Department of Electrical and Electronic Engineering
Imperial College London

EE4-69 Signal Processing and Machine Learning for Finance
Coursework 2020

Desmond Sy

CID: 01210056

April 4, 2020

Contents

1 Regression Methods

1.1 Processing stock price data in Python

1. The S&P 500 Index price and logarithmic price time series are plotted in figure 1 below:

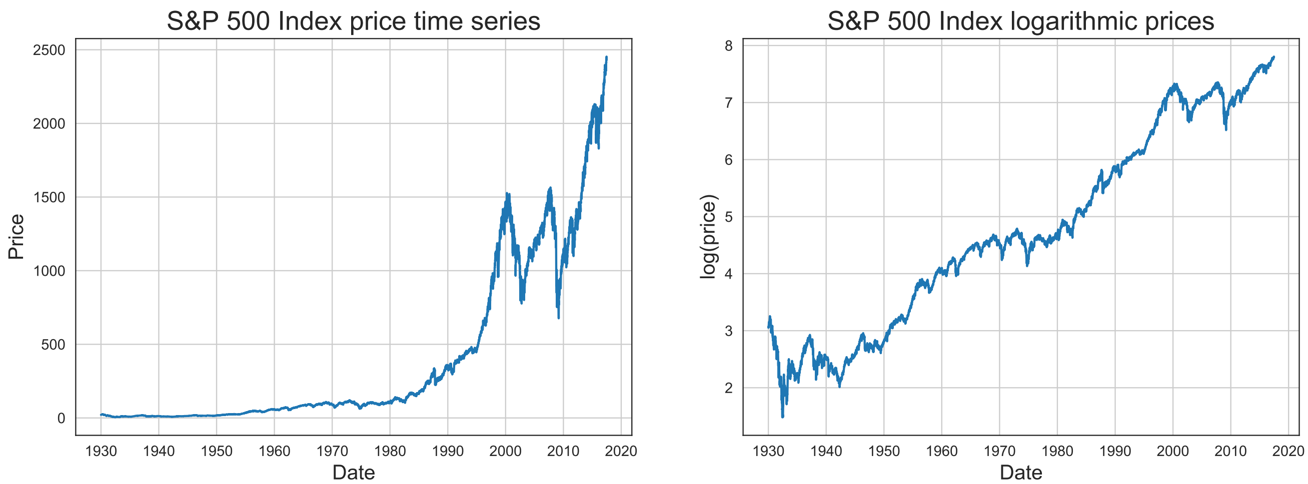


Figure 1: S&P 500 Index price time series (left) and logarithmic price time series (right)

2. Using a rolling window of 252 days, the mean and variance of the price and logarithmic price time series are plotted in figure 3 and figure 2:

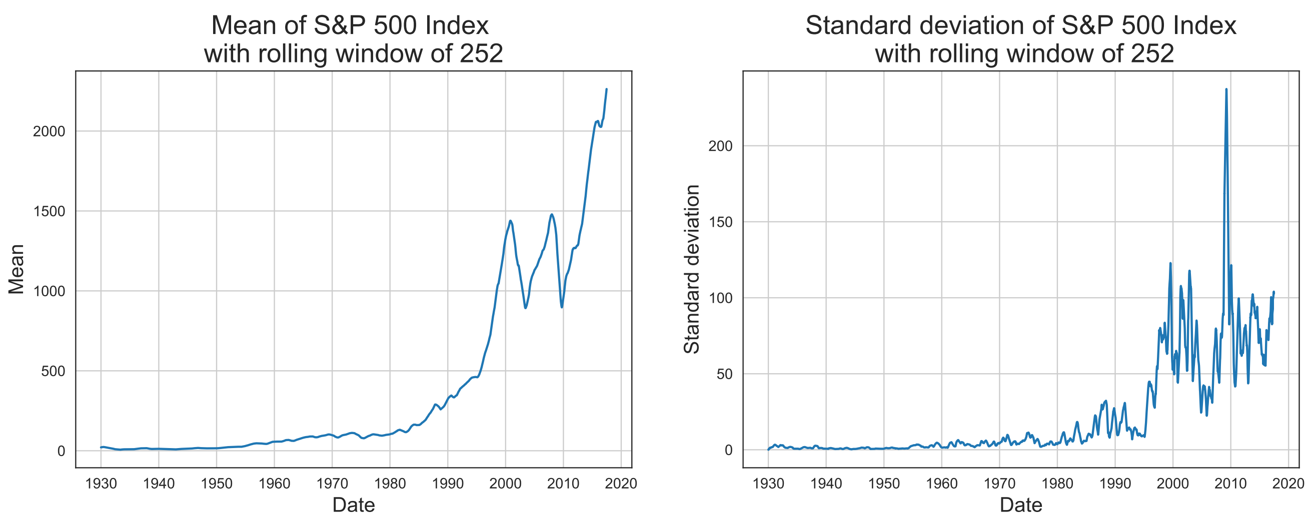


Figure 2: S&P 500 Index: Rolling mean of price time series (left) and rolling standard deviation of price time series (right)

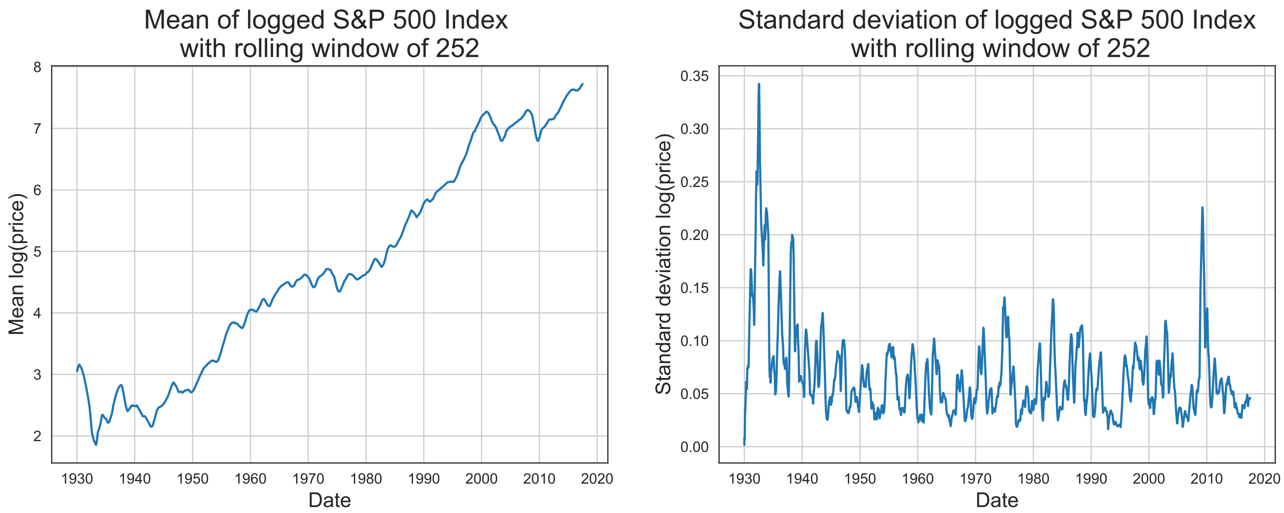


Figure 3: S&P 500 Index: Rolling mean of logarithmic price time series (left) and rolling standard deviation of logarithmic price time series (right)

The rolling price and logarithmic price time series suggests that S&P 500 may have been stationary in the past 30 years, but however this is an inaccurate observation as the scale of the plot in earlier time periods is too small compared with the maximum S&P 500 value. From the rolling mean and variance of the logarithmic time series, it can be seen that the mean of both plots are not constant with time, and therefore the S&P 500 is not a stationary function.

3. Simple return is defined as $R = \frac{P_t}{P_{t-1}} - 1$, and logarithmic return is defined as $\log(1 + R) = \log P_t - \log P_{t-1}$. The rolling mean and standard deviation of the simple returns and logarithmic returns are plotted in figure 4 and figure 5:

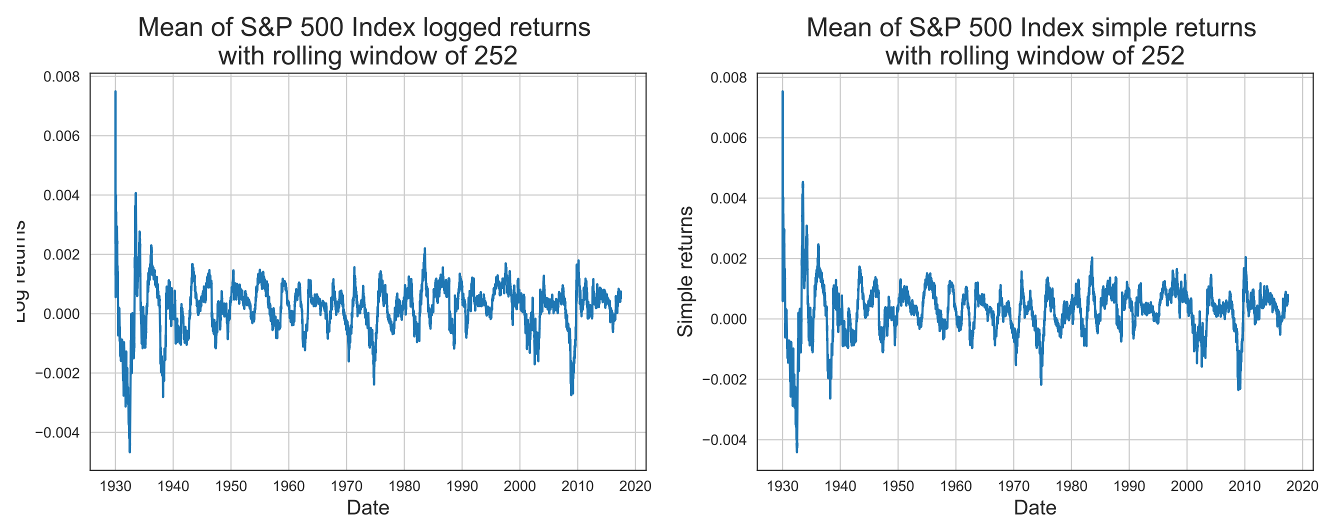


Figure 4: S&P 500 Index: Rolling mean of logarithmic returns and simple returns

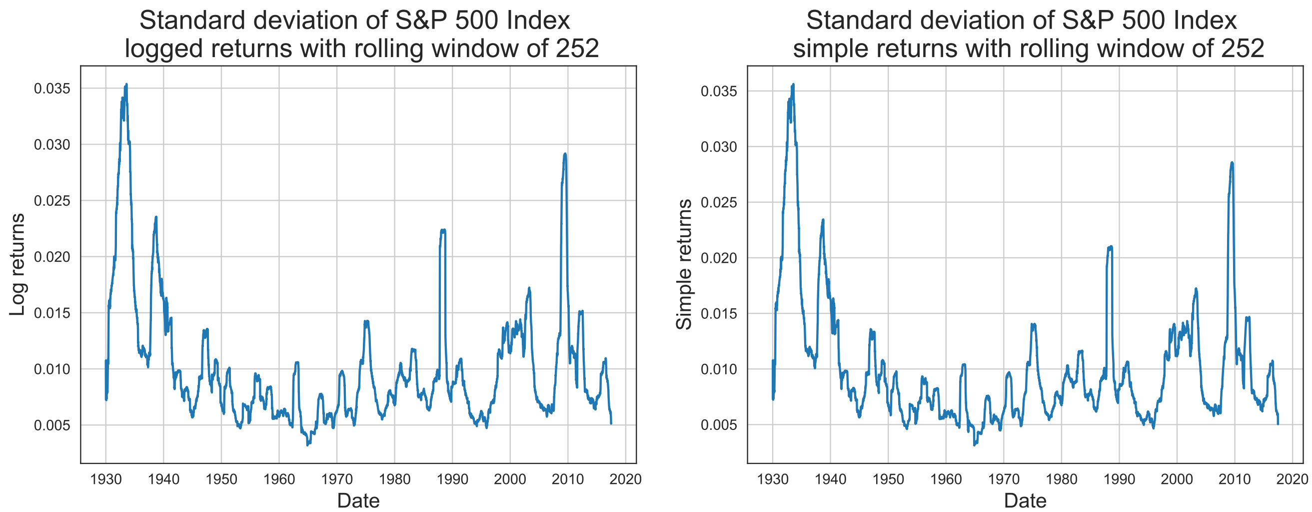


Figure 5: S&P 500 Index: Rolling standard deviation of logarithmic returns and simple returns

The sliding statistics of both simple and logarithmic returns are more stationary than the figures obtained in section 1.1.2. This is important because stationary processes are easier to analyse, and with stationary processes, the goal of forecasting the time series becomes much more straightforward.

4. There are several reasons why log returns are preferred over simple arithmetic returns. Firstly, logarithmic returns are additive, meaning that multi-period log returns can be summed up to get the total return, whilst simple returns are not time additive. Secondly, log returns are approximately stationary and invariant. The rationale behind this is that stock prices are assumed to follow a log-normal distribution, and so the arithmetic returns will also be. Taking a logarithmic transform results in normal returns, allowing the data to be easily interpreted.

The Jarque-Bera test is a goodness-of-fit test of whether sample data have the skew and kurtosis matching a normal distribution. The p-value for both the logarithmic and simple return series are 0, suggesting that the observed data is inconsistent with the assumption that the null hypothesis (the observed data is Gaussian) is true. This is confirmed with the following distribution plots:

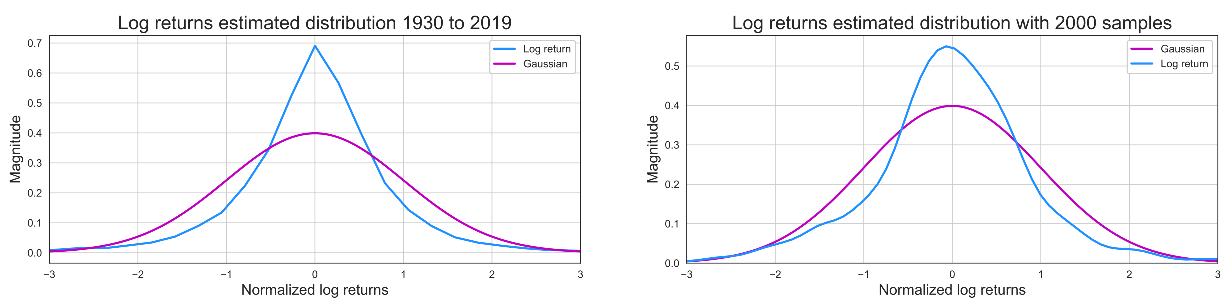


Figure 6: Distribution of logarithmic returns, from 1930 to 2019 (left), first 2000 samples (right)

As shown, the log return distribution does not follow a normal distribution. This is because when considering the entire time horizon, outliers in the data cause returns to be skewed. When considering a shorter time horizon of 2000 samples, we see that the distribution shape is closer to that of a normal distribution, as the samples are approximately stationary in the period.

5. The following table summarizes the differences between simple and logarithmic returns. Log returns are therefore more suitable for risk management, as it provides a more accurate and holistic picture of total returns, whereas simple returns creates a misleading positive return, when the overall change in stock price is \$0.

Time	Stock Price	Simple Returns	Log Returns
t=0	\$1	0	0
t=1	\$2	100%	0.693
t=2	\$1	-50%	-0.693
Total change	\$0	+50%	0

Table 1: Simple vs logarithmic returns

6. Logarithmic returns should only be used when dealing with intraday data. The assumption that returns follow a log-normal distribution over a longer time period is unrealistic because the log-normal distribution has positive skew, meaning that the right tail is fatter. However, returns data (for the S&P 500 as well as other major indices) is negatively skewed. This negative skewness indicates that investors expect frequent small gains, but a few large losses. In this case, simple returns should be used over logarithmic returns.

1.2 ARMA vs. ARIMA Models for Financial Applications

1. The S&P 500 time series is ploted in figure 7 below:

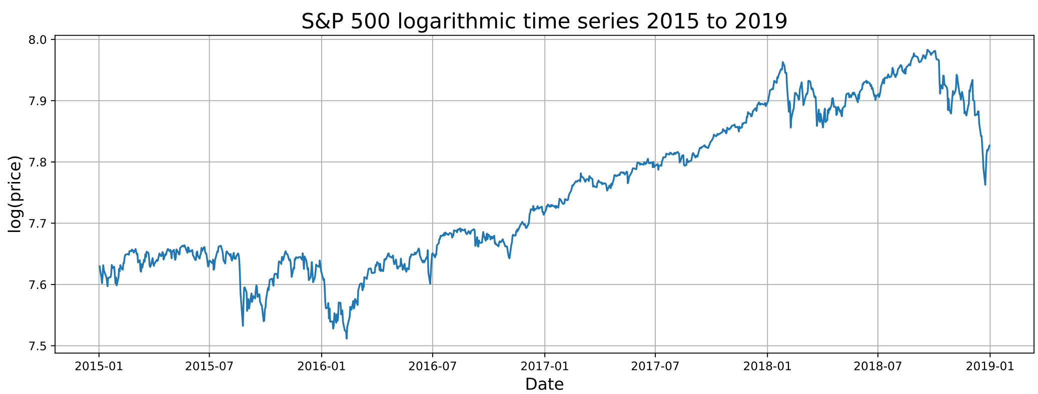


Figure 7: S&P 500 Index logarithmic price time series 2015-2019

The ARMA(p,q) model uses p autoregressive terms and q moving average terms. The AR part regresses the input signal $x[t]$ on its own lagged values, which can capture momentum and mean reversion effects observed in the markets. The MA part is a linear combination of white noise error terms occurring at various times in the past. This can be used to capture unexpected shocks observed as white noise.

The ARMA(p,q) is represented as:

$$x[t] = \sum_{i=1}^{p} a_i x[t-i] + \sum_{i=1}^{q} b_i \eta[t-i] + \eta[t] \tag{1}$$

The ARIMA(p,d,q) model applies a differencing d to the input signal $x[t]$. For instance, if $d = 1$, then $y[t] = x[t] - x[t-1]$ is used as the input signal instead of $x[t]$. An ARIMA model is therefore better for non-stationary data because applying a differencing transformation can help stabilize the mean by removing changes in the level of a time series, thereby eliminating trend and seasonality.

