

# Financial Signal Processing and Machine Learning [1]

Department of Electrical and Electronic Engineering

Imperial College London

Yao Lei Xu

YLX15

01062231

April 18, 2019

# Contents

<b>1</b>	<b>Regression Methods</b>	<b>4</b>
1.1	Processing Stock Price Data in Python . . . . .	4
1.1.1	Natural-Log Transform . . . . .	4
1.1.2	Stationarity . . . . .	4
1.1.3	Returns . . . . .	5
1.1.4	Advantages of log-returns . . . . .	6
1.1.5	Simple vs Log Returns (I) . . . . .	7
1.1.6	Simple vs Log Returns (II) . . . . .	7
1.2	ARMA vs. ARIMA Models for Financial Applications . . . . .	8
1.2.1	Suitability of ARMA and ARIMA Models . . . . .	8
1.2.2	ARMA Modelling . . . . .	9
1.2.3	ARIMA Modelling . . . . .	10
1.2.4	Prices vs Log-Prices with ARIMA . . . . .	10
1.3	Vector Auto-Regressive (VAR) Models . . . . .	12
1.3.1	Concise VAR . . . . .	12
1.3.2	Optimal VAR Coefficients . . . . .	13
1.3.3	Eigenvalues of VAR . . . . .	13
1.3.4	VAR Portfolio Analysis I . . . . .	14
1.3.5	VAR Portfolio Analysis II . . . . .	16
<b>2</b>	<b>Bond Pricing, CAPM and APT</b>	<b>18</b>
2.1	Examples of Bond Pricing . . . . .	18
2.1.1	Effective Rates . . . . .	18
2.1.2	Equivalent Rates (I) . . . . .	18
2.1.3	Equivalent Rates (II) . . . . .	18
2.2	Forward Rates . . . . .	19
2.2.1	Forward Rates Equivalent . . . . .	19
2.3	Duration of a Coupon- Bearing Bond . . . . .	20
2.3.1	Duration . . . . .	20
2.4	Capital Asset Pricing Model (CAPM) and Arbitrage Pricing Theory (APT) . . . . .	21
2.4.1	Non-Weighted Market Return . . . . .	21
2.4.2	Non-Weighted Rolling Beta . . . . .	22
2.4.3	Weighted Market Return . . . . .	22
2.4.4	Weighted Rolling Beta . . . . .	23
2.4.5	Arbitrage Pricing Theory . . . . .	23
<b>3</b>	<b>Portfolio Optimization</b>	<b>28</b>
3.1	Adaptive Minimum-Variance Portfolio Optimization . . . . .	28
3.1.1	Lagrangian Formulation . . . . .	28
3.1.2	Static Minimum-Variance Portfolio . . . . .	29
3.1.3	Adaptive Minimum-Variance Portfolio . . . . .	31
<b>4</b>	<b>Robust Statistics and Non-Linear Methods</b>	<b>33</b>
4.1	Data Import and Exploratory Data Analysis . . . . .	33
4.1.1	Descriptive Statistics . . . . .	33
4.1.2	Histograms and PDF . . . . .	34

4.1.3	Mean vs Median Z-Score Outlier Detection . . . . .	34
4.1.4	Impact of Outliers . . . . .	37
4.1.5	Box Plots . . . . .	39
4.2	Robust Estimators . . . . .	40
4.2.1	Custom Robust Estimators . . . . .	40
4.2.2	Complexity Analysis . . . . .	40
4.2.3	Breakdown Points . . . . .	41
4.3	Robust and OLS Regression . . . . .	42
4.3.1	OLS Regression . . . . .	42
4.3.2	Robust Regression (Huber Regression) . . . . .	42
4.3.3	Regression and Outliers . . . . .	43
4.4	Robust Trading Strategies . . . . .	44
4.4.1	Simple Moving Average-Cross Over . . . . .	44
4.4.2	Robust Moving Median Cross Over . . . . .	45
<b>5</b>	<b>Graphs in Finance</b>	<b>48</b>
5.1	S&P Stocks Selection . . . . .	48
5.2	Correlation Based Graphs . . . . .	49
5.3	Correlation Graph Analysis . . . . .	50
5.4	Spectral Similarity Graphs . . . . .	51
5.5	Price vs Return Graphs . . . . .	54

# 1 Regression Methods

## 1.1 Processing Stock Price Data in Python

### 1.1.1 Natural-Log Transform

The figure below shows the price of the SPX index since 1930 and the natural-log transform of the price. The natural-log transform compresses the range of the price data, which makes it easier to process, together with other advantages discussed later on.

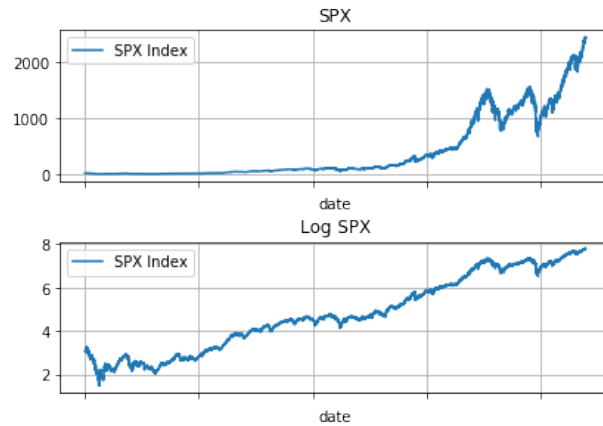


Fig. 1 Prices vs log-prices

### 1.1.2 Stationarity

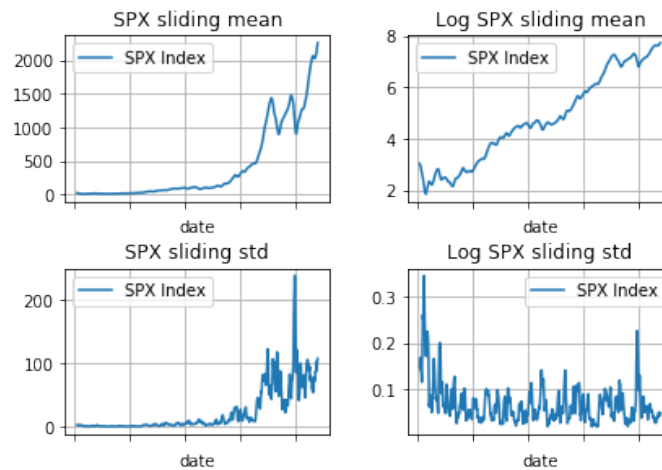


Fig. 2 Prices: Sliding mean and standard deviation

A stochastic signal is considered wide-sense stationary (WSS) if the first moment (the mean) is time invariant and the autocorrelation function is dependent on time differences only.

$$m_x(t) = m_x(t + T), \text{ for any } T \quad (1)$$

$$R_{xx}(t_1, t_2) = R_{xx}(t_1 - t_2), \text{ for any } t_1, t_2 \quad (2)$$

Stationarity is an important characteristic to explore in financial markets, since such characteristics would imply that the signal's average persists over time and that the autocorrelation is dependent on time differences only. This would justify investing strategies based on mean-reverting behaviour of the signal.

The figures below shows the sliding mean and standard deviation of the prices series (left column) and log-prices series (right column), which demonstrates the stationarity characteristics of the given time series. Overall, the sliding mean (or the moving average) smooths the signal, while the sliding standard deviation represents its volatility over the past year (252 trading days).

As shown in figure 2, the moving average increases over time for both the price and the log-price, which implies non-stationarity. However, the sliding standard deviation shows more consistent behaviour, with log-price's value oscillating about 0.068. Overall, the price signal is not stationary.

