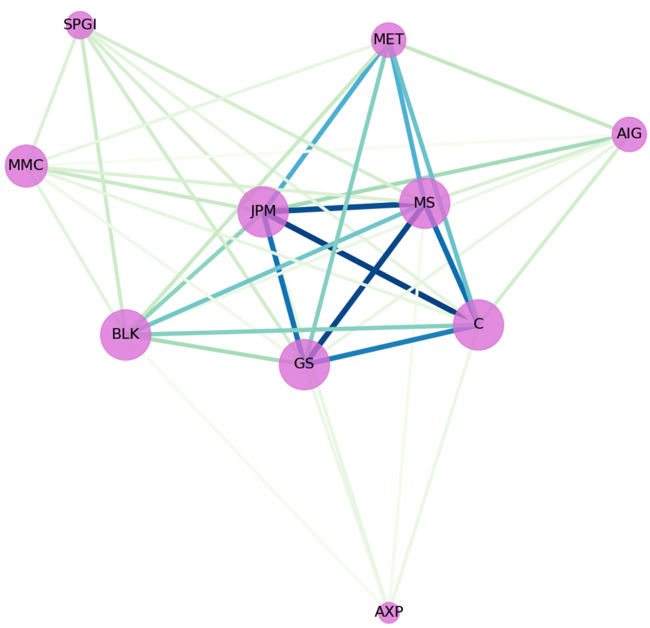


Figure 38: Spectral correlation graph of the 10 selected stocks. (re-ordered).

5.5 Using Prices vs. Returns for Graph Construction

In Figure 39, we constructed a graph using the correlation matrix of prices. Compared to the previous graphs in section 5.2, which were generated using the correlation matrix of returns, this graph does not convey much information about the intrinsic relationships of the considered companies. We only get little information about the general trends of the signals and don't learn much about the relationships and connections of firms between each other. In this case, we have only kept the edges with positive weights.

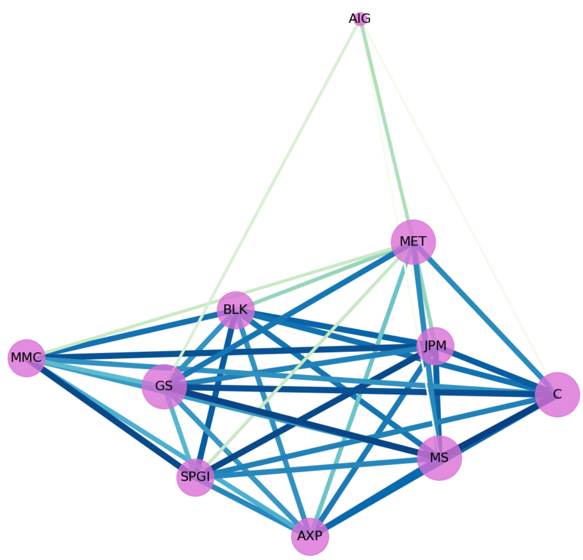


Figure 39: Prices correlation graph of the 10 selected stocks.

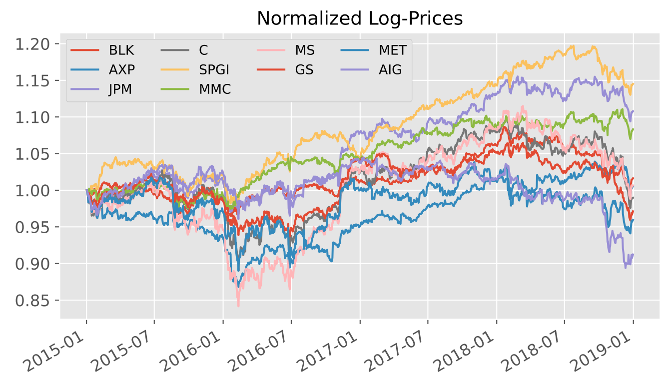


Figure 40: Normalised log-prices series for the 10 selected stocks.

In Figure 40, we have normalised the log-price signals. Observe that while most of the signals are upward trending, the price of AIG is downward trending and therefore has negative correlation with most of the other firms. This is visualized by the position of the AIG vertex in Figure 39 and the number of edges connected to it. Finally, because the price series have constant values in them, as they are not zero-mean signals, their spectra will be very high at low frequency values for all signals, making the entire analysis using spectra irrelevant. Therefore, we conclude that it would be more appropriate to use returns to conduct such analysis.