To check if the logarithmic price time series, the augmented Dickey-Fuller (ADF) test is applied. ADF tests for the null hypothesis that a unit-root is present in a time series sample. If a unit root is present in a time series sample, it is non-stationary. The more negative the ADF test statistic, the stronger the rejection of the hypothesis that a unit root is present, at some confidence level. The p-value of this test is obtained by:

`adfuller(snp_close)[1]`

A p-value of 0.6676 is obtained. Since $p > 0.05$, the null hypothesis cannot be rejected and the S&P 500 logarithmic time series has a unit root present, and is therefore non-stationary. Therefore, an ARIMA model is more appropriate for the S&P 500 price time series.

2. An AR(1) model is fitted to the closing logged price data. The result is shown in the figure below:

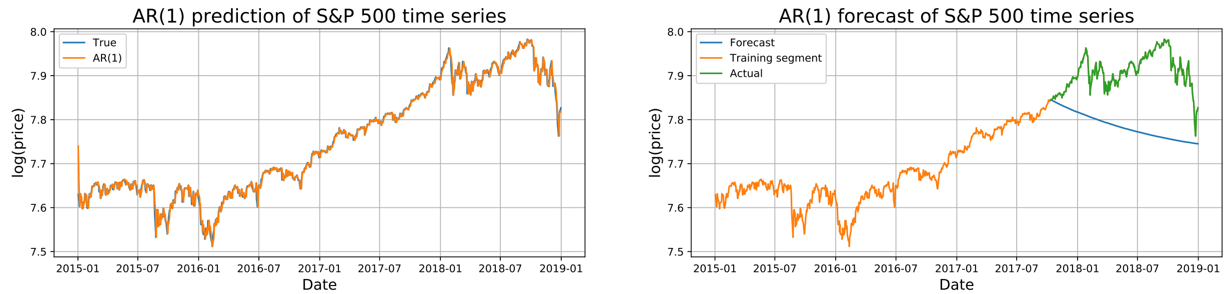


Figure 8: AR(1) prediction and forecast of S&P 500 logged price time series

The fitted AR(1) model has a constant parameter of 7.74 and $a_1 = 0.9974$. The ACF plot below shows the suitability of the AR(1) model as the time series is highly correlated to its previous data sample. In theory, higher order AR models can be used to fit the S&P 500 time series, but the first time lag component contains sufficient information needed to fit the S&P 500 series, and therefore models beyond AR(1) may not give significant improvements.

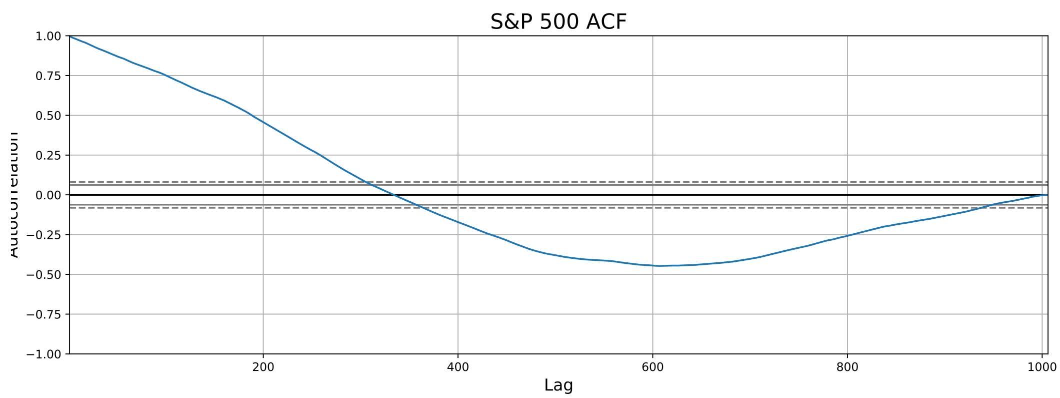


Figure 9: ACF plot

We see that the AR(1) prediction overlaps the true signal with an extremely small mean squared error of 8.62717e-05. The right subplot of figure 8 shows the forecasted data using 70% of the data. The AR(1) model only takes into account the previous day's return to generate a prediction, and therefore it generates a forecast that does not capture the trend at all.

3. Next, an ARIMA(1,1,0) model is fitted to the closing logged price data.

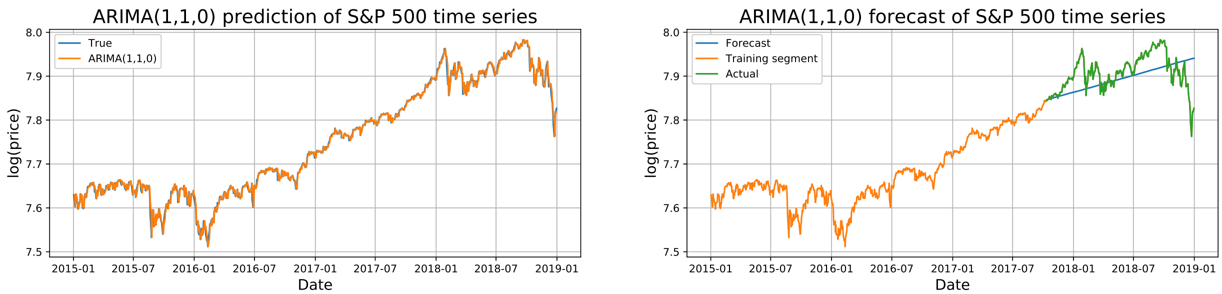


Figure 10: ARIMA(1,1,0) prediction and forecast of S&P 500 logged price time series

The fitted ARIMA(1,1,0) model has a constant parameter of 0.000196 and $a_1 = -0.008752$. Similar to the AR(1) model, the prediction overlaps the true signal but with a smaller mean squared error of 7.42831e-05. Compared with AR(1), the ARIMA forecast appears to be more accurate in the short term, correctly forecasting the trend. This is because the differencing parameter removes the stationarity in the data (this is explained in more depth in the next section), and therefore forecasts have informational value.

Although both ARMA and ARIMA give incorrect forecast predictions, ARIMA is better for capturing short term trends due to the incorporation of a differencing parameter.

4. As mentioned in section 1.1.4, there are many advantages when taking the logarithm of returns, including additivity, normality and stationarity. Taking the log of prices is necessary for the ARIMA model because the differencing process essentially transforms the logged prices into log returns, making the ARIMA predictions much more stationary, and therefore giving more accurate model parameters. The sample statistics are useful as descriptors of future behavior only if the series is stationary. If the series is increasing over time, the sample mean and variance will be growing with sample size, and they will underestimate the actual statistics in future periods. Hence, inaccuracies will occur if regression models are fitted to non-stationary data.

1.3 Vector Autoregressive (VAR) Models

1. The VAR(p) process is given by:

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

In expanded matrix form this is,

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{k,t} \end{bmatrix} = \begin{bmatrix} a^1_{1,1} & a^1_{1,2} & \cdots & a^1_{1,k} \\ a^{1,}_{2,1} & a^1_{2,2} & \cdots & a^1_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ a^1_{k,1} & a^1_{k,2} & \cdots & a^1_{k,k} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \vdots \\ y_{k,t-1} \end{bmatrix} + \cdots + \begin{bmatrix} a^p_{1,1} & a^p_{1,2} & \cdots & a^p_{1,k} \\ a^p_{2,1} & a^p_{2,2} & \cdots & a^p_{2,k} \\ \vdots & \ddots & \ddots & \vdots \\ a^p_{k,1} & a^p_{k,2} & \cdots & a^p_{k,k} \end{bmatrix} \begin{bmatrix} y_{1,t-p} \\ y_{2,t-p} \\ \vdots \\ y_{k,t-p} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \\ \vdots \\ e_{k,t} \end{bmatrix} \tag{3}$$

where each $y_{k,t}$ represents an interdependent time series. In contrast, the regular autoregressive model only consists of a single time series, regressed upon its historic values. Each individual time series is written out as:

$$\begin{gathered} y_{1,t} = c_1 + a^1_{1,1}y_{1,t-1} + a^1_{1,2}y_{2,t-1} + \cdots + a^1_{1,k}y_{k,t-1} + ... + a^p_{1,1}y_{1,t-p} + a^p_{1,2}y_{2,t-p} + \cdots + a^p_{1,k}y_{k,t-p} + e_{1,t} \\ y_{2,t} = c_2 + a^1_{2,1}y_{1,t-1} + a^1_{2,2}y_{2,t-1} + \cdots + a^1_{2,k}y_{k,t-1} + ... + a^p_{2,1}y_{1,t-p} + a^p_{2,2}y_{2,t-p} + \cdots + a^p_{2,k}y_{k,t-p} + e_{2,t} \\ \vdots \\ y_{k,t} = c_k + a^1_{k,1}y_{1,t-1} + a^1_{k,2}y_{2,t-1} + \cdots + a^1_{k,k}y_{k,t-1} + ... + a^p_{k,1}y_{1,t-p} + a^p_{k,2}y_{2,t-p} + \cdots + a^p_{k,k}y_{k,t-p} + e_{k,t} \end{gathered}$$

Here, $\mathbf{Y}$ is a $K \times T$ matrix corresponding of K time series and T samples:

$$\mathbf{Y} = \begin{bmatrix} y_{1,t} & y_{1,t-1} & \cdots & y_{1,t-(T-1)} \\ y_{2,t} & y_{2,t-1} & \cdots & y_{2,t-(T-1)} \\ y_{3,t} & y_{3,t-1} & \cdots & y_{3,t-(T-1)} \\ \vdots & \vdots & \ddots & \vdots \\ y_{k,t} & y_{k,t-1} & \cdots & y_{k,t-(T-1)} \end{bmatrix} \tag{4}$$

Or equivalently:

$$\mathbf{Y} = [\mathbf{y}_t \quad \mathbf{y}_{t-1} \quad \cdots \quad \mathbf{y}_{t-p}] \tag{5}$$

Given $\mathbf{B} = [\mathbf{c} \quad \mathbf{A}_1 \quad \mathbf{A}_2 \quad \ldots \quad \mathbf{A}_p]$, $\mathbf{y}_t$ can be written as:

$$\mathbf{y}_t = [\mathbf{c} \quad \mathbf{A}_1 \quad \mathbf{A}_2 \quad \ldots \quad \mathbf{A}_p] \begin{bmatrix} 1 \\ \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-p} \end{bmatrix} \tag{6}$$

Similarly, $\mathbf{y}_{t-1}$ is:

$$\mathbf{y}_{t-1} = [\mathbf{c} \quad \mathbf{A}_1 \quad \mathbf{A}_2 \quad \ldots \quad \mathbf{A}_p] \begin{bmatrix} 1 \\ \mathbf{y}_{t-2} \\ \mathbf{y}_{t-3} \\ \vdots \\ \mathbf{y}_{(t-1)-p} \end{bmatrix} \tag{7}$$

From this pattern, $\mathbf{Y} = \mathbf{BZ} + \mathbf{U}$ can be inferred as:

$$\mathbf{Y} = \mathbf{BZ} + \mathbf{U} = [\mathbf{c} \quad \mathbf{A}_1 \quad \mathbf{A}_2 \quad \ldots \quad \mathbf{A}_p] \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \mathbf{y}_{t-1} & \mathbf{y}_{t-2} & \cdots & \mathbf{y}_{t-T} \\ \mathbf{y}_{t-2} & \mathbf{y}_{t-3} & \cdots & \mathbf{y}_{(t-1)-T} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{t-p} & \mathbf{y}_{(t-1)-p} & \cdots & \mathbf{y}_{(t-T)-p} \end{bmatrix}}_{\mathbf{Z}} + \left[\mathbf{e}_t \quad \mathbf{e}_{t-1} \quad \cdots \quad \mathbf{e}_{(t-p)-(T-1)}\right] \tag{8}$$

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

2. The optimal coefficients $\mathbf{B}_{opt}$ are obtained by setting $\mathbf{U} = 0$, giving:

$$\mathbf{Y} = \mathbf{B}_{opt}\mathbf{Z} \tag{9}$$

To make $\mathbf{Z}$ invertible, we have to multiply it with its transpose $\mathbf{Z}^T$. In doing so, both sides of equation 9 have to be multiplied by $\mathbf{Z}^T$:

$$\mathbf{YZ}^T = \mathbf{B}_{opt}\mathbf{ZZ}^T \tag{10}$$

Taking the inverse of $\mathbf{ZZ}^T$ and multiplying it to both sides gives the optimal coefficients $\mathbf{B}_{opt}$:

$$\begin{aligned}\mathbf{B}_{opt}\mathbf{ZZ}^T(\mathbf{ZZ}^T)^{-1} &= \mathbf{YZ}^T(\mathbf{ZZ}^T)^{-1} \\ \mathbf{B}_{opt} &= \mathbf{YZ}^T(\mathbf{ZZ}^T)^{-1}\end{aligned} \tag{11}$$

3. A VAR(1) process is written as:

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \mathbf{e}_t \tag{12}$$

The previous time instant:

$$\mathbf{y}_{t-1} = \mathbf{A}\mathbf{y}_{t-2} + \mathbf{e}_{t-1} \tag{13}$$

Combining equations 12 and 13 together:

$$\mathbf{y}_t = \mathbf{A}^2\mathbf{y}_{t-2} + \mathbf{A}\mathbf{e}_{t-1} + \mathbf{e}_t \tag{14}$$

A general trend emerges from this substitution. Eigendecomposition can be performed on the matrix $\mathbf{A}$, where $\mathbf{Q}$ and $\mathbf{\Lambda}$ represent the eigenvector matrix and diagonal eigenvalue matrix respectively.

$$\mathbf{y}_t = \mathbf{Q}\mathbf{\Lambda}^k\mathbf{Q}^{-1}\mathbf{y}_{t-k} + \sum_{i=0}^{k}\mathbf{A}^i\mathbf{e}_{t-i} \tag{15}$$

The eigenvalue matrix has the following form:

$$\mathbf{\Lambda}^k = \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k^k \end{bmatrix} \tag{16}$$

From this, we observe that if $k \to \infty$, the eigenvalues of $\mathbf{A}$ become unstable and diverge for $|\lambda_i^k| > 1$. Therefore, for stability we require all eigenvalues $|\lambda_i^k|$ to be less than 1.

4. A VAR(1) model was fitted on the MA(66) detrended time series of 5 stocks: CAG, MAR, LIN, HCP and MAT.

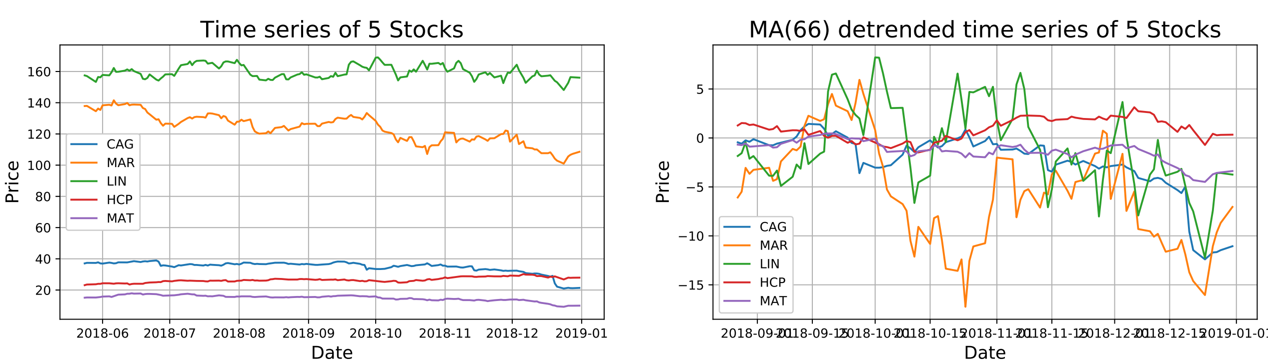


Figure 11: CAG,MAR,LIN,HCP,MAT time series

The eigenvalues of the regression matrix are:

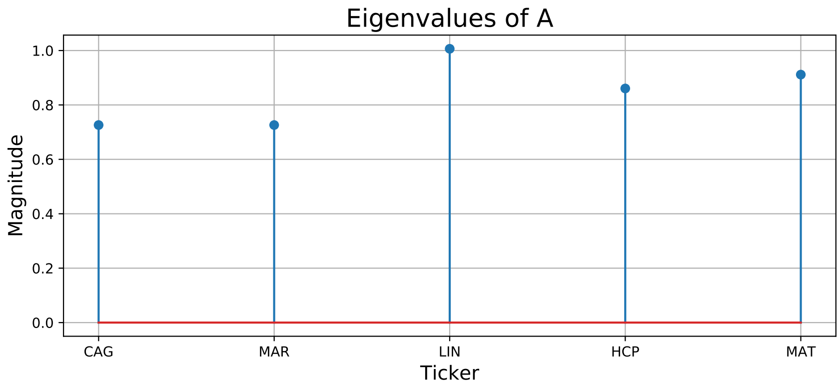


Figure 12: Eigenvalue stem plot

Ticker	Absolute Eigenvalue
CAG	0.726094
MAR	0.726094
LIN	1.006360
HCP	0.860519
MAT	0.911445

Table 2: Absolute eigenvalues of the regression matrix for CAG, MAR, LIN, HCP and MAT

Of all the stocks in table 2, LIN has an absolute eigenvalue of 1.006360, which suggests that the LIN time series is unstable and therefore should not be used to generate a portfolio.

The following heatmap shows the price and log returns correlations between the 5 stocks. The darker the color, the more positive the correlation between the stocks listed on the x and y axis. The color scale next to each correlation plot shows the correlation magnitude.