### 1.1.3 Returns

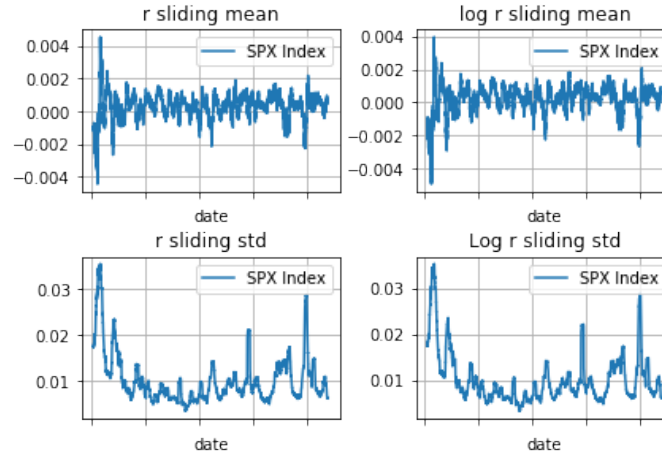


Fig. 3 Returns: Sliding mean and standard deviation

In stead of looking at the price data, it is often more useful to look at the return characteristics of the signal in terms of stationarity. In this case, the simple return is defined as the percentage increase at each time step

$$r_{simple} = \frac{p_t - p_{t-1}}{p_{t-1}} \quad (3)$$

while the log return is defined as the difference between successive log values

$$r_{log} = \ln \frac{p_t}{p_{t-1}} = \ln(p_t) - \ln(p_{t-1}) \quad (4)$$

Figure 3 shows the sliding window analysis applied to the returns signal. By looking at the returns, the signal is much more well-behaved, with the moving average for both returns oscillating about 0, hence showing more stationarity.

### 1.1.4 Advantages of log-returns

In quantitative finance, prices are often assumed to be log-normally distributed in the short-run (in the long-run it is more skewed due to the market being generally upward trending), which means that the log return  $\log(\frac{p_{t+1}}{p_t}) = \log(1 + r_{simple})$  is normally distributed. This can be shown visually with the histograms below, which plots the distribution of daily returns. This additional advantage of Gaussianity is essential for signal processing, since most signal processing models and other statistical techniques assume Gaussianity (e.g. Z-score). In addition, by looking at returns, it creates a metric that makes assets comparable with each other, which is essential for multi-dimensional cross-assets analysis.

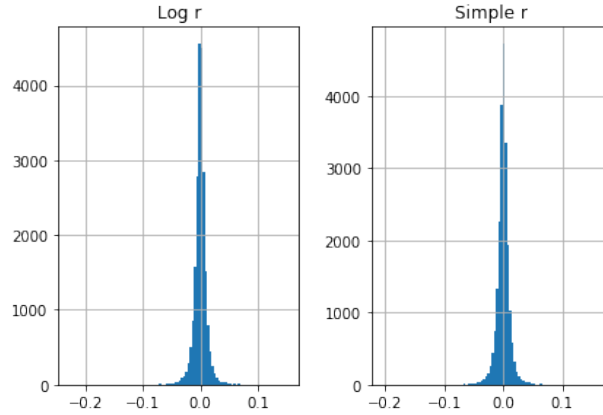


Fig. 4 Returns distribution

To formally test Gaussianity, it is possible to use the "Jarque-Bera" test which checks to what extent the sample data's skewness (how symmetric is the sample data) and kurtosis (how fat is the tail of the distribution) matches a normal distribution, which has skewness 0 and kurtosis 3. The statistic is defined as:

$$JB = \frac{n}{6} \left( S^2 + \frac{(C - 3)^2}{4} \right) \quad (5)$$

where  $n$  is the number of observations,  $S$  is the sample skewness and  $C$  is the sample kurtosis. Therefore, the closer the data sample follows a normal distribution with  $S=0$  and  $C=3$ , the closer to 0 is the JB statistic.

Figure 5 below shows the difference between the JB test statistic for simple returns and log returns. It is clear from the graph that as number of data points increases, the greater is the difference between the JB value for simple returns and log returns, meaning that simple returns deviates further away from the normal distribution. This proves that the log-returns shows more Gaussianity than simple returns. In addition, the p-value for the JB statistic is around 0 for both cases for number of data points greater than 300, which shows high confidence about the above conclusion.

Finally, the JB test statistic shows more normality in the short run for both the log and simple returns, which is expected as the upward trend in the market over the long run skews the return distribution.

For signal processing purposes, log-returns are often preferred for several reasons other than its normality. For instance, it is time additive for calculating compounded returns (explored in the next section) which is computationally and numerically stable. In addition, it's better suited

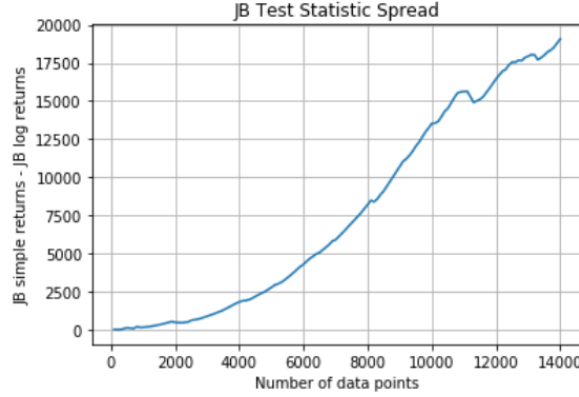


Fig. 5 JB Test Statistic Spread

for mathematical analysis since  $\frac{p_{t+1}}{p_t} = e^{\ln(\frac{p_{t+1}}{p_t})}$  and exponents are easier to manipulate with integration and differentiation in stochastic calculus. Finally, for small values of  $r_{simple}$ , the log return  $\log(\frac{p_{t+1}}{p_t}) = \log(1 + r_{simple}) \approx r_{simple}$ , which is a valid assumption for the short term.

### 1.1.5 Simple vs Log Returns (I)

If a purchased stock doubles and halves its values in the following days from £1 to £2 to £1, the simple and log returns would be [ 1.00 -0.50] and [ 0.69 -0.69] respectively. Given that the sum of logarithmic returns is zero and that the purchased stock did not change value at all, it is implied that arithmetic sum of log-returns is more suitable than simple returns to describe the value gained by the asset over time.

$$p_t = p_{t-k} \exp\left(\sum_{i=1}^k r_{log_k}\right) \quad (6)$$

$$p_t = p_{t-k} \prod_{i=1}^k (1 + r_{simple_k}) \quad (7)$$

This has several computational advantages (mathematical tractability and numerical stability) since logarithms and exponentials are easier to manipulate with calculus and addition of small number is more stable than multiplication.

### 1.1.6 Simple vs Log Returns (II)

It is important to emphasize that log-normality exists only in the short term, so it is not ideal for long-term analysis. In addition, the log-returns are additive across time, but not across assets (portfolio). In the portfolio formulation, simple returns are preferred as they are additive across assets.

## 1.2 ARMA vs. ARIMA Models for Financial Applications

### 1.2.1 Suitability of ARMA and ARIMA Models

The following figures show the daily closing log-price of the S&P500 over the last 4 years, as well as its sliding mean and standard deviation. By the same analysis carried out in section 1, it is a non-stationary signal due to the presence of the trend.

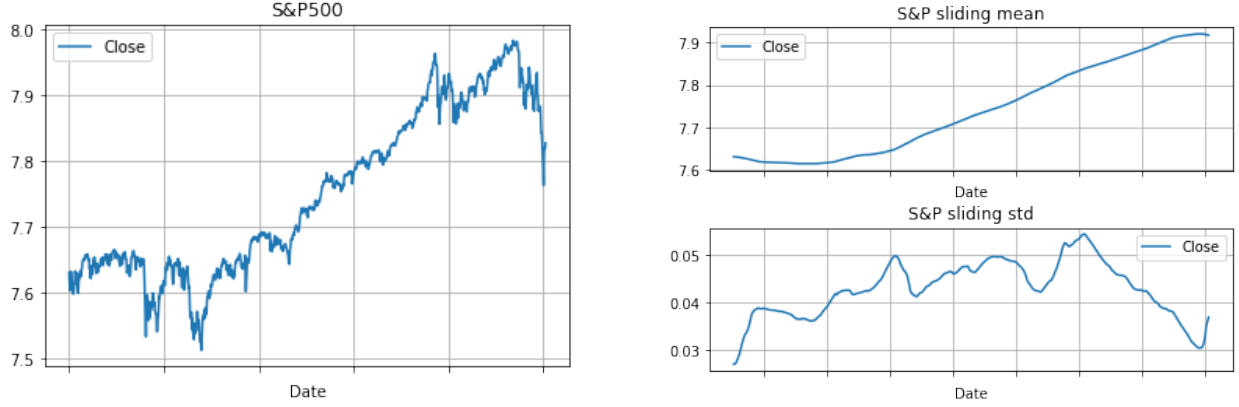


Fig. 6 S&P500 closing prices

An auto-regressive (AR) process is a process for which the future values depends on its past values plus noise, and is mathematically defined as:

$$x[t] = a_1x[t-1] + \dots + a_px[t-p] + \eta[t] = \sum_{i=1}^p a_ix[t-i] + \eta[t] \quad (8)$$

where the autocorrelation coefficients are found by finding the vector  $\mathbf{a}$ , which is estimated using the Yule-Walker equations:

$$\mathbf{a} = \mathbf{R}_{xx}^{-1} \mathbf{r}_{xx} \quad (9)$$

where  $\mathbf{R}_{xx}$  and  $\mathbf{r}_{xx}$  are formed by the autocorrelation functions such that:

$$\mathbf{R}_{xx} = \begin{bmatrix} r_0 & r_{-1} & \dots & \eta^2 \\ r_1 & r_0 & \dots & r_{p-2} \\ \dots & \dots & \dots & \dots \\ r_{p-1} & r_{p-2} & \dots & r_0 \end{bmatrix} \quad (10)$$

$$\mathbf{r}_{xx} = \begin{bmatrix} r_1 \\ r_2 \\ \dots \\ r_p \end{bmatrix} \quad (11)$$

The error term from the auto-regressive model can be further modelled as a linear combination of its past values, which forms the moving-average (MA) part of an auto-regressive moving-average (ARMA) model, which is defined as:

$$x[t] = \sum_{i=1}^p a_ix[t-i] + \sum_{i=1}^q b_i\eta[t-i] + \eta[t] \quad (12)$$



In the financial sector in particular, the AR part of the ARMA model aims to explain the momentum and mean reversion effects of the markets (often explained with participation effect), while the MA part captures the "noise" that is caused by exogenous shocks (such as news) that can't be explained by the past data.

However, ARMA models assumes stationarity in the data, which means that it would perform badly with a signal such as the S&P 500 prices series since there is a strong trend present. For this reason, an ARIMA model is more appropriate since the integrating part (I in ARIMA) applies a differencing operation which removes the elements of non-stationarity.

### 1.2.2 ARMA Modelling

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 AR(1). 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 (on the left) and of last 4 years from 2015 to 2019 (figure on the right), 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. Finally, there is a large residual peak at the beginning due to non-stationarity.

In addition, the model parameter in this case is 7.740, which is greater than 1. This means that the process is fundamentally explosive and non-stationary, which is to be expected due to the uptrend present in the SPX (stability of the AR process is explored in 1.3.3).

In practice, the given AR(1) model makes assumptions that often does not reflect reality. For instance, the process is assumed to be a pure martingale by having an auto-regression of order 1. Also, effects of shocks is also not modelled by the white noise moving average, which is also not reflective of reality. Finally, the prediction is basically a lagged version of the signal, which suggests that the above model is not really useful in practice, as it assumes that the next day's price is today's price.

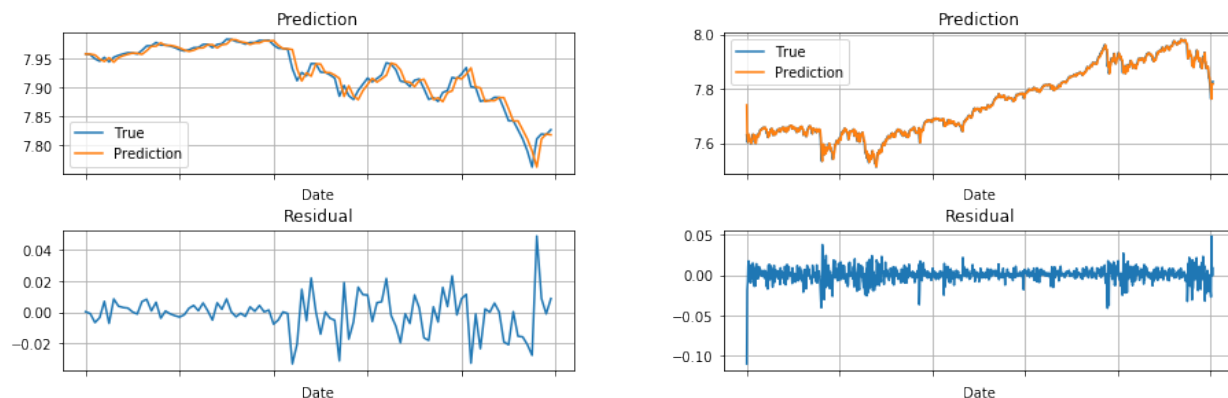


Fig. 7 ARMA log-prices

### 1.2.3 ARIMA Modelling

This section will use an ARIMA(1,1,0) to fit the time series, which essentially applies an order-1 AR model as before on an initial differencing (of order 1) on the log-price series to remove non-stationarity. Figure below shows that there is a slight improvement for this model (mean absolute residual of 0.005848, which is an improvement of 2%), and there is no peak present at beginning unlike the ARMA model.

Finally, the AR parameter of this model is -0.008751, which is less than 1 in magnitude, meaning that the returns series is stable and stationary unlike the explosive ARMA model from the previous section. The parameter is also negative, which means that successive returns tend to be negatively correlated. Finally, the parameter is near 0 in value, which means that there is little correlation between successive returns, and that the model isn't good enough for practical purposes.

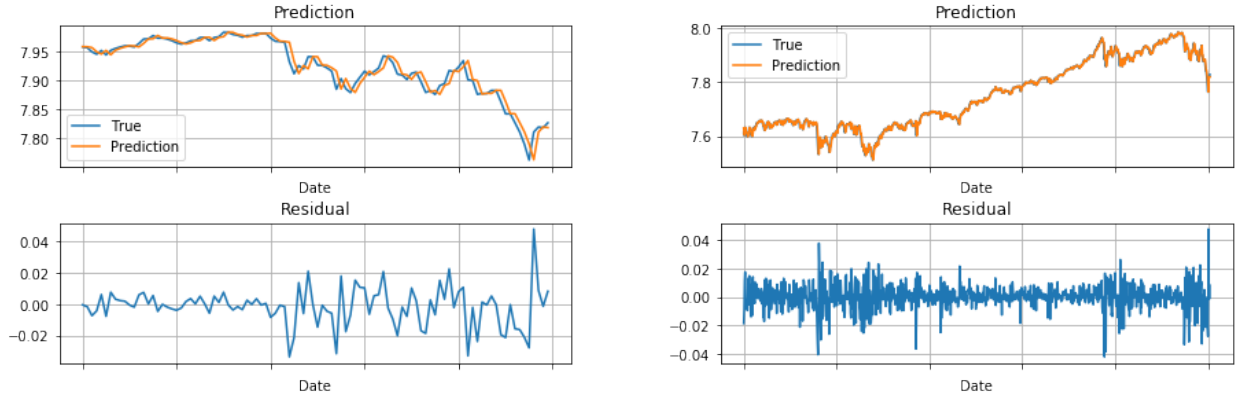


Fig. 8 ARIMA log-prices

The residual characteristics of the ARIMA model is slightly better than the ARMA model since the differencing improves upon the non-stationarity of the trend present in the data. However, the performance remains similar since the volatility caused by the exogenous events is still not captured by the MA term.