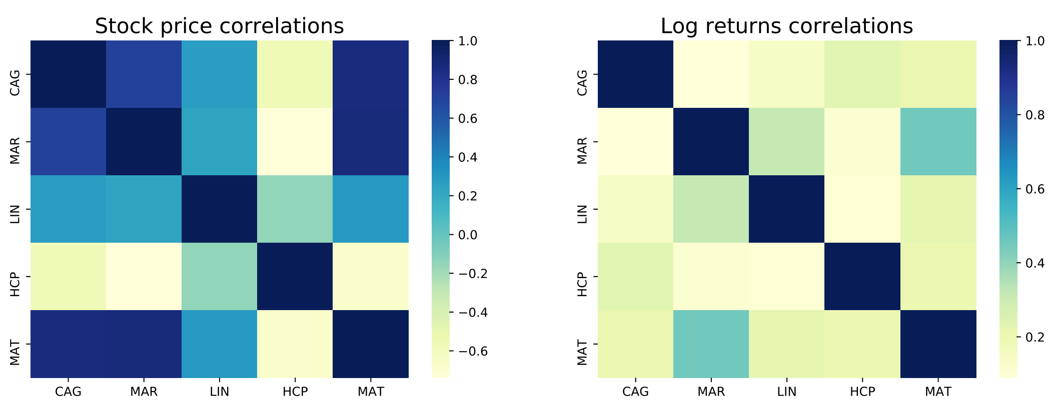


Figure 13: Stock price and log returns correlations between CAG, MAR, LIN, HCP and MAT

When looking at the stock price correlations, we see that MAT, CAG and MAR are extremely correlated to each other. This is likely due to the fact that these 3 stocks are grouped together in the consumer discretionary sector, and therefore it is expected that these stocks are affected by the same global macro factors, hence move in the same direction as each other. HCP appears to be negatively correlated with every other stock, and LIN exhibits little to no correlation with the other stocks, as indicated by the color magnitude. According to these observations, a possible portfolio could invest 40% into of MAT, CAG and MAR, 40% into HCP and the remaining 20% in LIN. This exploits the fact that HCP has a negative correlation with the other stocks, and the remaining 20% in LIN is relatively unaffected by movements in the 4 stocks. However, the eigenvalue analysis from the VAR(1) model suggests that LIN price series causes instabilities and is unsuitable to use in a portfolio. Taking this into account, a revised portfolio may invest 50% in MAT, CAG and MAR, and 50% in HCP.

5. The time series of the real estate stocks are plotted in figure 14. There does not appear to be an obvious joint directional movement of the stocks, contrary to the belief that same sector stocks should move in the same direction.

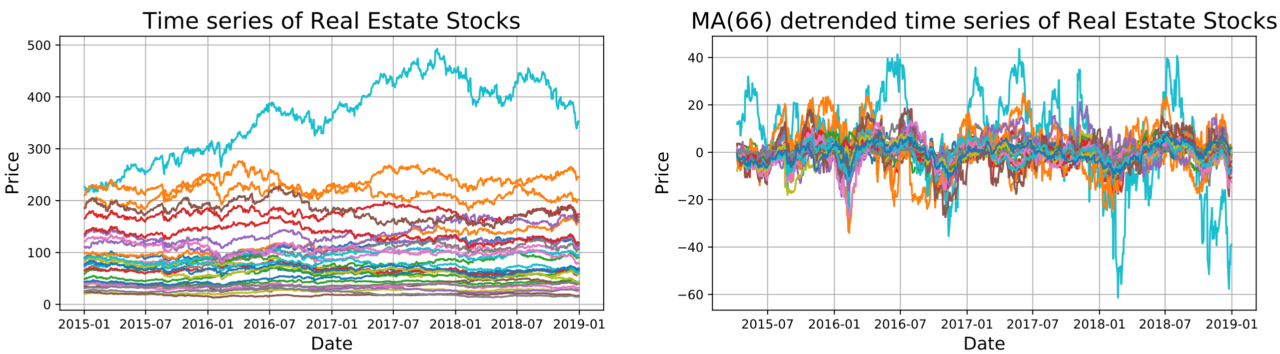


Figure 14: Time series and detrended time series of real estate stocks

By examining the correlations between the real estate stocks (Figure 15), it can be observed that the stocks exhibit a range of correlation strengths. This means that creating a portfolio consisting entirely of real estate stocks is a fairly diversified investment, due to the mix of price correlations observed.

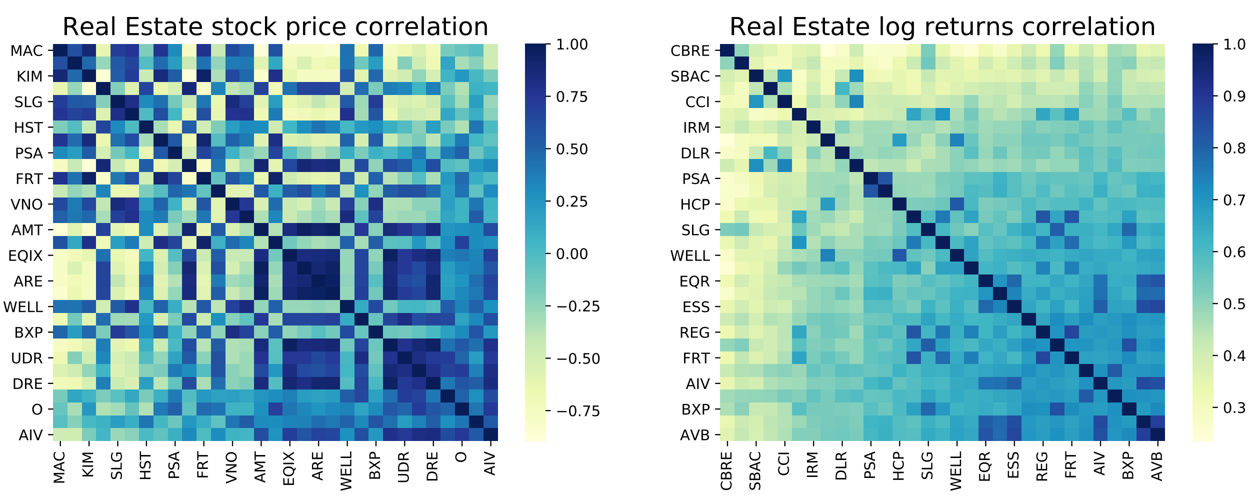


Figure 15: Stock price and log returns correlations between stocks in the real estate sector

In contrast, energy stocks are much more correlated to each other than real estate stocks (Figure 16). In this case, investing in the energy sector could be a good idea because its movements will likely depend on similar news sources (oil inventories, global energy demand etc) and therefore portfolios exploiting directional betting strategies could work well. However due to the unpredictability of market movements, it is not advised to invest in a single sector as random shocks will lead to large portfolio losses.

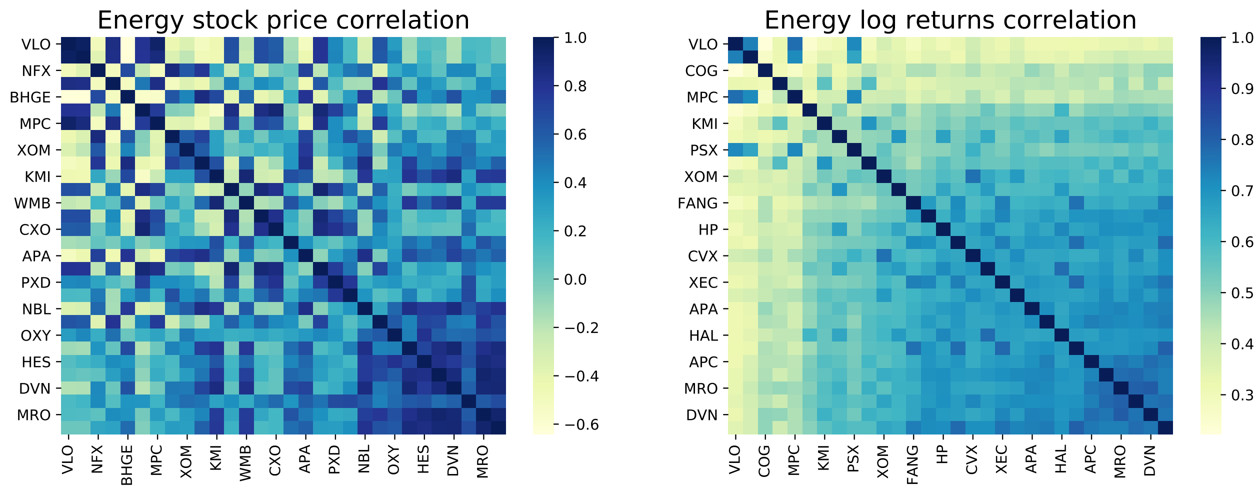


Figure 16: Stock price and log returns correlations between stocks in the energy sector

2 Bond Pricing

2.1 Examples of bond pricing

1. An investor receives USD 1100 in one year in return for an investment of USD 1000 now.

a) Annual compounding

$$\begin{aligned} 1000(1+r) &= 1100 \\ r &= 10\% \end{aligned} \tag{17}$$

b) Semi-annual compounding

$$\begin{aligned} 1000(1+\frac{r}{2})^2 &= 1100 \\ r &= 9.762\% \end{aligned} \tag{18}$$

c) Monthly compounding

$$\begin{aligned} 1000(1+\frac{r}{12})^{12} &= 1100 \\ r &= 9.569\% \end{aligned} \tag{19}$$

d) Continuous compounding

$$\begin{aligned} \lim_{N\to\infty} 1000(1+\frac{r}{N})^N &= 1100 \\ e^r &= 1.1 \\ r &= 9.531 \end{aligned} \tag{20}$$

2. What rate of interest with continuous compounding is equivalent to 15% per annum with monthly compounding?

$$\begin{aligned} e^r &= (1+\frac{15\%}{12})^{12} \\ r &= 14.907\% \end{aligned} \tag{21}$$

3. A deposit account pays 12% per annum with continuous compounding, but interest is actually paid quarterly. How much interest will be paid each quarter on a USD 10,000 deposit?

First calculate the equivalent interest rate with quarterly compounding:

$$\begin{aligned} e^{0.12} &= (1+\frac{R}{4})^4 \\ R &= 12.18\% \end{aligned} \tag{22}$$

The interest paid each quarter is therefore:

$$10000 * \frac{R}{4} = 10000 * \frac{12.18\%}{4} = 304.55 \tag{23}$$

2.2 Forward rates

a) Yes, assuming that the expected change in inflation in the second year is below 9%. Otherwise, the net value of the investment would be negative.

b) The 5% one year rate and 7% two year rate are based on today's spot rate curve, i.e. if you were to lend for one or two years today, you would expect to receive a 5% or 7% interest rate in return. The 9% strategy is where a forward rate of 9% is locked in today, for the lending rate between years 1 and 2, such that no arbitraging opportunities exist. Mathematically, the following equality is to be satisfied:

$$(1+s_1)(1+f_{1,2}) = (1+s_2)^2 \tag{24}$$

Therefore, the return of lending at 5% for one year today and at 9% the following year must equal to the return of lending at 7% for two years from today.

$$(1 + 5\%)(1 + 9\%) \equiv (1 + 7\%)^2 \tag{25}$$

c) The forward rate contract can be used to hedge interest rate risk by fixing a rate today. The lender is protected if the one year spot rate in the following year goes down, as they have secured a rate higher than the market rate.

However, if the one year spot rate in the following year is greater than the locked in forward rate today, as an investor you cannot benefit from this favorable fluctuation in interest rate.

d)

2.3 Duration of a coupon-bearing bond

a) Duration is calculated by taking the time weighted average of present values of the cash flows at the corresponding time period. The duration of the 1% bond is:

$$D = 0.0124 + 0.0236 + 0.0337 + 0.0428 + 0.0510 + 0.0583 + 6.5377 = 6.7595 \tag{26}$$

The duration is less than the bond maturity. This is expected as there are periodic coupon payments. For a zero coupon bond, the duration is equal to the maturity.

b) Modified duration is a measure of the sensitivity of the bond price to a change in interest rate. It is the derivative of the present value of a bond with respect to yield ($\frac{dPV}{dy}$). The simplified modified duration of the 1% bond is:

$$D_M = \frac{D}{1 + yield} = \frac{6.7595}{1.05} = 6.438 \tag{27}$$

The modified duration takes into account for changing yield to maturities (which is influenced by a change in interest rates). A first order taylor series expansion of $\frac{dPV}{dy}$ gives:

$$\Delta y \approx -D_M P \Delta y \tag{28}$$

Therefore, if yield increases or decreases by 1%, the price will decrease or increase by 6.438%.

c) Similar to bonds, pension liabilities have an inverse relationship to interest rates. An interest rate decrease will increase liabilities, and an interest rate increase will decrease liabilities. The amount of the increase or decrease can be estimated using the duration of the liabilities, to prevent unexpected changes in interest rates.

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

2.4.1 Market returns per day

The market returns for each day are plotted in figure 17 below. This is computed by taking the average returns of all 157 stocks per day.

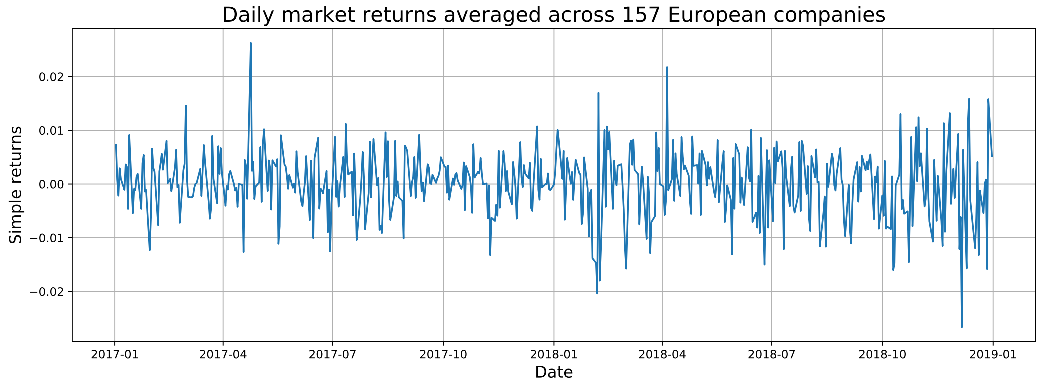


Figure 17: Market returns over time

2.4.2 Rolling Beta

The beta of a stock indicates how risky it is relative to the underlying market it is in. A high beta stock suggests that it is much more volatile than the market, i.e. its movements are β times the market movement. In contrast, a low beta stock ($0 < \beta < 1$) will reduce the overall risk of the portfolio. Beta is computed by:

$$\beta = \frac{Cov\left(R_i, R_m\right)}{var\left(R_m\right)} \tag{29}$$

where R_i corresponds to the returns of stock i, with a rolling window of 22 applied, and R_m is the overall market return. The following figure shows the rolling beta for all 157 companies:

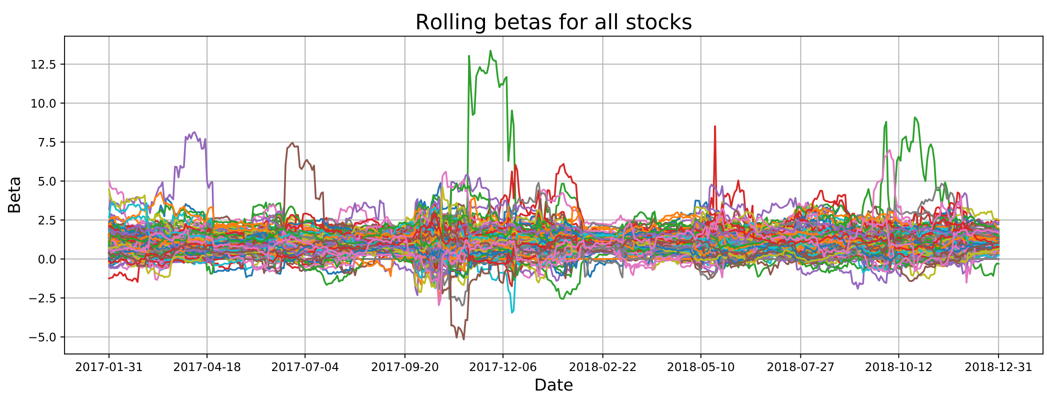


Figure 18: Rolling Beta

2.4.3 Cap-weighted market returns

The cap-weighted market return weights the stock returns with its corresponding market capitalization, with the weighting coefficient $\frac{mcap_i}{\sum_i \text{mcap}_i}$ for each stock i. This value is calculated for each company and summed together, which forms the relation:

$$R_m = \sum_i \frac{\text{mcap}_i \times \text{ret}_i}{\sum_i \text{mcap}_i} \tag{30}$$

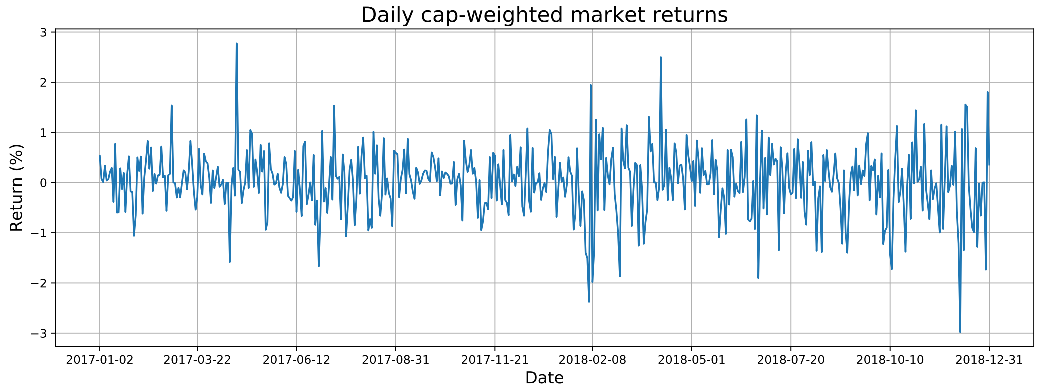


Figure 19: Cap-weighted market returns over time

2.4.4 Cap-weighted rolling Beta

The cap-weighted beta is plotted below. Instead of using daily market returns, the daily cap-weighted market returns are used instead.

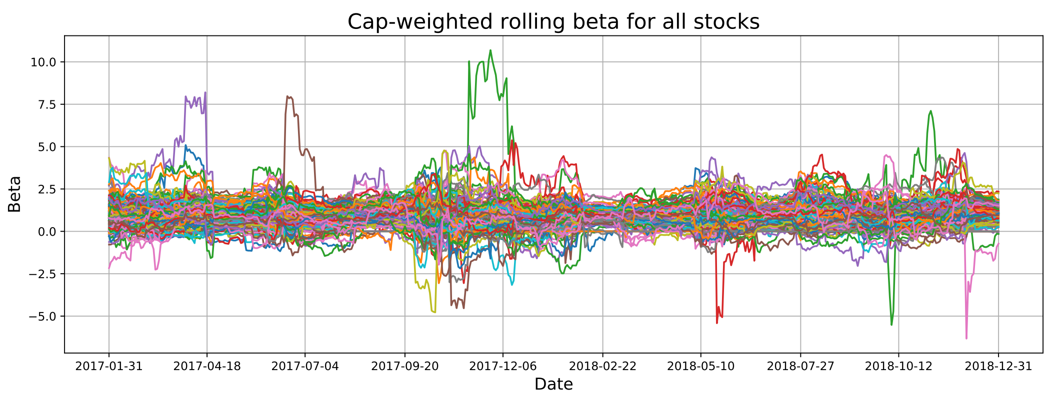


Figure 20: Cap-weighted rolling Beta

When taking the rolling beta and cap-weighted rolling betas of the largest-cap stock in the given data, the figure below is obtained. On average, the cap-weighted beta gives higher beta values, which suggests that its dominant capitalization has an even larger effect relative to the underlying market index, meaning that it is even riskier than if the regular beta were used.

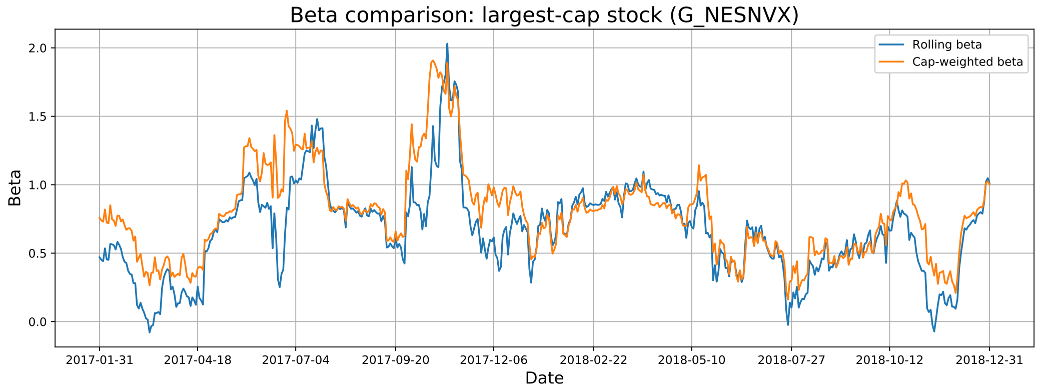


Figure 21: Rolling vs cap-weighted rolling beta for largest-cap stock

2.4.5 Arbitrage Pricing Theory (APT)

APT is a generalized version of the CAPM model, which incorporates more than a single factor to predict a stock's return. With APT, the expected stock return can be predicted using the linear relationship between a number of macroeconomic factors or indices, where sensitivity to changes in each factor is represented by a factor-specific beta coefficient. The APT regression predicts one a, R_m and R_s per day:

$$\mathbf{r}_i(t) = a + \mathbf{b}_{m_i}(t)R_m + \mathbf{b}_{s_i}(t)R_s + \epsilon_i \tag{31}$$

where $\mathbf{r}_i$, $\mathbf{b}_{m_i}$, $\mathbf{b}_{s_i} \in \mathbb{R}^{157\times 1}$. ϵ_i is the residual of the regression.

a) Figure 22 below shows the estimated values of a, Rm and Rs from linear regression.

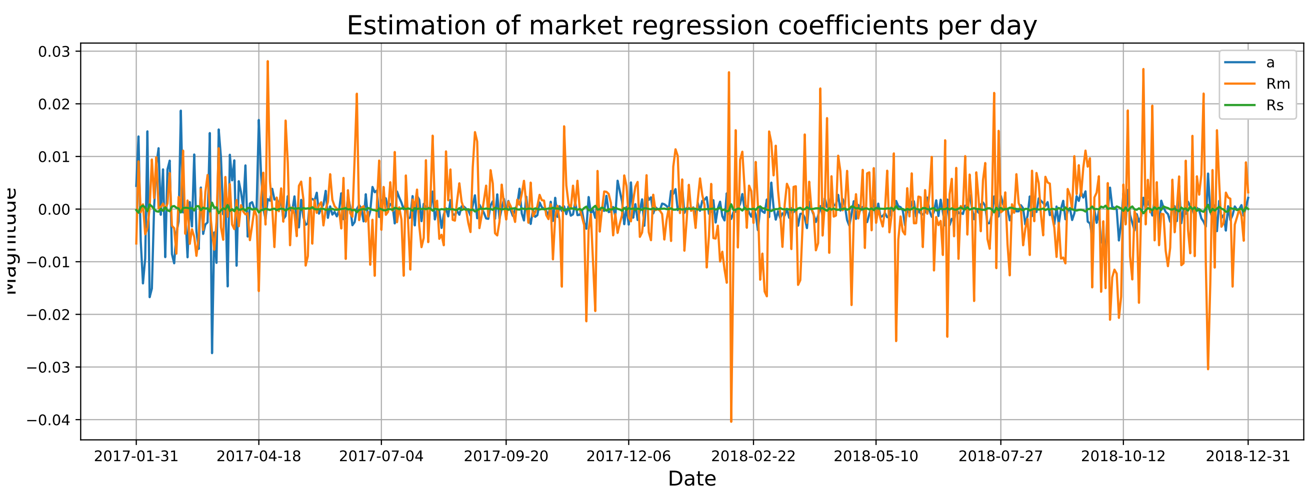


Figure 22: Estimation of a, Rs and Rm

b) The mean and standard deviation of a, R_m and R_s are shown below:

	mean	std
a	-0.000068	0.003663
Rm	-0.000306	0.007825
Rs	0.000017	0.000234

Figure 23: Mean and standard deviation of a, Rm and Rs

As shown from figure 22, the magnitude of the estimated Rm matches closely to that of the actual Rm in figure 17. The estimated standard deviation is relatively large, which could indicate that the market is relatively volatile during this period, and is a significant component to the predicted stock return.

Rs has a relatively low magnitude and standard deviation compared to a and Rm, which means that the company's size (market cap) is an insignificant factor in predicting each individual stock return.

The coefficient a represents the risk-free rate of return. As shown, during the initial time steps it has a relatively large value but it slowly converges to much lower values thereafter. In reality, T-Bond yields are used as a proxy to the risk-free rate and do not fluctuate as much as described in the plot. Regardless, it is less of a contributor to the stock return prediction than Rm and from this plot we can conclude that the overall daily market returns is the most significant component in predicting daily stock returns.

c) The correlation between a stocks simple return and specific return is plotted below. The specific return is defined as the return on an asset or investment over and above the expected return that cannot be explained by common factors. It is the return coming from the asset or investment's own merits, rather than the merits common to other correlated investments. The high correlation between the stocks specific return and daily returns (with the exception of four stocks which have a correlation below 0.6) indicate that the stocks daily return is highly attributed to the company's own performance factors.

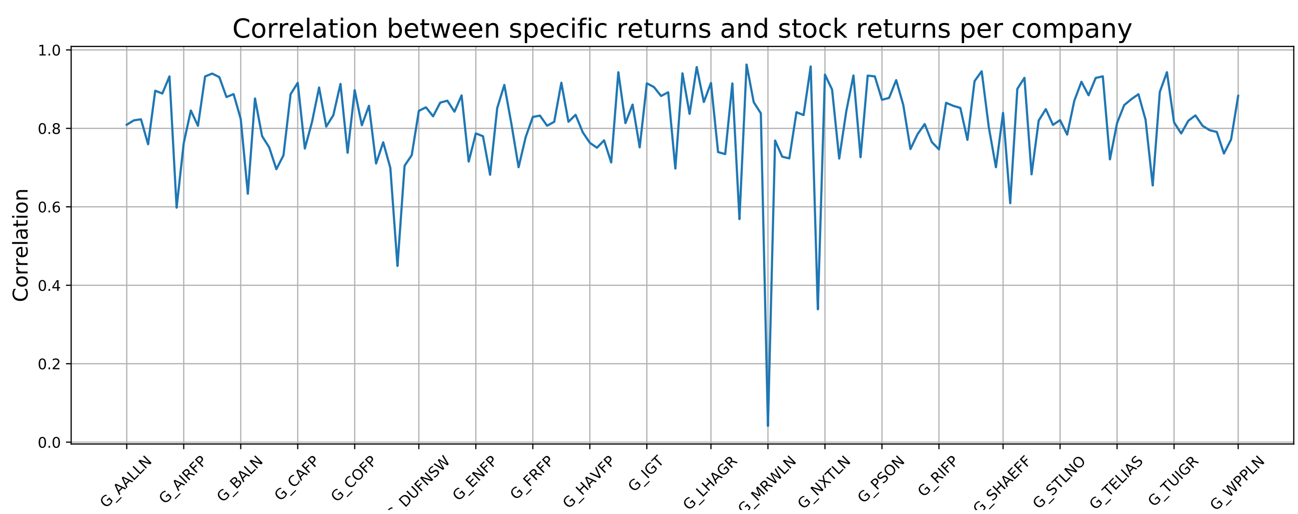


Figure 24: Correlation between simple return and specific return

d) A rolling window of 22 is applied to the matrix $\mathbf{R}$ before calculating the covariance matrix. For each window stride rolled, one covariance matrix is generated. The eigenvalues of each of these covariance matrices are calculated and plotted below. Throughout the measured duration, the eigenvalues are extremely small but positive, which suggests that $\mathbf{R}_m$ and $\mathbf{R}_s$ are unstable metrics.

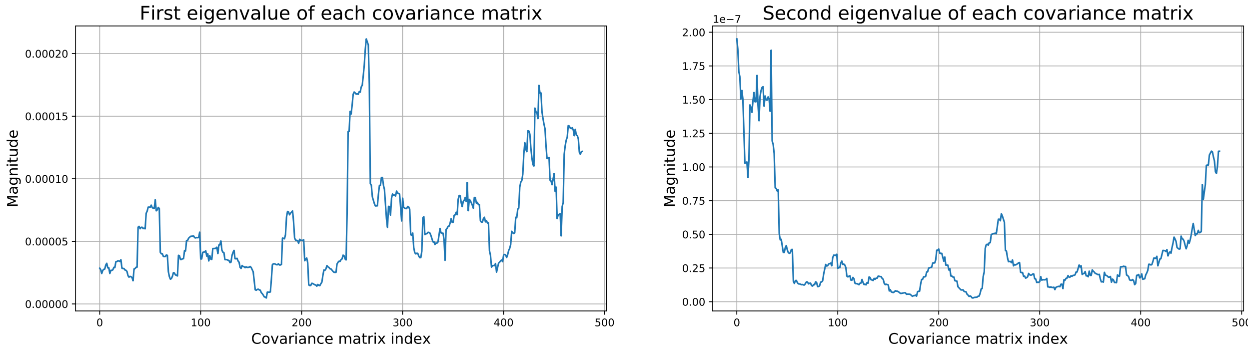


Figure 25: First and second eigenvalues of all calculated covariance matrices

e) The figure below shows the variance for each principal component in the specific return matrix. The first 7 principal components contain 53.33% of the total variance, suggesting that only 7 stocks are responsible for more than 50% of portfolio volatility. Ideally, the risk should be more spread out over all the significant principal components, such that the risk is much more spread out over all assets. After eigendecomposition of the covariance matrix, the eigenvalue matrix, by mathematical definition, contains the variance components of dataset feature space.

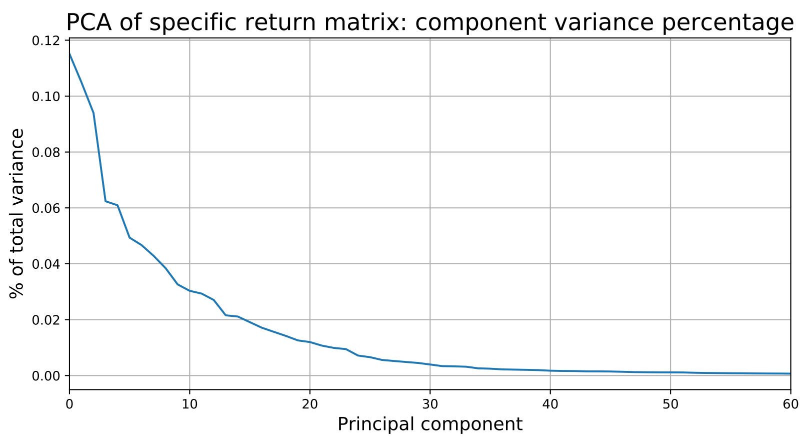


Figure 26: Specific returns matrix PCA

By selecting the number of principal components, n eigenvectors with the highest variance are selected, and the original dataset is projected onto the principal components to reduce the dimension of the dataset. Therefore, each principal component contains a percentage of the total variance.

3 Portfolio Optimization

3.1 Adaptive minimum-variance portfolio optimization

3.1.1 Optimal weights

A portfolio consists of n assets with a weight allocation matrix $\mathbf{w} = [w_1, \cdots, w_n]^T$. The return of the entire portfolio at any time instant t is therefore $\bar{r}[t] = \mathbf{w}^T\mathbf{r}[t]$. The objective of portfolio optimisation is to minimize the overall variance of the portfolio, to ensure investments are at minimal risk. The optimisation problem is as follows:

$$\begin{aligned} \min_{\mathbf{w}} \quad & J(\mathbf{w}, \mathbf{C}) = \frac{1}{2}\mathbf{w}^T\mathbf{C}\mathbf{w} \\ \text{s.t.} \quad & \mathbf{w}^T\mathbf{e} = 1 \end{aligned} \tag{32}$$

where $\mathbf{e} \in \mathbb{R}^{M\times 1}$.

The Lagrangian of the optimisation problem is:

$$\min_{\mathbf{w},\lambda} J(\mathbf{w}, \lambda, \mathbf{C}) = \frac{1}{2}\mathbf{w}^T\mathbf{C}\mathbf{w} - \lambda\left(\mathbf{w}^T\mathbf{e} - 1\right) \tag{33}$$

Differentiate the Lagrangian with respect to $\mathbf{w}$ and λ and set the derivative to zero:

$$\frac{\partial J}{\partial \mathbf{w}} = \mathbf{w}^T\mathbf{C} - \lambda\mathbf{e}^T = 0 \tag{34}$$

$$\frac{\partial J}{\partial \lambda} = \mathbf{w}^T\mathbf{e} - 1 = 0 \tag{35}$$

Solve equation 34 for $\mathbf{w}$:

$$\begin{aligned} \mathbf{w}^T\mathbf{C} = (\mathbf{C}\mathbf{w})^T &= \lambda\mathbf{e}^T \\ (\mathbf{C}\mathbf{w})^T &= \lambda\mathbf{e}^T \\ \mathbf{w} &= \lambda\mathbf{C}^{-1}\mathbf{e} \\ \mathbf{w}^T &= \lambda\mathbf{e}^T\mathbf{C}^{-1} \end{aligned} \tag{36}$$

Note that the covariance matrix is symmetric, and therefore $\mathbf{C} = \mathbf{C}^T$. Substitute equation 36 into equation 35:

$$\begin{aligned} \lambda\mathbf{e}^T\mathbf{C}^{-1}\mathbf{e} &= 1 \\ \lambda &= \frac{1}{\mathbf{e}^T\mathbf{C}^{-1}\mathbf{e}} \end{aligned} \tag{37}$$

Substituting this back into equation 36 give the optimal weights:

$$\mathbf{w}_{opt} = \frac{\mathbf{C}^{-1}\mathbf{e}}{\mathbf{e}^T\mathbf{C}^{-1}\mathbf{e}} \tag{38}$$

Using the optimal weights, the theoretical variance is calculated as:

$$\sigma^2_{opt} = \mathbf{w}^T_{opt}\mathbf{C}\mathbf{w}_{opt} = \lambda\mathbf{e}^T\mathbf{C}^{-1}\mathbf{C}\lambda\mathbf{C}^{-1}\mathbf{e} = \lambda^2\mathbf{e}^T\mathbf{C}^{-1}\mathbf{e} = \lambda = \frac{1}{\text{sum}\left(\mathbf{C}^{-1}\right)} \tag{39}$$

3.1.2 Minimum-variance portfolio

The following plot shows the cumulative return for the test set, using both the optimal weights and equal weights for each stock. The cumulative returns for the optimally weighted portfolio outperforms the equally weighted portfolio for the first half of the data (training data), generating an excess return of around 1-2%. However in the second half of the data (test set), the equally weighted portfolio outperforms the optimal weight portfolio, until the end of 2018 when the optimal weight portfolio makes a small gain despite a market correction.

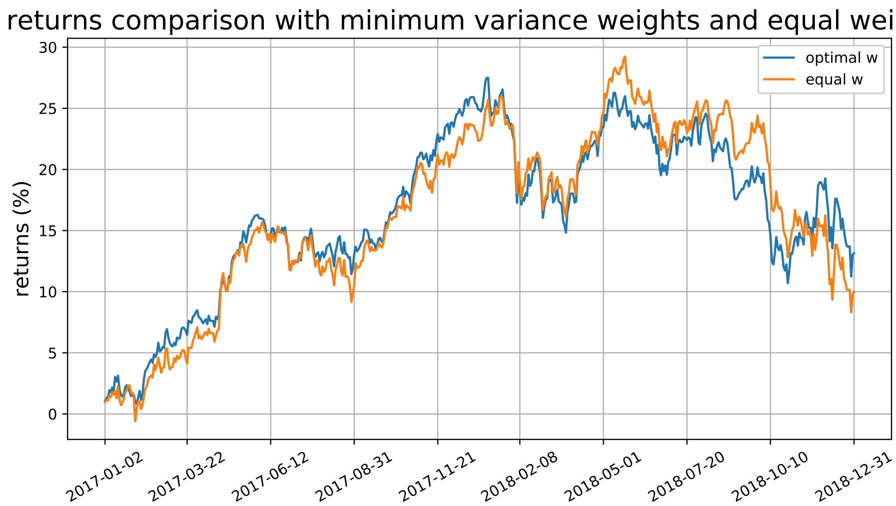


Figure 27: Cumulative returns with optimal weights and equal weights, test set

The actual variance for the optimally weighted portfolio returns is 5.560e-05, for the equally weighted portfolio returns it is 5.867e-05, whereas the theoretical variance calculated by equation 39 is 2.851e-05. The optimally-weighted portfolio has a variance closer to the theoretical variance, which indicates that it is ever so slightly less risky than the equally weighted portfolio. When considering the entire time horizon, the optimally weighted portfolio generates an extra 3%.