In terms of physical meaning, the ARIMA model helps understand the final signal better. This is because applying the ARIMA model with integrating order of 1 is the equivalent of applying the ARMA model on the time-series differenced once. And since the differencing is applied on the log-prices, this means that the ARIMA model is predicting the next log-return as a function of its past values instead of predicting next log-price. This also make the values of this model comparable with other assets.

$$\ln(p_t) - \ln(p_{t-1}) = \ln \frac{p_t}{p_{t-1}} = r_{log} \quad (13)$$

This therefore also explains why a first order differencing of  $\ln(p_t) - \ln(p_{t-1})$  makes the signal stationary, since the log-returns are more normally distributed as shown in the histogram from the previous sections.

### 1.2.4 Prices vs Log-Prices with ARIMA

As briefly mentioned in the previous section, the ARIMA model applies the ARMA model to the time-series differenced once. This is much more meaningful in the context of log-prices since the difference of log prices is the log-return.

$$r_{log} = \ln \frac{p_t}{p_{t-1}} = \ln(p_t) - \ln(p_{t-1}) \quad (14)$$

Therefore, the ARIMA model is fitting an auto regressive model on successive returns, which not only makes more physical sense but also improves the quality of modelling by making it stationary. This is in addition to all the benefits of using log-prices as mentioned in the section 1.1.4.

The figure below shows the results of ARIMA model applied to the prices series directly (but brought back to the log space afterwards to make the comparison sensible). For this model, the MAR is of 0.005849, which is about 0.02% worse than the ARIMA applied to log-prices.

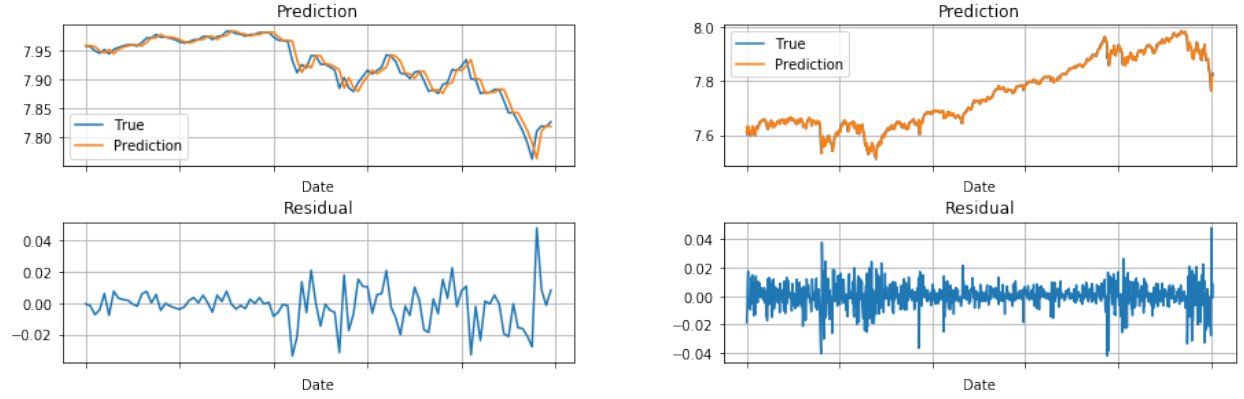


Fig. 9 ARIMA prices

### 1.3 Vector Auto-Regressive (VAR) Models

#### 1.3.1 Concise VAR

The multivariate extension of the AR model is the vector auto-regressive (VAR) model, where the next time value is dependent on the past values of multiple variables, with the residual being captured by an error term  $\mathbf{e}_t$  as shown in the equation below. This is particularly useful when multiple time-series exhibit auto-regression characteristics across variables, meaning that there is correlation in time (auto-correlation) and in space (across variables). The VAR(p) process is mathematically defined as:

$$\mathbf{y}_t = \mathbf{c} + \mathbf{A}_1 \mathbf{y}_{t-1} + \dots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{e}_t \quad (15)$$

where  $\mathbf{y}_i \in \mathbb{R}^{K \times 1}$  for  $i$  in  $[t-p, t]$  are vectors with data on  $k$  assets,  $\mathbf{c} \in \mathbb{R}^{K \times 1}$  is a vector of constants,  $\mathbf{A}_j \in \mathbb{R}^{K \times K}$  for  $j$  in  $[1, p]$  are the parameter matrices, and  $\mathbf{e}_t \in \mathbb{R}^{K \times 1}$  is the error vector.

However, by taking a bases view of the matrix, it is possible to re-write the above model concisely by composing the following matrices:

$$\mathbf{U} = \mathbf{e}_t \quad (16)$$

$$\mathbf{Y} = \mathbf{y}_t \quad (17)$$

$$\mathbf{B} = [\mathbf{c} \quad \mathbf{A}_1 \quad \mathbf{A}_2 \quad \dots \quad \mathbf{A}_p] \quad (18)$$

$$\mathbf{Z} = \begin{bmatrix} 1 \\ \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \dots \\ \mathbf{y}_{t-p} \end{bmatrix} \quad (19)$$

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

Therefore, a concise matrix form of the VAR model can be written as:

$$\mathbf{Y} = \mathbf{BZ} + \mathbf{U} \quad (20)$$

In addition, it is possible to generalize the VAR model further by extending the above equation to the multi-period case such that multiple  $T$  time-steps are modelled. This means adding additional columns for additional times steps to  $\mathbf{Y}$ ,  $\mathbf{Z}$ , and  $\mathbf{U}$ . This way, the above equation holds for matrices of dimensions  $\mathbf{B} \in \mathbb{R}^{K \times (KP+1)}$ ,  $\mathbf{Z} \in \mathbb{R}^{(KP+1) \times T}$ ,  $\mathbf{U} \in \mathbb{R}^{K \times T}$ , and  $\mathbf{Y} \in \mathbb{R}^{K \times T}$  (Hence, equation 15 is a special case of equation 20 for  $T = 1$ ). The matrices for multi steps predictions are therefore changed to:

$$\mathbf{Y} = [\mathbf{y}_t \quad \mathbf{y}_{t+1} \quad \dots \quad \mathbf{y}_{t+T-1}] \quad (21)$$

$$\mathbf{U} = [\mathbf{e}_t \quad \mathbf{e}_{t+1} \quad \dots \quad \mathbf{e}_{t+T-1}] \quad (22)$$

$$\mathbf{Z} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \mathbf{y}_{t-1} & \mathbf{y}_t & \dots & \mathbf{y}_{t-1+T} \\ \mathbf{y}_{t-2} & \mathbf{y}_{t-1} & \dots & \mathbf{y}_{t-2+T} \\ \dots & \dots & \dots & \dots \\ \mathbf{y}_{t-p} & \mathbf{y}_{t-p+1} & \dots & \mathbf{y}_{t-p+T} \end{bmatrix} \quad (23)$$

### 1.3.2 Optimal VAR Coefficients

To find the solution to the linear equation  $\mathbf{Y} = \mathbf{BZ} + \mathbf{U}$ , the least-squares method is used to minimize the squared elements of  $\mathbf{U}$ , since it essentially represents the error in the model. Therefore, this problem can be set up as an optimization problem with the following objective function  $J(\mathbf{B})$ :

$$J(\mathbf{B}) = (\mathbf{BZ} - \mathbf{Y})(\mathbf{BZ} - \mathbf{Y})^T \quad (24)$$

$$J(\mathbf{B}) = \mathbf{BZZ}^T\mathbf{B}^T - 2\mathbf{YZ}^T\mathbf{B}^T + \mathbf{YY}^T \quad (25)$$