3.1.3 Adaptive minimum-variance portfolio

A minimum-variance portfolio strategy is implemented by making the portfolio weights adaptive.

4 different window lengths are used for comparison. For the adaptive weight regime, the past M days daily returns are used to compute the covariance matrix, which is then used to calculate the optimal weights for the current day. For the non-adaptive optimal weights and equal weight regimes, the covariance matrix is computed using the entire time horizon considered, and therefore the window length stated only applies to the calculation of the adaptive weights.

The plots below show the portfolio returns when 3 different weight regimes are used. As shown, the adaptive weighting portfolio outperforms the other two weighting regimes, which shows that the adaptive strategy is more resistant and robust to market corrections (towards the end of 2018). Additionally, a lower adaptive rolling window length outperforms a higher adaptive rolling window length, where the ending cumulative returns for $M = 15$ is roughly 0.65, whereas for $M = 50$ it is roughly 0.25, a significant reduction in portfolio gains.

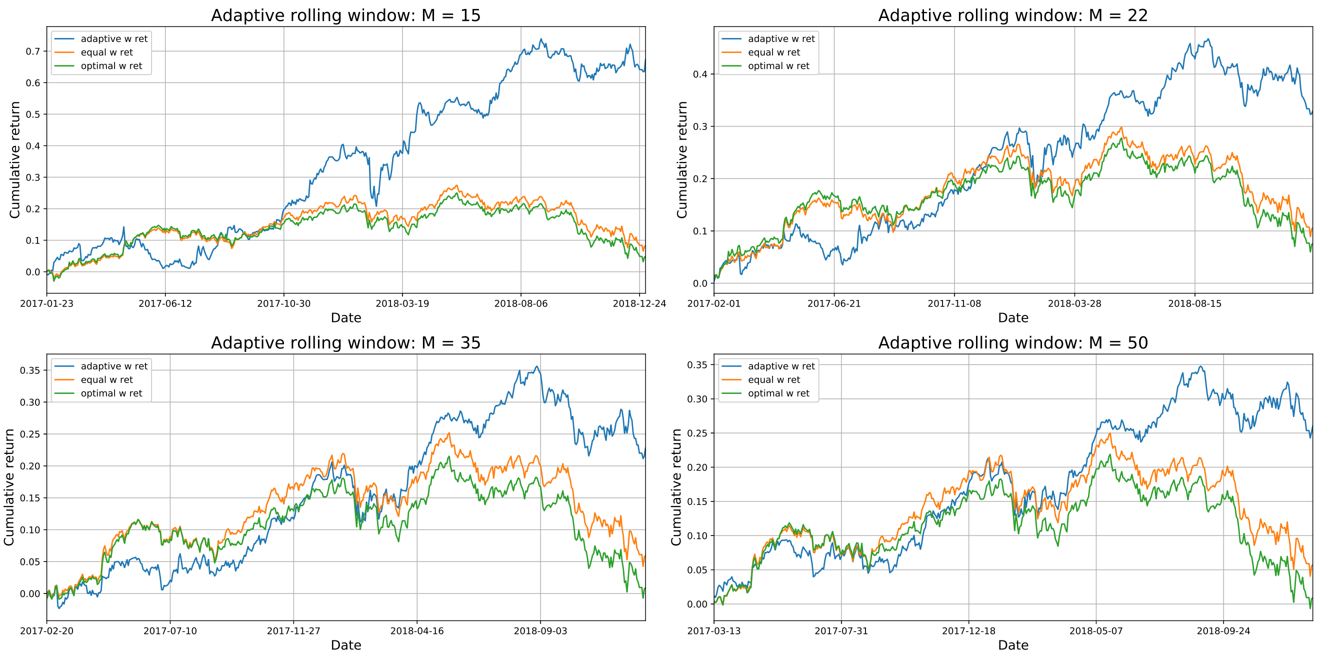


Figure 28: Cumulative returns comparison with adaptive weights, non-adaptive optimal weights and equal weights

The following figure shows the portfolio returns variance distribution for different window lengths. From the result it can be inferred that the portfolio returns become riskier as the window length used increases, as described by the rightward shift in mean and fatter tailed distributions. This may be because using a larger rolling window results in less data points, and therefore this loss of information may lead to inaccuracies.

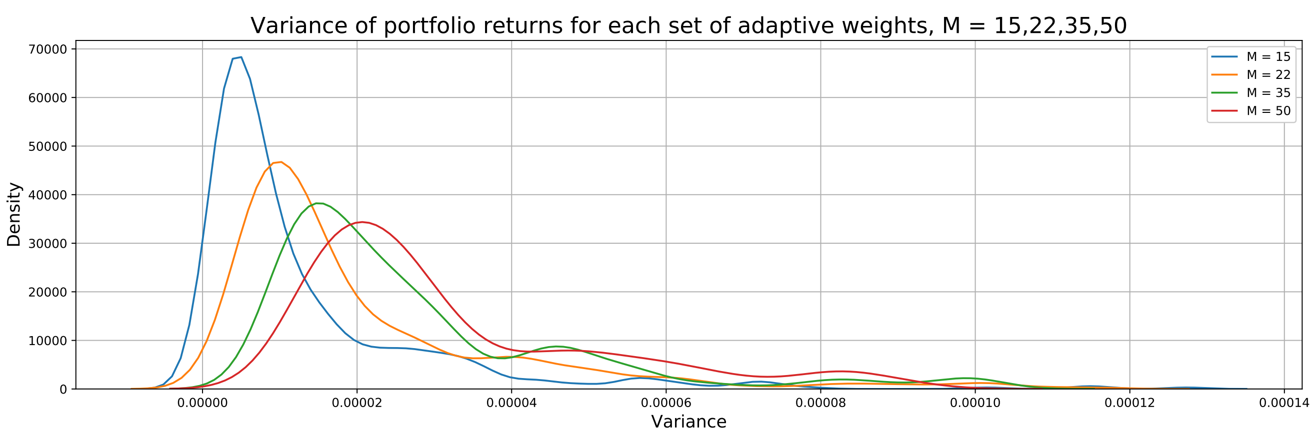


Figure 29: Variance of portfolio returns for each set of adaptive weights, different window lengths comparison

The log of portfolio returns variances are plotted below. For the adaptive weight regime, for each window length the average of all portfolio returns variances is calculated. The averaged variance appears to plateau beyond $M = 35$, indicating that window lengths beyond 35 are inefficient. It also appears that the variance could be even lower with a smaller M, but however for $M < 12$ the covariance matrix is singular and does not work. Perhaps the portfolios generated for $M < 12$ would have negative variance, hence the covariance matrix singularity. As expected, the portfolio returns variance for the other two weight regimes remain constant and are close to identical.

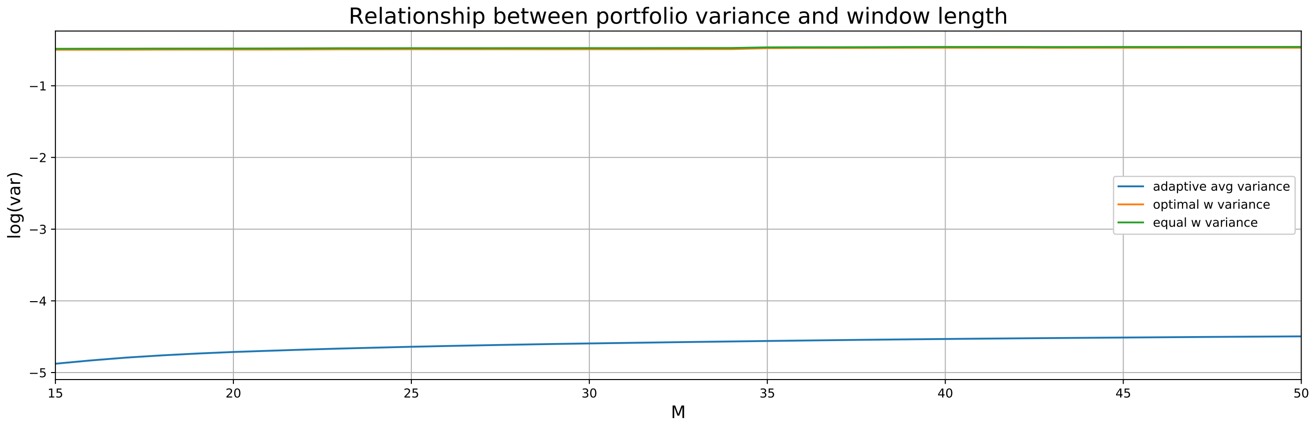


Figure 30: Portfolio returns variance comparison with different weighting regimes

4 Robust Statistics and Non Linear Methods

4.1 Data Import and Exploratory Data Analysis

1. The key descriptive statistics of each stock is summarized in the table below:

		aapl	ibm	jpm	djx
Open	mean	1.877448e+02	1.385335e+02	1.088017e+02	2.498985e+04
	std	1.896082e+02	1.395785e+02	1.097429e+02	2.521502e+04
	median	1.858598e+02	1.373998e+02	1.077656e+02	2.475778e+04
High	mean	1.877495e+02	1.384423e+02	1.086902e+02	2.498860e+04
	std	1.853015e+02	1.297227e+02	1.047819e+02	2.510840e+04
	median	3.268706e+07	5.209121e+06	1.467776e+07	3.596328e+09
Low	mean	2.229176e+01	1.225519e+01	5.385203e+00	8.596203e+02
	std	2.241331e+01	1.206441e+01	5.222890e+00	8.242440e+02
	median	2.214014e+01	1.235136e+01	5.466754e+00	8.944732e+02
Close	mean	2.229318e+01	1.217575e+01	5.328294e+00	8.603737e+02
	std	2.422847e+01	1.180154e+01	6.167677e+00	1.246267e+03
	median	1.426229e+07	3.340601e+06	5.342836e+06	6.444814e+08
Adj Close	mean	1.863500e+02	1.429650e+02	1.094900e+02	2.498165e+04
	std	1.877400e+02	1.441350e+02	1.107350e+02	2.518751e+04
	median	1.850450e+02	1.422550e+02	1.078150e+02	2.478033e+04
Vol	mean	1.863750e+02	1.430200e+02	1.091750e+02	2.501842e+04
	std	1.822775e+02	1.325396e+02	1.045752e+02	2.504856e+04
	median	2.891265e+07	4.238900e+06	1.368895e+07	3.507530e+09

Figure 31: Key descriptive statistics of AAPL, IBM, JPM and DJI

2. The following figures summarize the returns histogram and pdf of AAPL, IBM, JPM and DJI index. The one-day returns distribution is much closer to a Gaussian distribution than that of the adjusted close distribution. For all of the adjusted close plots, most of the distributions appear to be the sum of two distributions with different means and large variances. In contrast, the one-day returns distribution is centred around zero and the variances are very similar across all four assets. However, when looking closer at the returns pdf, the number of outliers is significantly higher than the adjusted close pdf.

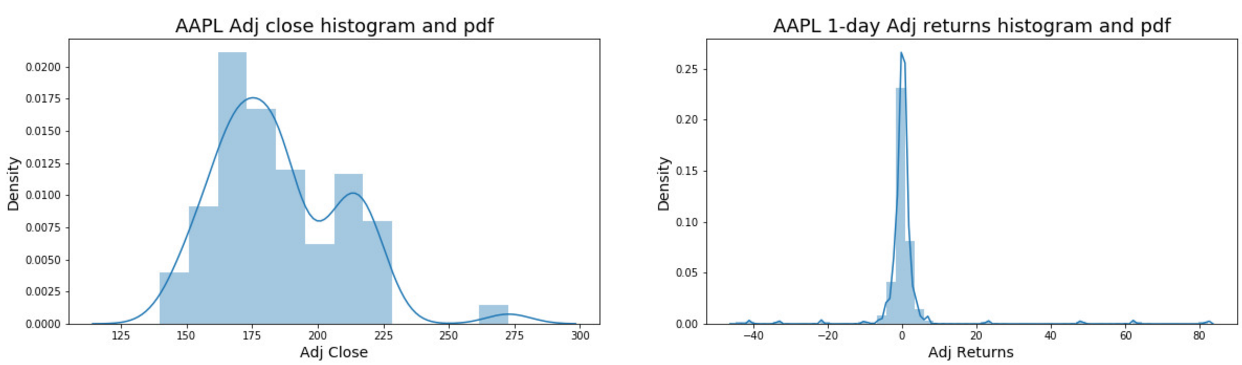


Figure 32: AAPL histogram and PDF for adj close and returns

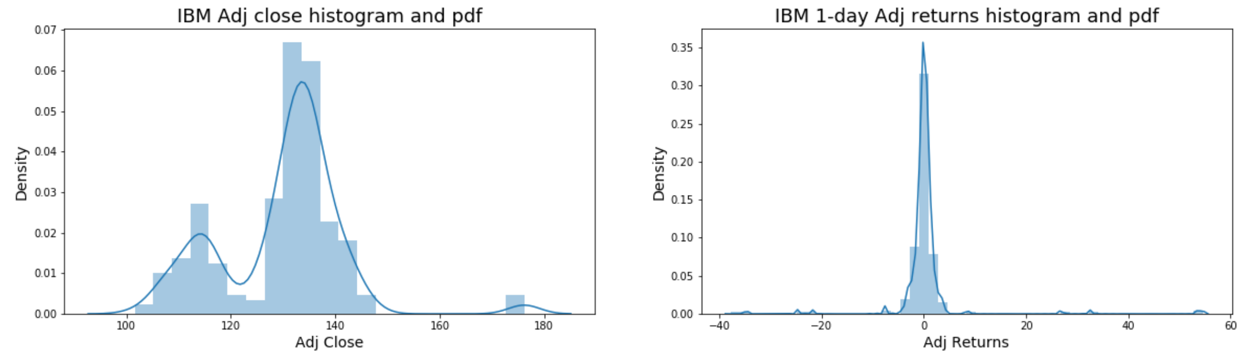


Figure 33: IBM histogram and PDF for adj close and returns

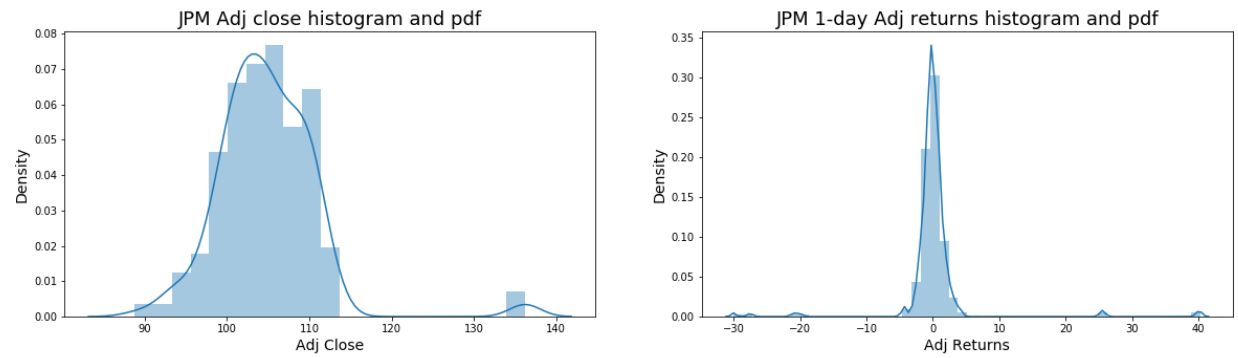


Figure 34: JPM histogram and PDF for adj close and returns

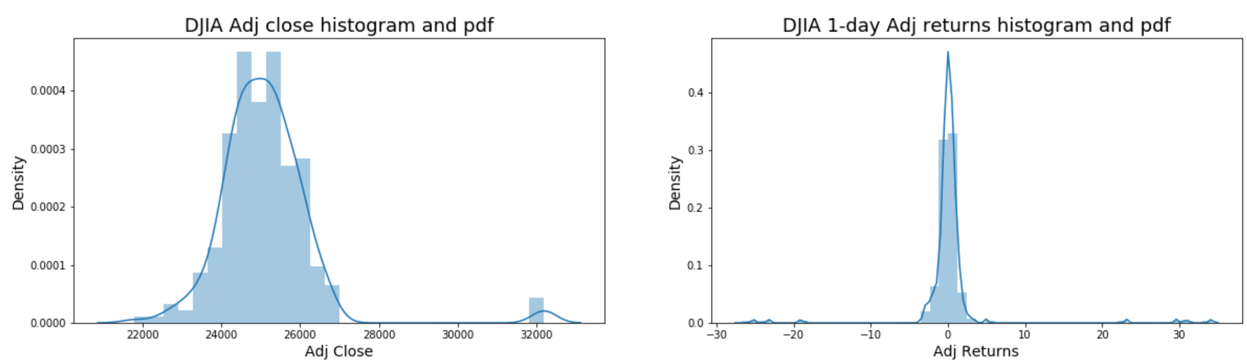


Figure 35: DJIA histogram and PDF for adj close and returns

3. The following figures show anomaly detection methods applied to each of the 4 stocks.

The plots on the left use a 5-day rolling mean along with a $\pm 1.5\times$ standard deviation relative to the rolling mean. The plots on the right use a 5-day rolling median along with a $\pm 1.5\times$ mean absolute deviation relative to the rolling median.

It turns out that when a constant of 1.5 is used, then according to the IQR method, any data which lies beyond 2.7 standard deviations from the mean, on either side, shall be considered as outlier. In common practice, anything beyond 3 standard deviations from either side of the mean is deemed an outlier, but it requires a constant of 1.7 to be achieved, which is less 'symmetrical'. Hence for this method of outlier detection, a constant of 1.5 is widely accepted.

The upper and lower bounds of the moving averages wrap around very closely to the adjusted close price series, whereas the MAD bound is a lot more distinct and noticeable. The median is a more robust estimator of central tendency as outliers will have negligible effects on the sample median, which will be observed in the next section. The rolling mean is much smoother than the rolling median as the moving average effectively acts as a low pass filter, removing high frequency components in the original adj close series.

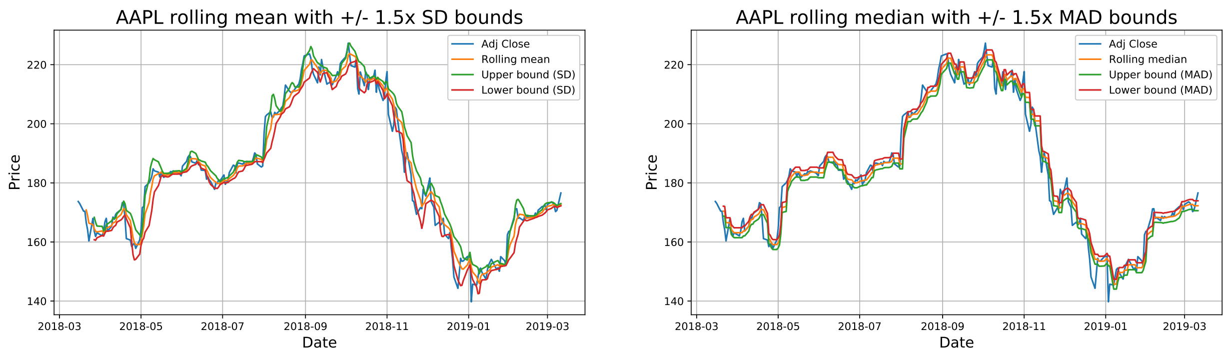


Figure 36: AAPL anomaly detection with moving average and median

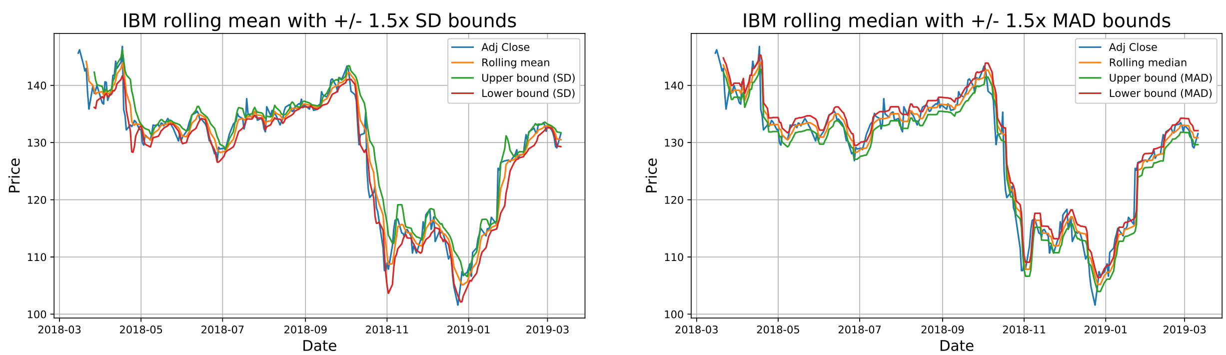


Figure 37: AAPL anomaly detection with moving average and median

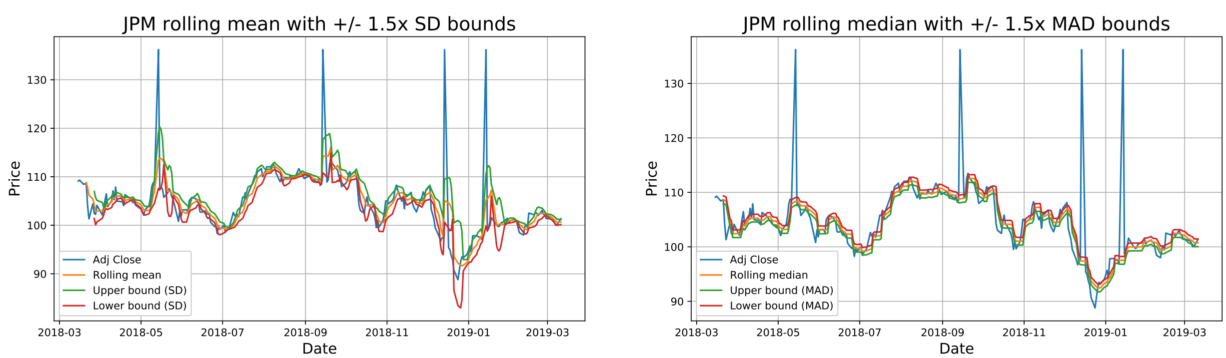


Figure 38: JPM anomaly detection with moving average and median

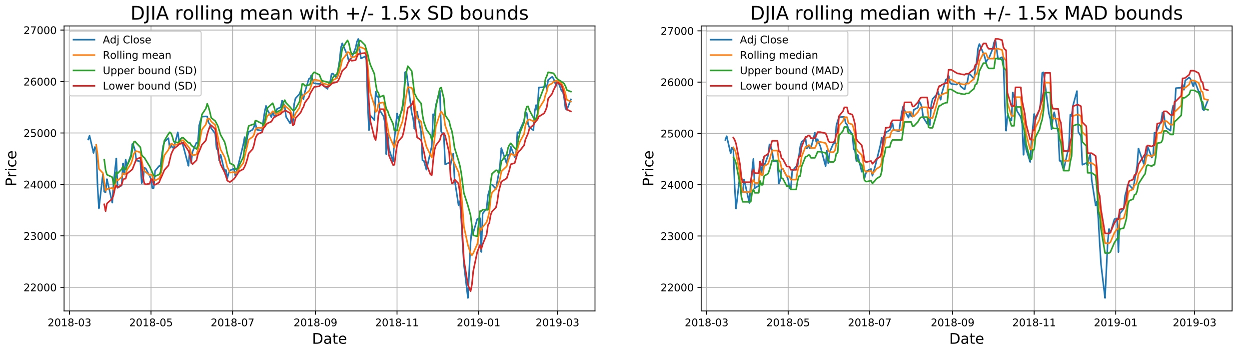


Figure 39: DJIA anomaly detection with moving average and median

4. The following figures show anomaly detection methods applied to each of the 4 stocks, with outlier points introduced at 4 specific dates. For the rolling mean plots on the left, it is observed that at the outliers, the magnitude of the rolling mean, upper bound and lower bound are all offset by the outlier due to the sudden increase in magnitude. On the other hand, the rolling median plots are completely unaffected and preserve the same shape as seen in the previous section, as if there no outliers at all in the data. Furthermore, robust statistics work best with non-Gaussian data (e.g. the adjusted close), and therefore in this case it is clear that the median is a much more robust statistic in being resistant to outliers.

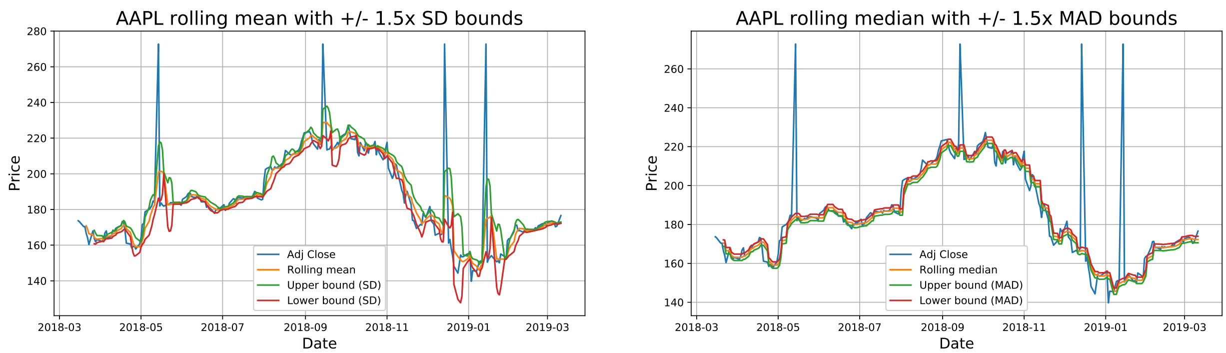


Figure 40: AAPL (with outliers) anomaly detection with moving average and median

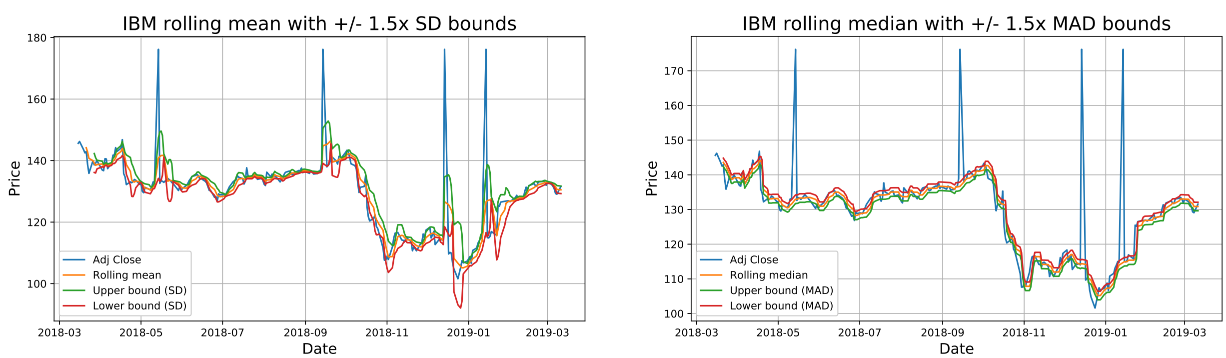


Figure 41: IBM (with outliers) anomaly detection with moving average and median

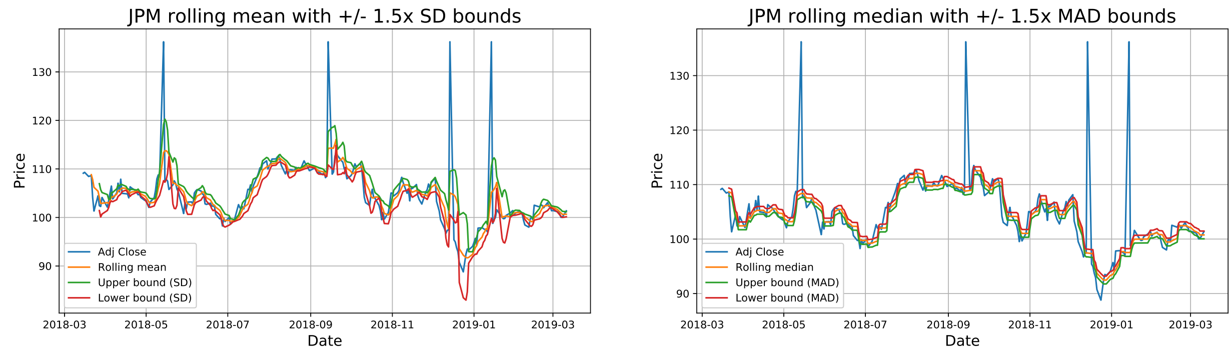


Figure 42: JPM (with outliers) anomaly detection with moving average and median

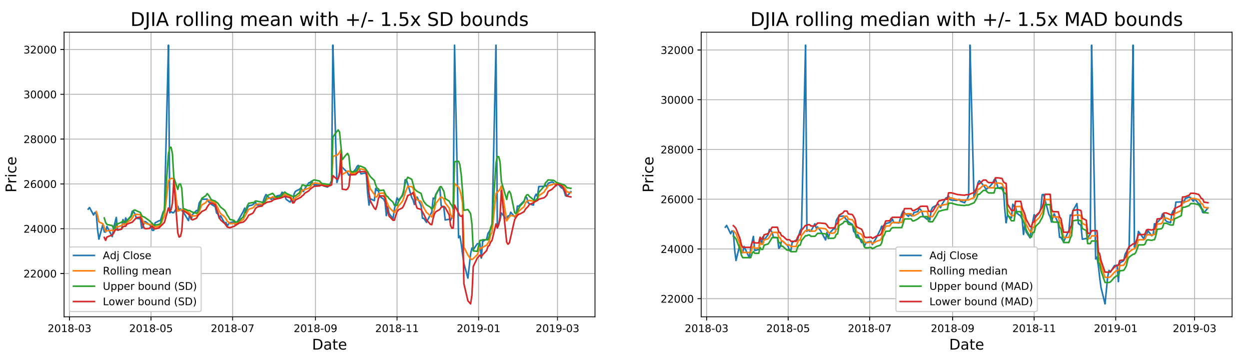


Figure 43: DJIA (with outliers) anomaly detection with moving average and median

5. A box plot is used to describe statistical information within data.

   1. The line splitting the box is the location of the median, which is equal to the 50^{th} (Q2) percentile.

   2. The skewness of the data is represented by the location of the median relative to the edges of the box, which represent the 25^{th} (Q1) and 75^{th} (Q2) percentile of the data respectively.

   3. The length of the box represents the interquartile range (IQR), which is where 50% of the data lies. It is therefore a measure of statistical dispersion.

   4. The whiskers at the end represent an upper and lower bound defined by Q1 - 1.5(IQR) and Q3 + 1.5(IQR), and any outliers beyond these bounds are shown as a diamond.

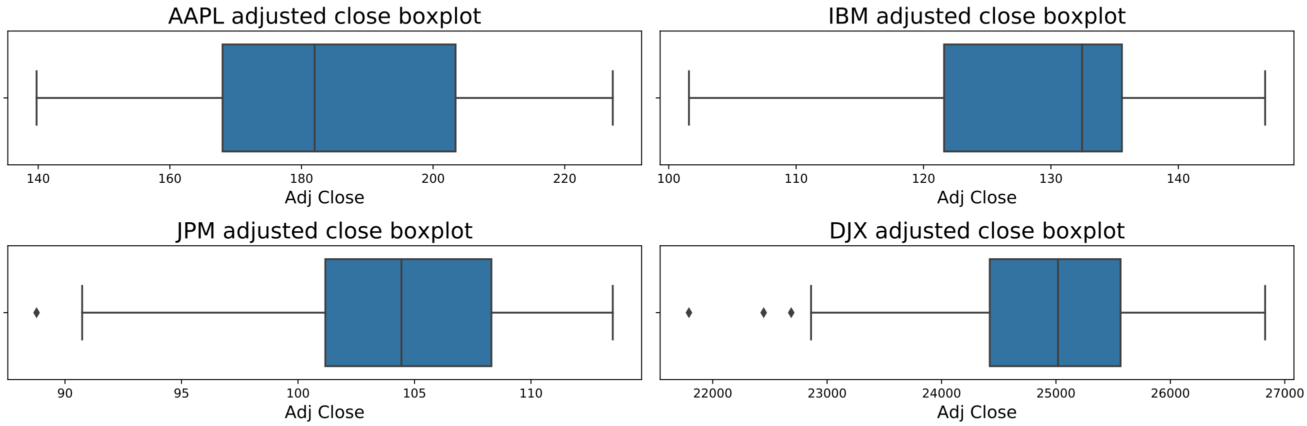


Figure 44: Box plot of 4 stocks

The box plots for AAPL and JPM are positively skewed and AAPL has a larger data variance as indicated by its longer box length. The box plot indicates that AAPL has no outliers, whereas JPM has one. In contrast, IBM data is negatively skewed with similar variability to that of JPM as dictated by the IQR distance. Lastly, the DJI box plot shows that it has no skew and follows a normal distribution. These findings do not provide any extra information than what the histogram and pdf already tell us from section 4.1.2.

4.2 Robust Estimators

1. The following functions are implemented to calculate the following robust estimators:

```python
import numpy as np
from scipy.stats import iqr

#robust location estimator: median
def median(x):
    return np.median(x)

#robust scale estimator: IQR
def IQR(x):
    return iqr(x)

#robust scale estimator: MAD
def MAD(x):
    return np.median( abs(x - np.median(x)) )
```

2. This section assesses the computational complexity of each of the robust estimators.

Robust location estimator (median):

- To find the median of an unsorted list, selection algorithms can be used. For example, the median of medians algorithm divides the list into sublists and finds each sublists median to obtain the unsorted lists median, and has a computational complexity of $\mathcal{O}(n)$ for all cases.

Robust scale estimator (IQR):

- To find the interquartile range of a list, the list has to be sorted and the 25th and 75th percentiles are extracted. The computational complexity is therefore determined by the complexity of the sorting algorithm used. The fastest sorting algorithms are quicksort or merge sort, which have complexities of $\mathcal{O}(n \log n)$. The worst-case runtime of quicksort with a complexity of $\mathcal{O}(n^2)$ is easily avoided by using an appropriate choice of the pivot.

Robust scale estimator (MAD):

- The median absolute deviation calculation requires an initial sorting of the unsorted list, finding the median, n subtractions from the median, n absolute value calculations, and lastly another median calculation. The computational complexity is dependent on the sorting algorithm used, and is therefore $\mathcal{O}(n \log n)$.

3. The breakdown point of an estimator is the proportion of incorrect observations (outliers) it can handle before giving an incorrect result. The higher the breakdown point, the more robust an estimator is. The maximum breakdown point is defined at 0.5, because if more than 50% of observations are incorrect, it is not possible to distinguish between the underlying and contaminated distribution.

Median

- If we have n data points and let a minority of them $\lfloor (n-1)/2 \rfloor$ to infinity leaving the rest fixed, then the median stays with the majority. The median changes, but does not become arbitrarily bad. The asymptotic breakdown point for the median is therefore $\frac{n-1}{2n} \approx 0.5$.

IQR

- The breakdown point of IQR is 0.25, because the maximum allowable number of outliers in either the top or bottom 25% of the data is 25%.

MAD

- The breakdown point of the median absolute deviation is limited by the breakdown point of the median. This is because the second median applied has no effect on the robustness. Therefore, the breakdown point of MAD is 0.5.

4.3 Robust and OLS regression

1. Linear and Huber regression are used to estimate the relationship between the stock returns and DJIA returns. The linear regression fit is computed by minimising the squared error of the residual. In contrast, the Huber regression fit is computed using the Huber M-estimator, which attempts to get the best of both the least-squares estimator as used in OLS, and the absolute deviation estimator. The advantage of using the least squares estimator is that it is easy to find the minimum, and the absolute deviation estimator is more robust to outliers. It attempts to minimise the following error function:

$$H(\varepsilon) = \begin{cases} \varepsilon^2/2 & \text{for } |\varepsilon| \leq k \\ k|\varepsilon| - k^2/2 & \text{for } |\varepsilon| > k \end{cases} \tag{40}$$

The residual is defined as ε. For residuals less than or equal to k, the least-squares estimator is used, and for residuals greater than k the absolute deviation estimator is used. The default value of k is 1.345σ, which gives the Huber M-estimator a 95% efficiency.

The corresponding R^2 score is computed for both linear regression and Huber regression, which is a statistic that summarizes how close the data are to the fitted line.