The problem can be optimized by setting the derivative of  $J(\mathbf{B})$  to zero, which yields the optimal coefficients  $\mathbf{B}_{opt}$  as the solutions to the of the VAR model:

$$\frac{\delta J(\mathbf{B})}{\delta \mathbf{B}} = 2\mathbf{BZZ}^T - 2\mathbf{YZ}^T = 0 \quad (26)$$

$$\mathbf{BZZ}^T = \mathbf{YZ}^T \quad (27)$$

$$\mathbf{B}_{opt} = \mathbf{YZ}^T(\mathbf{ZZ}^T)^{-1} \quad (28)$$

### 1.3.3 Eigenvalues of VAR

Consider a VAR(1) process, that is:

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \mathbf{e}_t \quad (29)$$

$$\mathbf{y}_{t-1} = \mathbf{A}\mathbf{y}_{t-2} + \mathbf{e}_{t-1} \quad (30)$$

$$\mathbf{y}_{t-2} = \mathbf{A}\mathbf{y}_{t-3} + \mathbf{e}_{t-2} \quad (31)$$

$$\dots \quad (32)$$

By substituting them in a recursive form, the equation for  $\mathbf{y}_t$  can be re-written for up to a lag of  $p$  as:

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \mathbf{e}_t \quad (33)$$

$$\mathbf{y}_t = \mathbf{A}(\mathbf{A}\mathbf{y}_{t-2} + \mathbf{e}_{t-1}) + \mathbf{e}_t \quad (34)$$

$$\mathbf{y}_t = \mathbf{A}^2\mathbf{y}_{t-2} + \mathbf{A}\mathbf{e}_{t-1} + \mathbf{e}_t \quad (35)$$

$$\dots \quad (36)$$

$$\mathbf{y}_t = \mathbf{A}^p\mathbf{y}_{t-p} + \sum_{i=0}^{p-1} \mathbf{A}^i\mathbf{e}_{t-i} \quad (37)$$

Since the eigenvalues of matrix  $\mathbf{A}$  follows the property  $\mathbf{A}^p\mathbf{v} = \lambda^p\mathbf{v}$ , for the VAR model to be stable the eigenvalues of the matrix  $\mathbf{A}$  must be less than 1 in absolute value such that  $\lim_{p \rightarrow \infty} \lambda^p = 0$ , otherwise the power will explode in value.

### 1.3.4 VAR Portfolio Analysis I

This section will explore how the VAR models can be used for portfolio construction. First, the closing price of assets CAG, MAR, LIN, HCP, MAT are examined. The price series is detrended by subtracting the prices series from their moving average of 66 data points (number of trading days in a quarter) to fabricate zero-mean stationary time serieses, which are plotted in the figure below (However, notice that LIN contains very few data point, so most of the time series is deleted when analysing the VAR model).

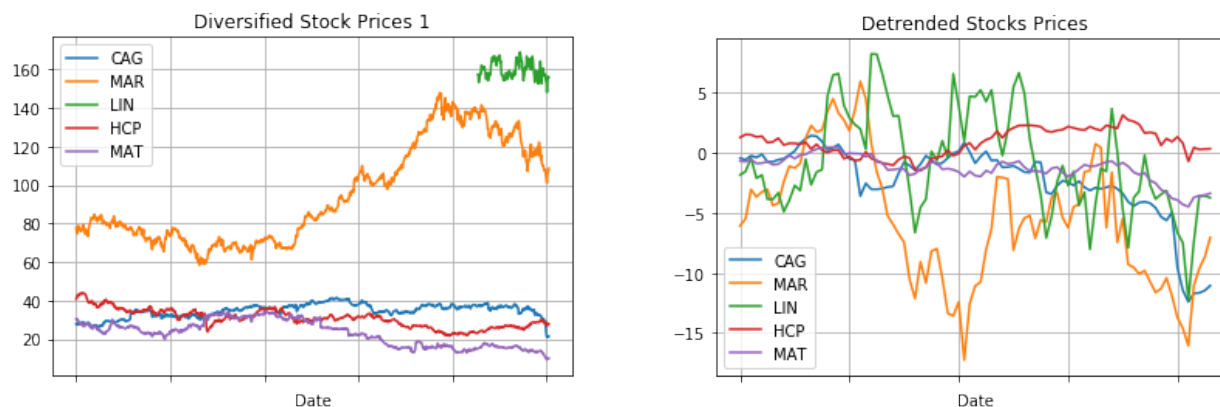


Fig. 10 *Diversified stocks ensemble*

The table below shows the parameters grid of the fitted VAR(1) model. As expected, the diagonal of the matrix below contains the highest values, which implies that the time series tends to be positively auto-correlated with its own lagged value, but not so much with other assets since they are near 0 or negative. The relationship between the parameters below and correlation is intuitive since positive auto-regression parameters of order 1 implies past prices positively influences future prices.

	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

This is also shown in figure 11 which plots the time-series value of CAG against its own lagged value (with VAR parameter of 0.873, correlation coefficient of 0.966), showing a strong linear relationship that reflects the high diagonal values shown in the table above. Conversely LIN and lagged HCP tend to be inversely correlated, which is reflected in its negative parameter value in the table above (VAR parameter of -0.401, correlation coefficient of -0.284).

The above VAR parameters matrix can therefore be interpreted almost as a covariance matrix given its relationship with correlation. Hence, by applying the eigenvalue and eigenvector analysis, it is possible to extract the direction and magnitude of highest variance (with the greatest eigenvalue) within the variables. However, even if the VAR parameters and covariance matrix share similar characteristics, this analysis fundamentally differs from PCA in many ways. For instance, the above matrix is asymmetric, which yields eigenvalues that are imaginary.

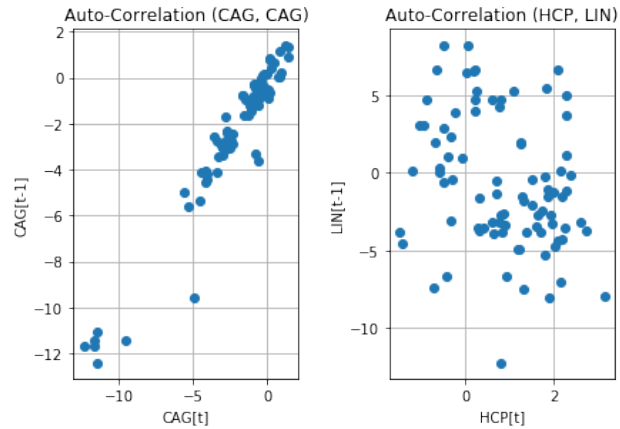


Fig. 11 Price series correlations

Overall, the eigenvalues of the above matrix average about 0.846 in magnitude, with the highest being 1.006 and the smallest being 0.726. Therefore, the assets tend to be uncorrelated with each other, which is expected since they are from different industry sectors as shown in the table below. Therefore, it would make sense to construct a portfolio using these stocks (the reason is explored in the next section). In addition, given that most of the eigenvalues have positive real parts, the given stocks form an unstable system.

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

However, the above example contains the stock LIN which has less than 20% of the data compared to other stocks in the portfolio. To make sure the analysis carried above is correct, the same analysis is repeated with APD, which is in the same GICS sub industry as LIN but has more data.