The plot below shows the regression without outliers present in the data.

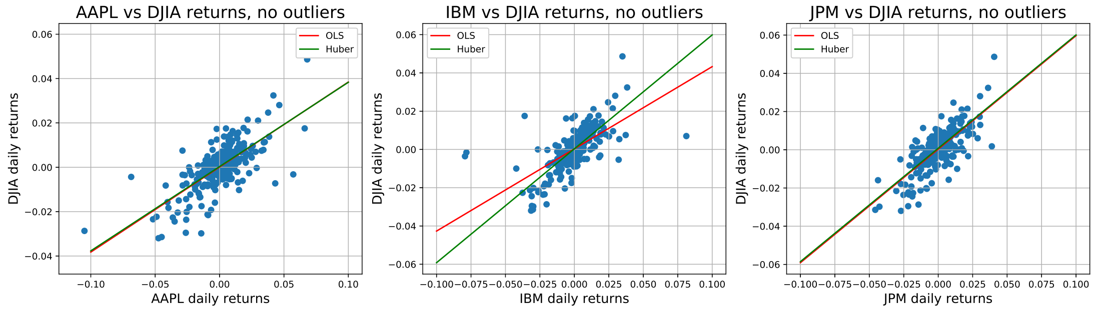


Figure 45: Linear and Huber regression between individual stock returns vs DJIA returns, no outliers

Stock	R^2 OLS	R^2 Huber
AAPL	0.5048	0.5043
IBM	0.4122	0.3511
JPM	0.5505	0.5484

Table 3: R^2 score: no outliers

2. The plot below shows the regression with the 4 outliers in section 4.1.4 present in the data.

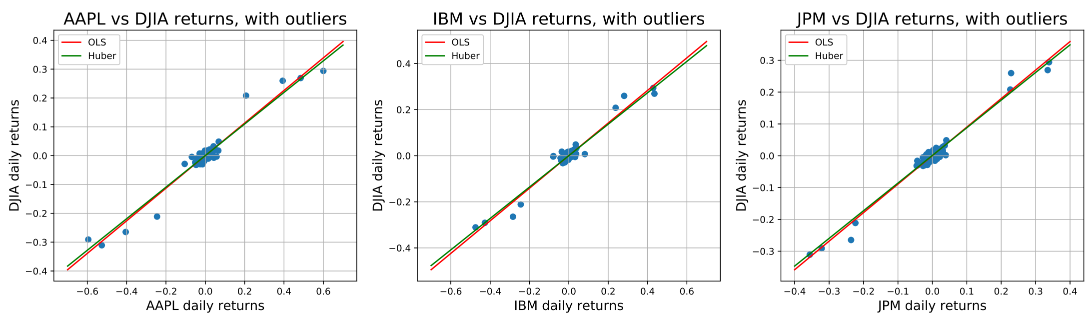


Figure 46: Linear and Huber regression between individual stock returns vs DJIA returns, with outliers

Stock	R^2 OLS	R^2 Huber
AAPL	0.9381	0.9372
IBM	0.9434	0.9422
JPM	0.9616	0.9605

Table 4: R^2 score: with outliers

To further test the effectiveness and robustness of the Huber regressor, 20% of the total dataset are made into outliers, all with a value of 1.2*max.

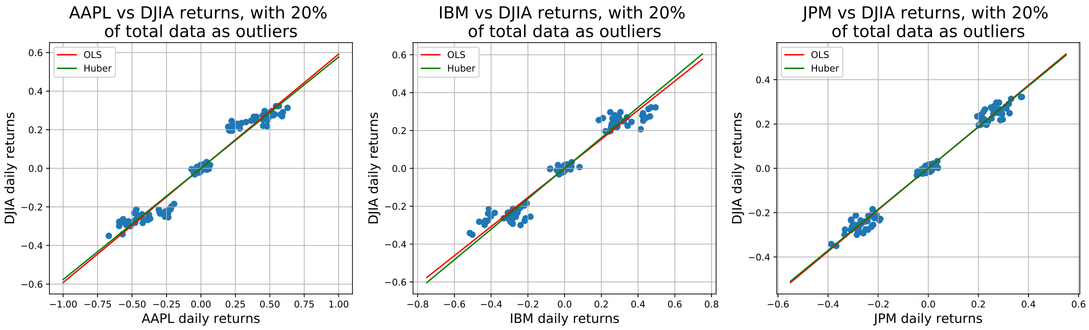


Figure 47: Linear and Huber regression between individual stock returns vs DJIA returns, with 20% of total data as outliers

Stock	R^2 OLS	R^2 Huber
AAPL	0.9518	0.9512
IBM	0.9512	0.9491
JPM	0.9844	0.9843

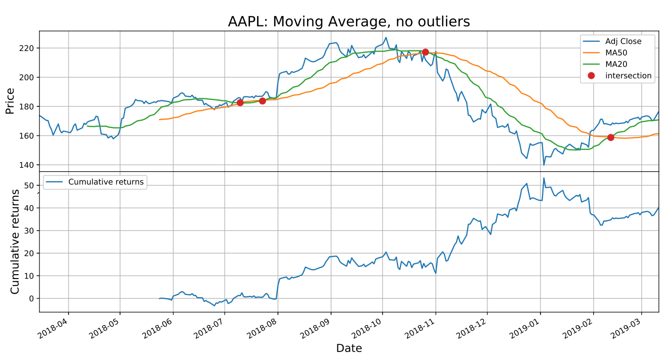
Table 5: R^2 score: with 20% of total data as outliers

3. From the R^2 scores of the regressions with and without outliers, OLS outperforms Huber in all cases. Changing the outlier percentage does not make the Huber regression perform any better. An explanation for this could be because the outliers create a symmetric distribution around the true value of the returns, and therefore it does not affect the loss function calculation.

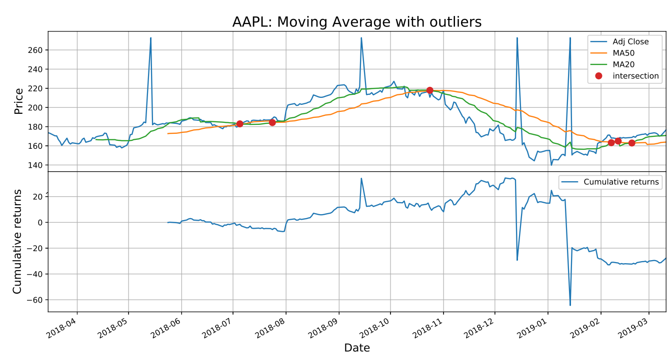
4.4 Robust Trading Strategies

1. The moving average crossover strategy is used to generate buy/sell signals. It involves the crossover between a fast moving average line and a slow moving average line. In particular, when the fast MA line (MA20) crosses the slow MA line (MA50) from below, long the stock, and when the fast line crosses the slow line from above, short the stock.

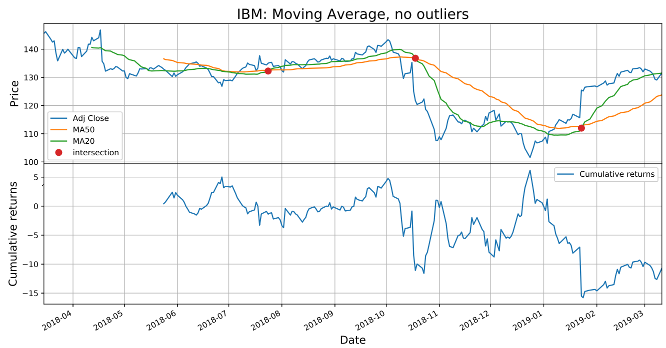
The following figures show the moving average crossover with a red marker representing the point of intersection. Additionally, the cumulative returns is plotted to show the strategy performance. Each daily action (units of stock to long or short) is defined to be +100 (long) or -100 (short) depending on the MA levels.



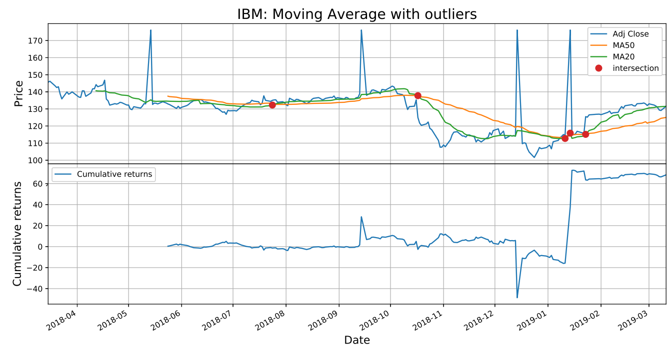
(a) AAPL moving average strategy with no outliers



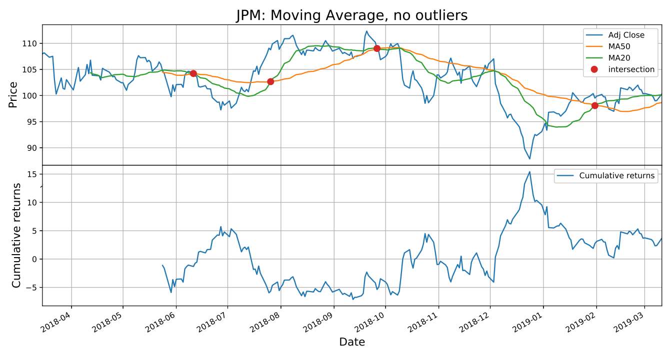
(b) AAPL moving average strategy with outliers



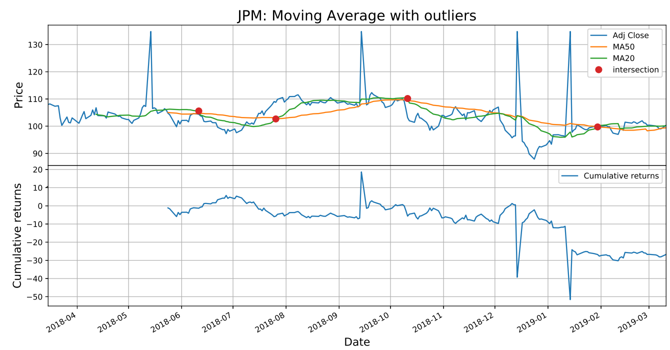
(a) IBM moving average strategy with no outliers



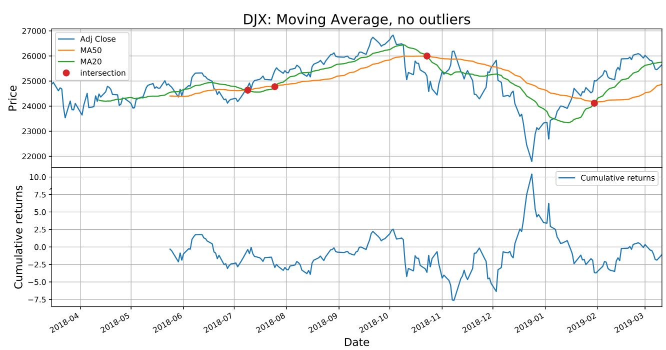
(b) IBM moving average strategy with outliers



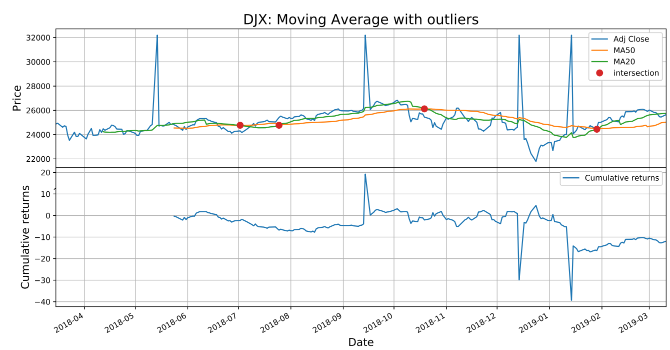
(a) JPM moving average strategy with no outliers



(b) JPM moving average strategy with outliers



(a) DJIA moving average strategy with no outliers



(b) DJIA moving average strategy with outliers

Observations

- The moving average strategy is extremely profitable with AAPL stock. The MA crossovers correctly identify turning points and the resulting trades generate high returns, especially the short positions taken between the 3rd and 4th red dots.

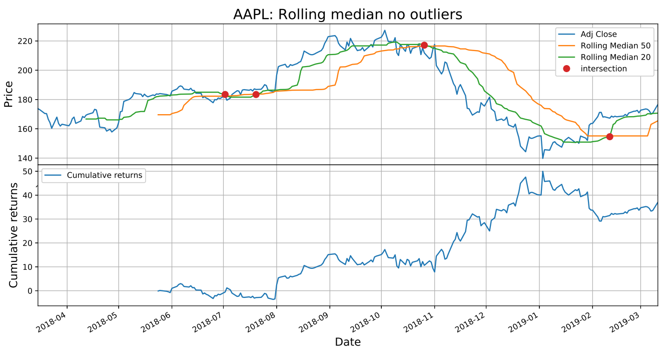
- With outliers, it makes a small profit spike in the middle when MA20 >MA50, meaning that only long positions are taken. However when the outliers are in locations where MA20 <MA50, the strategy takes a huge loss, which brings the cumulative returns to the negative territory.

- For all 4 stocks, the moving average strategy is able to recognize the market correction that happened towards the end of 2018, and it is able to profit by taking short positions.

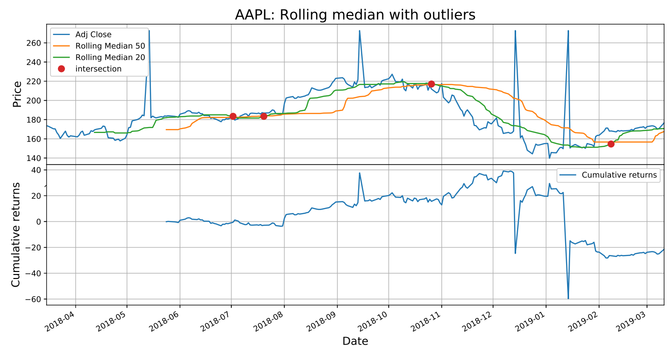
- IBM stock is relatively flat between the first two intersection dots, and therefore the cumulative returns are around 0 during that time. It identifies a turning point with the third intersection dot, but however it is one month late to the actual trough, causing sharp drop in returns before the MA20 finally crosses the MA50 line.

- With outliers, it is a similar story to AAPL stock. In all cases we have positive outliers (spikes), and therefore if the outlier lies in a region where MA20 >MA50, positive returns will be made and vice versa.
- To conclude, all moving averages are lagging indicators and will always be "behind" the price. Therefore, moving average indicators only work well when prices are following a trend, which is most clearly shown by AAPL stock. If prices are not trending, moving average signals can be misleading.

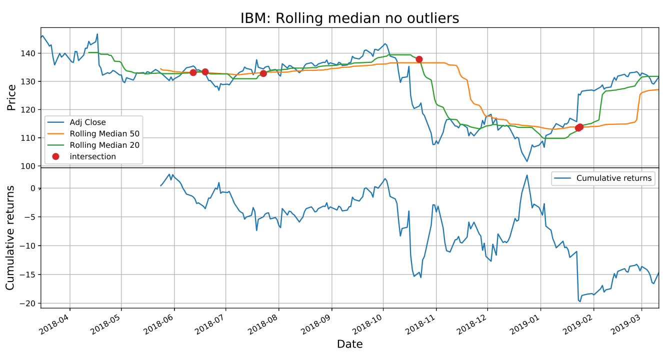
2. The above plots are repeated for a rolling median strategy, both with and without outliers.



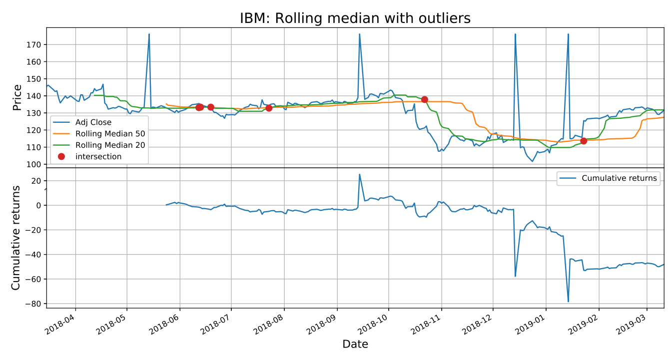
(a) AAPL rolling median strategy with no outliers



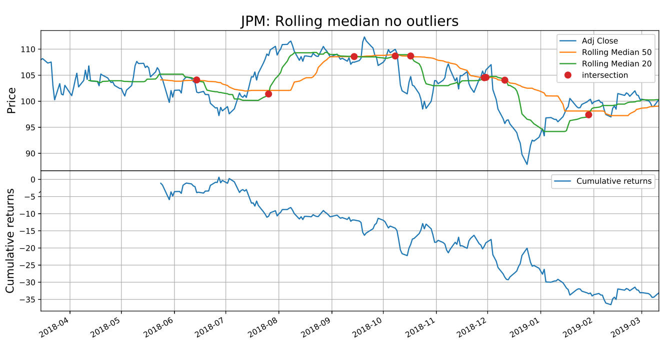
(b) AAPL rolling median strategy with outliers



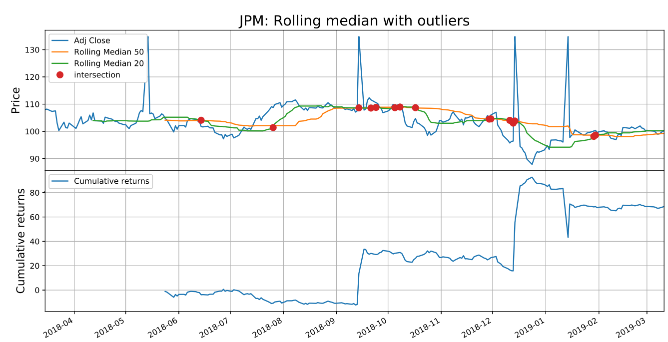
(a) IBM rolling median strategy with no outliers



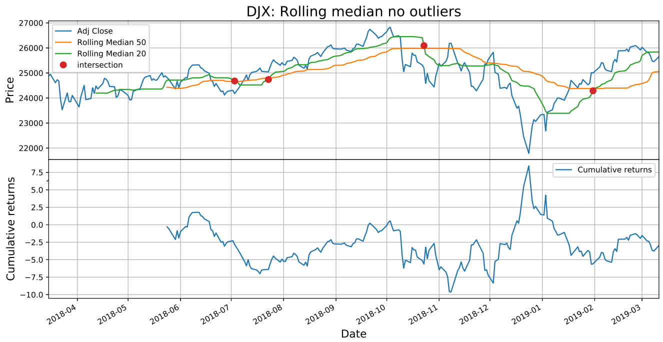
(b) IBM rolling median strategy with outliers



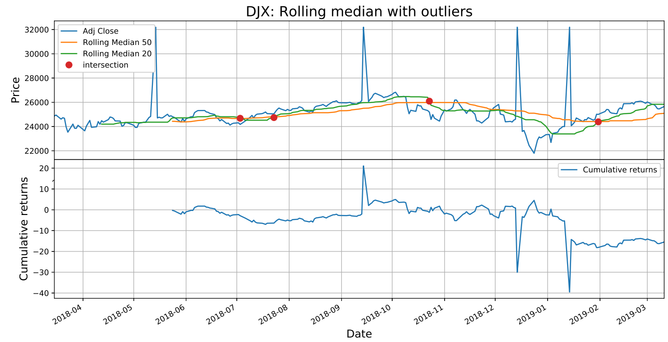
(a) JPM rolling median strategy with no outliers



(b) JPM rolling median strategy with outliers



(a) DJIA rolling median strategy with no outliers



(b) DJIA rolling median strategy with outliers

Observations

- For AAPL, the moving median result is the same as the moving average result. The intersection points are roughly the same, and this is largely due to the clear uptrend and downtrend during the considered time horizon, making both strategies profitable.
- JPM stock shows a drastic difference in cumulative returns compared with its moving average counterpart. The cumulative returns are on a steady decline, reaching as low as -35. The JPM rolling median has a lot more crossovers, its robustness makes it respond quicker to price movements. However, due to the lagged nature of rolling statistics, its response time is not fast enough to profit from short-term stock volatility.
- The effect of outliers is largely the same as the results with moving averages. The rolling median statistic could be more effective in predicting the trend for stocks that are more volatile (higher beta), allowing it to be resistant to sharp fluctuations. The drawback is that, as mentioned above, it will not be able to profit from these sharp movements at all.
- In order to reduce the lag in simple moving averages/rolling medians, the exponentially weighted moving average can be used. EMA's reduce the lag by applying more weight to recent prices relative to older prices, and can therefore be used to profit from shorter-term volatility.

5 Graphs in Finance

1. 10 random stocks in the financials sector are selected. In the list, there are two credit rating agencies (MCO, SPGI), two bank holding companies (HBAN, RF), one investment banking company (MS), three financial services companies (NTRS, CMA, AXP), one mutual funds company (BEN) and one conglomerate (BRK). Each of these companies have very different operations to each other, including the ones in the same category. The aim is to uncover the correlations between a range of companies that span several financial subsectors. The figures below show the raw prices and logarithmic returns of the 10 stocks. On first glance, the raw prices of each individual stock seem to be very correlated with one another.

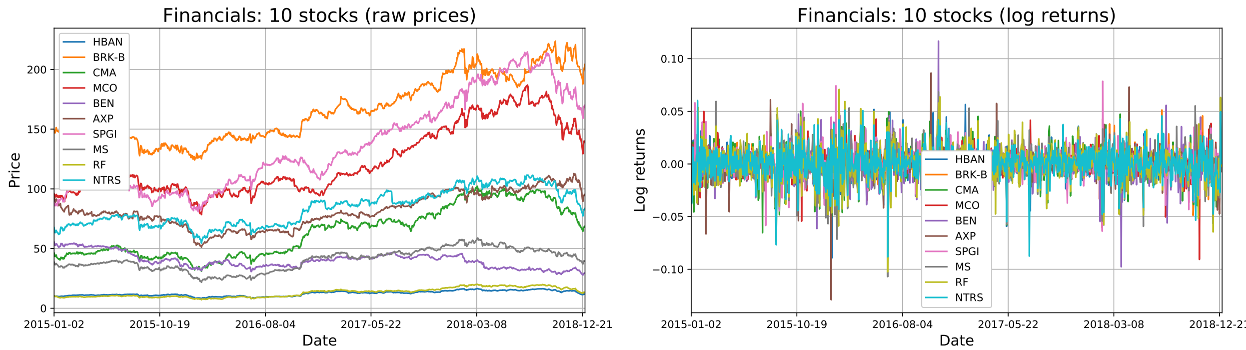


Figure 56: Price series and log returns of 10 financials stocks

2. The correlation matrix is used show the relationships between the different stocks in the matrix. If the correlation between two stocks is above a certain threshold (0.7 in this case), the two stocks are 'connected'. The graph represents this connection as an edge between these two stocks, which are represented by circles (nodes). If the correlation is below 0.7, then there is no edge connecting the stock nodes. The dependencies between stocks are then clearly visible through the clustering of nodes and their connections.

	HBAN	BRK-B	CMA	MCO	BEN	AXP	SPGI	MS	RF	NTRS
HBAN	1.000000	0.642797	0.822725	0.508116	0.591466	0.502601	0.483623	0.763018	0.846334	0.747781
BRK-B	0.642797	1.000000	0.631688	0.618854	0.614926	0.569307	0.607013	0.725367	0.643430	0.703221
CMA	0.822725	0.631688	1.000000	0.552695	0.609888	0.523950	0.523980	0.789339	0.881705	0.770088
MCO	0.508116	0.618854	0.552695	1.000000	0.552588	0.506859	0.779344	0.644927	0.534393	0.601682
BEN	0.591466	0.614926	0.609888	0.552588	1.000000	0.484832	0.530021	0.686703	0.609815	0.649299
AXP	0.502601	0.569307	0.523950	0.506859	0.484832	1.000000	0.462673	0.577062	0.518446	0.559687
SPGI	0.483623	0.607013	0.523980	0.779344	0.530021	0.462673	1.000000	0.617912	0.494936	0.575780
MS	0.763018	0.725367	0.789339	0.644927	0.686703	0.577062	0.617912	1.000000	0.799078	0.769494
RF	0.846334	0.643430	0.881705	0.534393	0.609815	0.518446	0.494936	0.799078	1.000000	0.776709
NTRS	0.747781	0.703221	0.770088	0.601682	0.649299	0.559687	0.575780	0.769494	0.776709	1.000000

Figure 57: Log returns correlation matrix

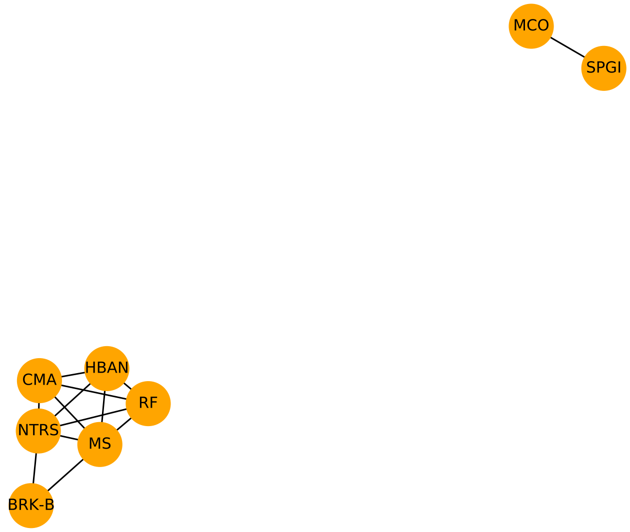


Figure 58: Log returns correlations graph

3. The topology of the graph is heavily affected by the nature of data. All of the stocks are from the financial sector, and therefore some of these log returns are correlated with one another, as indicated by the southwest clustering in figure 58. The MCO-SPGI cluster is interesting because the graph was able to distinguish the fact that MCO (Moody's Corporation) and SPGI (S&P Global Inc) are not only financial institutions, but also credit rating agencies, and hence its returns follow a similar trend. AXP (American Express) had below threshold correlations with every other stock in the list, this is because it is predominantly a credit card and charge card company, a different subsector within the financial services industry than the other stocks. Therefore it makes sense that its returns are uncorrelated.

By performing a shuffle operation to the ordering of the time series, we can see that the correlation matrix (Figure 59) contains the same values. It may be hard to tell from the correlation matrix, but plotting the graph (Figure 60 shows the exact same clustering as figure 58. Therefore, a reordering of the graph's vertices (nodes) would not change the edges and clustering of the data.

	BEN	CMA	SPGI	AXP	NTRS	MS	RF	HBAN	MCO	BRK-B
BEN	1.000000	0.609888	0.530021	0.484832	0.649299	0.686703	0.609815	0.591466	0.552588	0.614926
CMA	0.609888	1.000000	0.523980	0.523950	0.770088	0.789339	0.881705	0.822725	0.552695	0.631688
SPGI	0.530021	0.523980	1.000000	0.462673	0.575780	0.617912	0.494936	0.483623	0.779344	0.607013
AXP	0.484832	0.523950	0.462673	1.000000	0.559687	0.577062	0.518446	0.502601	0.506859	0.569307
NTRS	0.649299	0.770088	0.575780	0.559687	1.000000	0.769494	0.776709	0.747781	0.601682	0.703221
MS	0.686703	0.789339	0.617912	0.577062	0.769494	1.000000	0.799078	0.763018	0.644927	0.725367
RF	0.609815	0.881705	0.494936	0.518446	0.776709	0.799078	1.000000	0.846334	0.534393	0.643430
HBAN	0.591466	0.822725	0.483623	0.502601	0.747781	0.763018	0.846334	1.000000	0.508116	0.642797
MCO	0.552588	0.552695	0.779344	0.506859	0.601682	0.644927	0.534393	0.508116	1.000000	0.618854
BRK-B	0.614926	0.631688	0.607013	0.569307	0.703221	0.725367	0.643430	0.642797	0.618854	1.000000

Figure 59: Reordered log returns correlation matrix

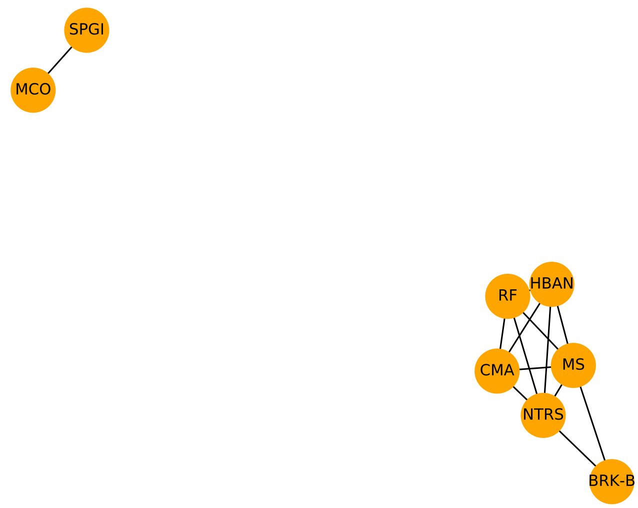


Figure 60: Reordered log returns correlations graph

4. The alternative distance metric used is the spectral distance. This is computed by taking the fft of the daily returns of each stock. The following code snippet is used to calculate the spectral distance matrix:

```
ret_fft = np.zeros((len(tickers),len(tickers)))
for i in range(len(tickers)):
    ret_fft[i,:] = np.fft.fft(logret[tickers].dropna().iloc[i,:]) #compute fft

ret_fft = pd.DataFrame(data=ret_fft,columns=tickers,index=tickers)

spect_dist = pd.DataFrame(columns=tickers.values,index=tickers.values)
for i in range(len(tickers)):
    for n in range(len(tickers)):
        spect_dist.iloc[i,n] = np.linalg.norm(ret_fft.iloc[i] - ret_fft.iloc[n]) #compute spectral distance

#plot graph
corr_thresh = 0.2
links = spect_dist.stack().reset_index()
links.columns = ['stock1', 'stock2','corr']
links_filtered=links.loc[(links['corr'] < corr_thresh) & (links['stock1'] != links['stock2'])]

plt.figure(figsize=(6,5))
G=nx.from_pandas_edgelist(links_filtered, 'stock1', 'stock2')
nx.draw(G, with_labels=True, node_color='orange', node_size=800, edge_color='black', linewidths=1, font_size=10)
```

The graph is generated by taking a reverse threshold of 0.2, i.e. spectral distances greater than 0.2 will be ignored and treated as dissimilar points. Again, the resulting network shows 2 clusters, but this time the clusters are connected by SPGI and RF. This suggests that the frequency components of the returns of the stocks in each cluster are similar to one another, meaning that they share similar trends in variability. This is a much more useful insight when building portfolios, as it can separate and cluster stocks with similar risks.

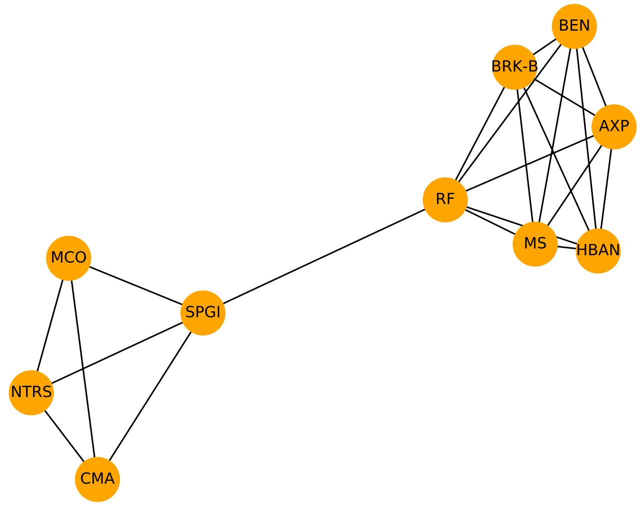


Figure 61: Log returns spectral distance graph

When performing the same shuffling operation as done in section 5.3, the resulting network topology is not the same. There is no clustering in this graph topology, and each node is densely connected to each and every other node on the diagram, forming an octagonal structure. This is because the Fourier transform calculation depends on the previous time steps. It suggests that the time series has been shuffled in a way such that the log returns of each stock fluctuate in very similar ways, and hence spectral distances are very similar. Using a lower distance threshold of 0.1 still generates a relatively dense network, but however the shuffling operation is not of much practical value.

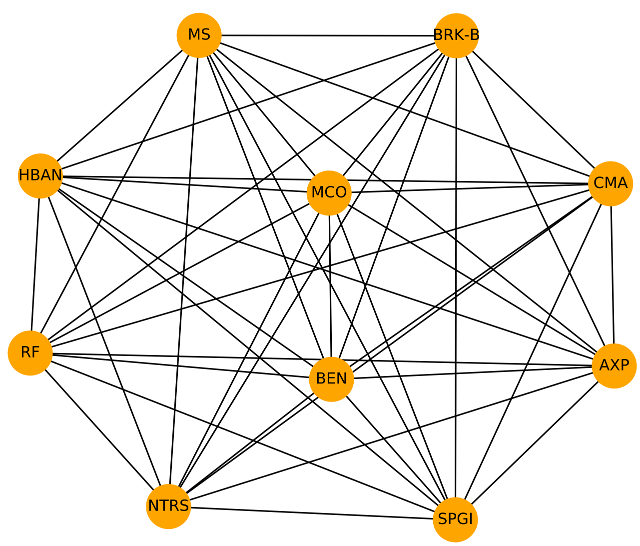


Figure 62: Shuffled log returns spectral distance graph

5. The results below are for raw prices. The correlation matrix values are much higher than with log returns, and this is further highlighted by the dense connectivity between graph nodes in figure 64. This is because prices are non-stationary, and for any statistics (including correlation) to make sense the underlying data has to be stationary. One does not typically care about high or low asset prices. Moreover, prices are not independent, early changes in the prices have more weight than later changes in the correlation calculation whereas with the returns each one has equal importance. Therefore, raw prices should not be used in correlation calculation.

	HBAN	BRK-B	CMA	MCO	BEN	AXP	SPGI	MS	RF	NTRS
HBAN	1.000000	0.899359	0.965112	0.873734	-0.151160	0.868741	0.872249	0.947844	0.974076	0.973168
BRK-B	0.899359	1.000000	0.944343	0.915740	-0.311736	0.891725	0.945038	0.856135	0.941461	0.918345
CMA	0.965112	0.944343	1.000000	0.927698	-0.273437	0.856647	0.950135	0.933288	0.995576	0.985824
MCO	0.873734	0.915740	0.927698	1.000000	-0.346642	0.885922	0.974046	0.838768	0.911528	0.915527
BEN	-0.151160	-0.311736	-0.273437	-0.346642	1.000000	-0.085244	-0.415324	0.027770	-0.278109	-0.233081
AXP	0.868741	0.891725	0.856647	0.885922	-0.085244	1.000000	0.845002	0.856276	0.852628	0.854321
SPGI	0.872249	0.945038	0.950135	0.974046	-0.415324	0.845002	1.000000	0.825941	0.934562	0.927322
MS	0.947844	0.856135	0.933288	0.838768	0.027770	0.856276	0.825941	1.000000	0.933286	0.931933
RF	0.974076	0.941461	0.995576	0.911528	-0.278109	0.852628	0.934562	0.933286	1.000000	0.982237
NTRS	0.973168	0.918345	0.985824	0.915527	-0.233081	0.854321	0.927322	0.931933	0.982237	1.000000

Figure 63: Raw prices correlation matrix

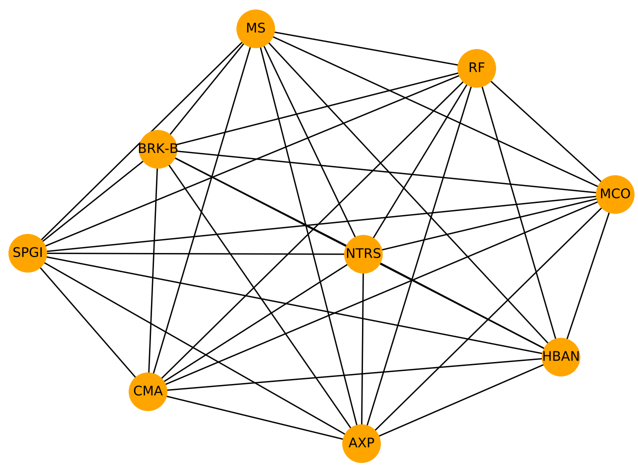


Figure 64: Raw prices correlations graph