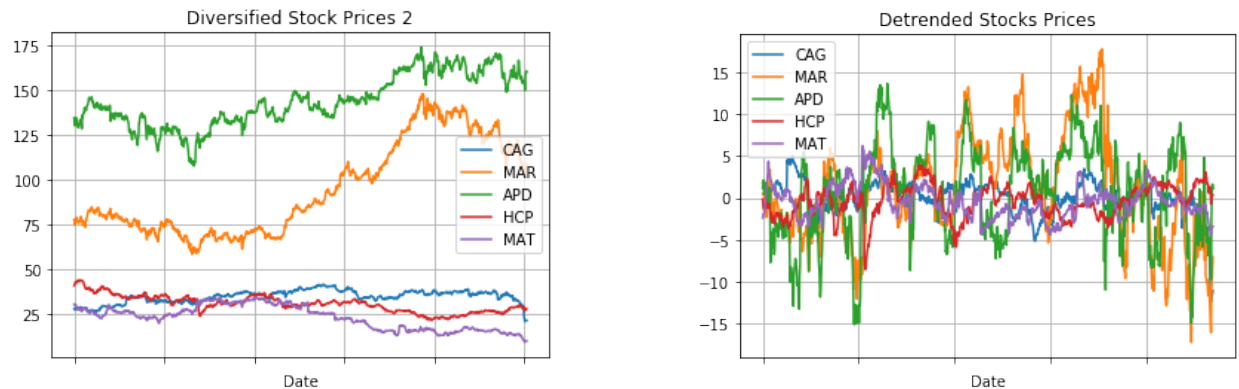


Fig. 12 Diversified stocks ensemble

For this example, the analysis shows similar pattern, with the diagonal containing the highest values, while other coefficients are closer to zero. The magnitude of eigenvalues in this case have average about 0.966, with the minimum being 0.942 and max being 0.979, which means that they are even more un-correlated than the example before.

	CAG	MAR	APD	HCP	MAT
L1.CAG	0.976601	0.003245	-0.043526	0.011506	0.001898
L1.MAR	0.002488	0.966825	0.019019	0.000366	-0.000140
L1.APD	-0.001881	0.006991	0.945082	-0.001559	-0.002263
L1.HCP	-0.029259	-0.071794	-0.023169	0.972232	-0.008258
L1.MAT	0.013399	-0.025500	0.014128	-0.010201	0.967046

### 1.3.5 VAR Portfolio Analysis II

	min eigenvalue	mean eigenvalue	max eigenvalue
Industrials	0.371246	0.763932	0.991721
Health Care	0.0921573	0.62144	0.994153
Information Technology	0.374081	0.809351	0.992738
Communication Services	0.752488	0.926293	0.982263
Consumer Discretionary	0.447563	0.810433	0.99065
Utilities	0.0421154	0.599877	0.985648
Financials	0.152575	0.631218	1.00434
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

The same eigenvalue analysis is carried out by grouping stocks of same sectors together. The min, mean and max eigenvalue magnitudes are presented in the table above. As expected, about 70% of the portfolio formed by grouping stocks from the same sector obtained eigenvalues for their VAR parameters matrix which are smaller than the ones mentioned in the previous section (100% for the case of APD instead of LIN). This means that the amount of variation in a given set of stocks is little, which implies that they tend to be more correlated in movement with each other, hence varying little in differences.

Generally, it is not advisable to build a portfolio by grouping the stocks together by their sector. This is because the set of assets chosen will have exposure to the same type of risks that can affect the particular sector (e.g. 2008 mortgage backed securities crisis affected the real-estate sector way more than other ones), which means that they are more correlated with each other.

The diversification premium of the portfolio is achieved when the stocks are un-correlated with each other, since the return of the portfolio is  $r_p = \sum_{i=1}^n w_i r_i$  and the standard deviation is  $\sigma^2 = \sum_{i,j=1}^n w_i \sigma_{ij} w_j$  where  $w_i, r_i$  is the weight and the return of an asset while  $\sigma_{ij}$  is the covariance of returns between two assets.

For instance, take 2 randomly chosen stocks, the return and the standard deviation of the portfolio is given by:

$$r_p = w_1 r_1 + (1 - w_1) r_2 \quad (38)$$

$$\sigma_p = \sqrt{w_1^2 \sigma_1^2 + 2w_1(1 - w_1)\rho\sigma_1\sigma_2 + (1 - w_1)^2 \sigma_2^2} \quad (39)$$

For simplicity, assume that both assets have return  $r$  and standard deviation  $\sigma$  and that  $w_1 = 0.5$ . The Sharpe ratio  $\frac{r_p}{\sigma_p}$  for perfectly correlated portfolio with  $\rho = 1$  is  $\frac{r}{\sigma}$ , where as for



perfectly uncorrelated (identically and independently distributed assets) portfolio with  $\rho = 0$  is  $\sqrt{2} \frac{r}{\sigma}$  (generally, for  $n$  iid assets, the Sharpe ratio is  $\sqrt{n} \frac{r}{\sigma}$ ).

Therefore, by forming a portfolio of different uncorrelated assets, it is possible to increase the Sharpe ratio, which means that there is more return per unit of risk taken. This is shown in the table below, where the Sharpe ratio is calculated for portfolios by sector and by aggregating the entire S&P500 using daily returns (which has one of the highest Sharpe ratio).

	return mean	return std	return sharpe
Information Technology	0.000635	0.011749	0.054007
Health Care	0.000526	0.010517	0.049984
Communication Services	0.000374	0.011035	0.033905
<b>S&amp;P500 Portfolio</b>	<b>0.000297</b>	<b>0.008799</b>	<b>0.033719</b>
Industrials	0.000283	0.009521	0.029746
Financials	0.000274	0.011037	0.024866
Utilities	0.000229	0.009506	0.024100
Materials	0.000208	0.010765	0.019299
Consumer Staples	0.000152	0.007968	0.019127
Consumer Discretionary	0.000146	0.009703	0.015098
Real Estate	0.000113	0.009610	0.011740
Energy	-0.000150	0.016196	-0.009262

## 2 Bond Pricing, CAPM and APT

### 2.1 Examples of Bond Pricing

#### 2.1.1 Effective Rates

Consider the following example: An investor receives USD 1100 in one year in return for an investment of USD 1000 now, gaining an effective rate  $r_{eff}$  of return of 10%. Under different compounding intervals, the nominal interest rates per annum will vary accordingly. Formally, the relationship between the effective interest rates  $r_{eff}$  and the nominal interest rate  $r$  is  $1 + r_{eff} = (1 + \frac{r}{m})^m$  where  $m$  is the number of compounding periods in a year. This means that  $r = m(1 + r_{eff})^{\frac{1}{m}} - m$ .

Therefore, for (a) annual, (b) semi-annual and (c) monthly compounding ( $m = 1, 2, 12$  respectively), the nominal rate of return per annum for an effective rate of 10% are:

- (a) 10.00%
- (b) 9.76%
- (c) 9.57%

Finally, up to the  $k^{th}$  period, the compounding is  $(1 + \frac{r}{m})^k = (1 + \frac{r}{m})^{mt}$ , and by using (d) continuous compounding where  $m$  approaches infinity, the rate becomes  $\lim_{m \rightarrow \infty} (1 + \frac{r}{m})^{mt} = e^{rt}$ . Hence, the continuous compounding rate of return after 1 year is  $r = \ln(1 + r_{eff})$  which is:

- (d) 9.53%

#### 2.1.2 Equivalent Rates (I)

If  $r = 15\%$  per annum for monthly compounding, the effective rate is  $r_{eff} = (1 + \frac{r}{12})^{12} - 1 = 16.08\%$ . The equivalent nominal rate for continuous compounding according to the equation in section 2.1.1. is therefore  $r = \ln(1 + 0.1608) = 14.91\%$

#### 2.1.3 Equivalent Rates (II)

If  $r = 12\%$  for continuous compounding, the effective rate is  $r_{eff} = e^r - 1 = 12.75\%$ . The equivalent nominal rate for quarterly compounding is therefore  $r = 4((1 + r_{eff})^{\frac{1}{4}} - 1) = 12.18\%$ . In terms of cash flow, this would generate for each quarter USD 30.45, 31.38, 32.33, 33.32 for an initial investment of USD 1000.

## 2.2 Forward Rates

### 2.2.1 Forward Rates Equivalent

#### 2.2.1.a. Comparison Principle

For the arbitrage opportunity to not exist, the following equivalence must hold:  $(1 + r_2)^2 = (1 + r_1)(1 + f_{1,2})$ , or  $(1 + r_j)^j = (1 + r_i)^i(1 + f_{i,j})^{j-i}$  in its general form, where  $f_{i,j}$  is called the implied forward rate. Therefore, the extra earnings has no additional benefits as it is the market expectation of what the 1 year spot rate will be next year. The investor should be neutral about investing for 1 year with  $r_1$  and 2 years with  $r_2$ .

#### 2.2.1.b. Implied and Market Forward Rates

On choosing between investing periods, the investor should consider the opportunity cost of investing elsewhere and liquidity rather than the rates offered for different investment periods, as the market expectation of the next year's spot rates are already included in the rates. This implied forward rate should be rather close to the market forward rate, which can be slightly different due to market imperfections.

#### 2.2.1.c. Advantages and Disadvantages of Forward Rates

The forward rate  $f_{1,2}$  is naturally higher than the rates  $r_1$  and  $r_2$  since the investor is taking the risk of future interest rates changes that can potentially impact the investment, which is both the advantage and disadvantage of forward rates since the rates can change for the better or worse.

#### 2.2.1.d. Investment Period Changes

To go to 2 year investments after 1 year, the rates will change according to the market expectations and can be calculated with the implied forward rate equation in section 2.2.1.a. using the current spot rate curves.

## 2.3 Duration of a Coupon- Bearing Bond

### 2.3.1 Duration

#### 2.3.1.a. Macaulay Duration

$$D = \sum_k t_k \frac{PV_k}{PV_{tot}} \quad (40)$$

The Macaulay Duration  $D$  as defined in the equation above is a measure of sensitivity of the bond price relative to change in yields. For the bond detailed in the table below,  $D$  is simply the sum of the last row's elements, which is 6.76. Note that for zero-coupon bonds, the duration is the same as the maturity date.

	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_k	9.52	9.07	8.64	8.23	7.84	7.46	717.79	768.55
PV_k/PV_tot	0.0124	0.0118	0.0112	0.0107	0.0102	0.0097	0.0934	1
t_k*PV_k/PV_tot	0.0124	0.0236	0.0337	0.0428	0.051	0.0583	6.5377	6.7595

#### 2.3.1.b. Modified Duration

$$D_m = \frac{1}{P(\lambda_0)} \frac{dP(\lambda)}{d\lambda} \text{ for } \lambda = \lambda_0 \quad (41)$$

Modified duration  $D_m$  is another measure of the bond sensitivity defined in the equation above, which is the derivative of the price-yield curve normalized by its price, which comes from the Taylor's series approximation of the price-yield curve.  $D_m$  can be proved to be related to the Macaulay duration as  $D_m = \frac{D}{1 + \frac{y}{m}}$ .

Given the yield  $y$  of 5%,  $D_m = \frac{D}{1 + \frac{y}{m}} = 6.4376\%$  where  $m = 1$  since the payments are annual. This value is slightly lower than the Macaulay duration due to the yield. Also, note that the two measures of duration would be the same under continuous compounding since  $m$  will tend to infinity.

#### 2.3.1.c. Advantages of Duration

The bonds with longer time to maturity are more sensitive to the change in yield due more coupon payments, which makes pension plans highly volatile. For instance, consider the above bond where yield increases by 1%. Approximately, the equation  $\frac{\Delta P}{P} \approx -D_m \Delta y$  holds, which implies a drop of 6.44%. This however can be overcome through Immunization, which means matching the desired price and duration of a bond portfolio via  $P = \sum_i n_i P_i$  and  $D = \sum_i n_i \frac{P_i}{P} D_i$  where  $i$  denotes the  $i^{th}$  bond,  $P$  the price,  $n$  the amount and  $D$  the duration (Macaulay or Modified).

In theory, the Taylor's series approximation can be expanded to  $2^{nd}$  order terms (bond convexity) and higher order terms and match as many terms as possible.

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

This section explores the application of CAPM and APT using the daily return of 157 European companies.

CAPM is an extension of the modern portfolio theory which prices individual assets in the presence of a risk-free asset, relating their rate of return to their beta/systematic risks (the risks that can't be diversified away). The CAPM is formulated as  $E[r_i] = r_f + \beta(E[r_m] - r_f)$ , where  $r_i$  is the return of the given company,  $r_m$  is the market return,  $r_f$  is the risk free rate and  $\beta = \frac{cov(r_i, r_m)}{var(r_m)}$ . Without taking the expectation, the CAPM is  $r_i = r_f + \beta(r_m - r_f) + \epsilon_i$ , where  $\epsilon_i$  is an uncorrelated zero-mean company specific risk random variable that can be diversified away since  $E[\epsilon_i] = 0$ .

APT is an extension of CAPM which makes fewer assumptions. The main difference is that instead of pricing the asset according to its systematic risk, it uses multiple macroeconomic and specific factors (size of the company, momentum, etc.) to decompose the return of a given asset. APT is formulated as  $r_i = \alpha_i + \sum_j \beta_{ij}F_j + \epsilon_i$  where  $\alpha_i$  is the excess return,  $\beta_{ij}$  is the exposure to the  $j^{th}$  factor  $F_j$  of the  $i^{th}$  company, and the  $\epsilon_i$  has the same definition from CAPM.

### 2.4.1 Non-Weighted Market Return

This section will estimate the market return as the average of company returns. In terms of data pre-processing, since there are some missing data present in the file, the companies with missing values are eliminated from the dataset, hence working with 141 companies in total. Then, the non-weighted market return is calculated by simply averaging the returns of all the companies per day. The returns are plotted in the figure below, with an average daily return of 0.004% and standard deviation of 0.00665, while the weighted volatility (1/Sharpe ratio, which is the standard deviation weighted by the average return) is around 141.



Fig. 13 Non-Weighted Market Returns

### 2.4.2 Non-Weighted Rolling Beta

Here the  $\beta$  is calculated on a rolling basis of 22 days window (1 trading month). This is done by calculating the covariance of each of the asset and the market normalized by the market variance. The rolling  $\beta$  is shown in the figure below. As shown in the histogram, most of the beta center around 1 meaning that most assets are positively correlated with the overall market (the market factor explains most of the individual returns), with a standard deviation of 0.71, which demonstrates low volatility overall. Finally, the peaks in the rolling beta graph are examples of idiosyncratic shocks associated with  $\epsilon$  discussed before, which can be due to news and other factors. Overall the zero-mean distribution means that these are risks that can be diversified away.

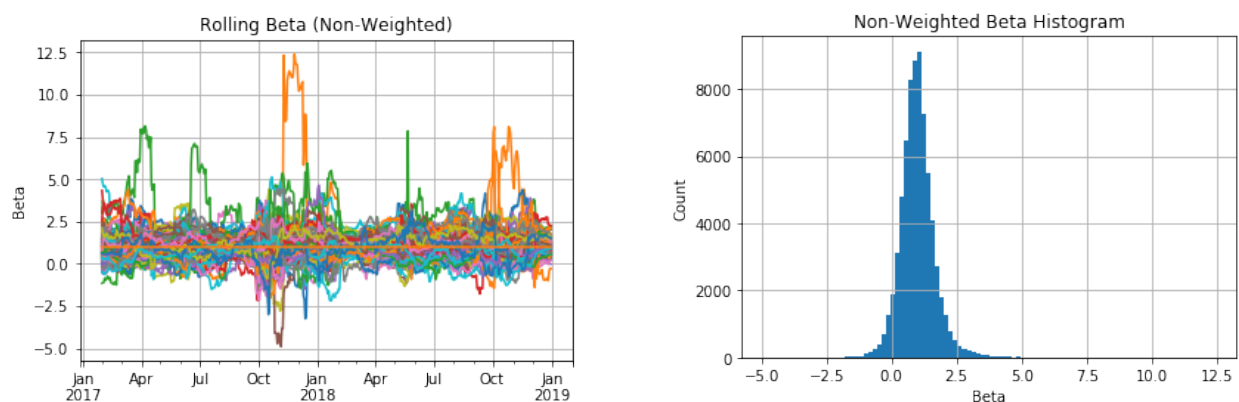


Fig. 14 *Non-Weighted Rolling Beta*

### 2.4.3 Weighted Market Return

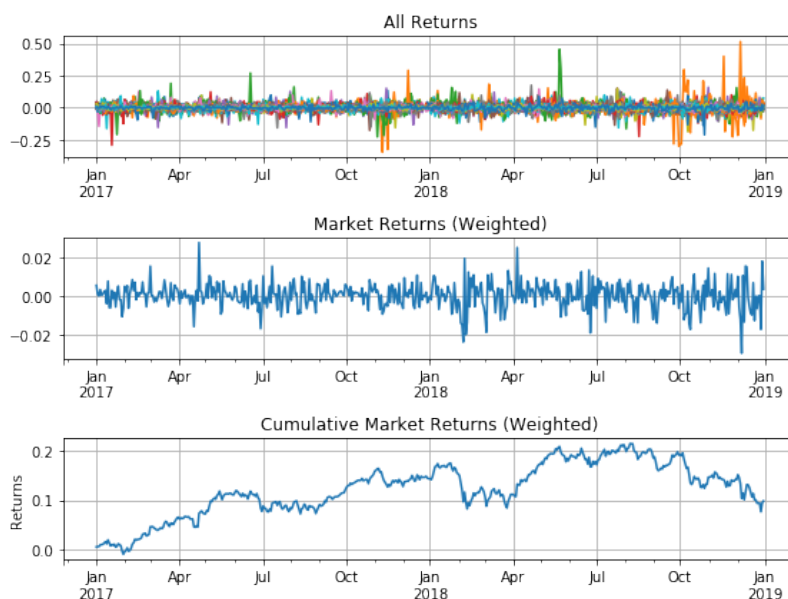


Fig. 15 *Weighted Market Returns*

The weighted market return is calculated as  $R_m = \sum_i r_i \frac{mcap_i}{\sum_i mcap_i}$ . This forms a portfolio that is weighted relative to market capitalization. This is hence forming a market portfolio, which according to the one-fund theorem, can form any efficient portfolio in the Markowitz sense (optimal mean-variance portfolio) by combining it with a risk-free asset.

The figure above shows the weighted market returns, which has a average return of 0.018% and standard deviation of 0.00660. Note that the 1/Sharpe ratio in this case is 35, which is much smaller than the non-weighted returns case of 141. This is because the small-cap stocks are riskier in nature, but their volatility is less expressed in the market portfolio since the exposure is inversely proportional to the market capitalization.

#### 2.4.4 Weighted Rolling Beta

The rolling  $\beta$  calculation is repeated for the cap weighted market returns. The figure below shows the rolling cap weighted beta and the overall distribution. This histogram too has a smaller standard deviation (0.69 instead of 0.71), which is expected due to the risky nature of small-cap companies.

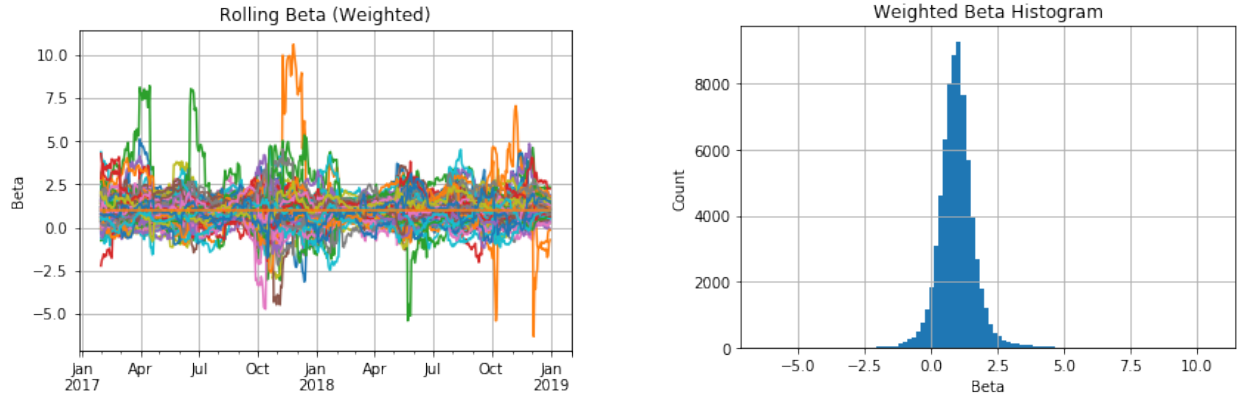


Fig. 16 *Weighted Rolling Beta*

#### 2.4.5 Arbitrage Pricing Theory

##### 2.4.5.a. Factor Returns Estimation

For this section, the APT is assumed to hold for a 2-factors model as follows:

$$r_i = \alpha + \beta_{m,i}R_m + \beta_{s,i}R_s + \epsilon_i \quad (42)$$

where  $\alpha$  is the excess return (which can be interpreted as the risk-free return),  $\beta_{m,i}$  is the exposure to the return  $R_m$  which is the return specific to the market return factor, while  $\beta_{s,i}$  is the exposure to the return  $R_s$  which is the return specific to the market capitalization factor and  $\epsilon_i$  is the specific return. Here the exposure to size is assumed to be  $\beta_{s,i} = \ln(\text{size})$ .

The Ordinary Least Squares method is used in this case to estimate one  $\alpha$ , one  $R_m$  and one  $R_s$  per day which satisfies the following matrix multiplication:

$$\mathbf{r} = \mathbf{X}\mathbf{b} + \mathbf{e} \quad (43)$$

$$\begin{bmatrix} r_1 \\ r_2 \\ \dots \\ r_n \end{bmatrix} = \begin{bmatrix} 1 & \beta_{m,1} & \beta_{s,1} \\ 1 & \beta_{m,2} & \beta_{s,2} \\ \dots & \dots & \dots \\ 1 & \beta_{m,n} & \beta_{s,n} \end{bmatrix} \begin{bmatrix} \alpha \\ R_m \\ R_s \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_n \end{bmatrix} \quad (44)$$

Hence, the solution to the OLS problem above is  $\hat{\mathbf{b}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{r}$ , which is computed for everyday.

#### 2.4.5.b. Analysis of Estimated Parameters

The figure below shows the daily values of the estimated parameters, while the table shows some statistics about the parameters. As shown in the table, the magnitude of  $\alpha$  is the biggest on average, which is not surprising as it reflects the average upward trend of the market. The market factor and the size factors play an important role as well. Although it is tempting to say that market factor is larger than the size factor in terms of magnitude and hence explains the returns better, it is important to notice also that the average exposure to the market factor is about 1, while the exposure to size is about 20.

Finally, most of the parameters are zero-mean random variables, with  $R_m$  exhibiting the largest mean normalized standard deviation, which reflects the chaotic nature of the market.

	mean	stddev	stddev/mean
Rm	-0.000285951	0.00791252	27.6709
Rs	0.000192386	0.00173227	9.00413
a	-0.00421565	0.0411911	9.771
Rm	0.00572882	0.00545938	0.952967
Rs	0.00129531	0.00116475	0.899203
a	0.0307899	0.0276514	0.898066

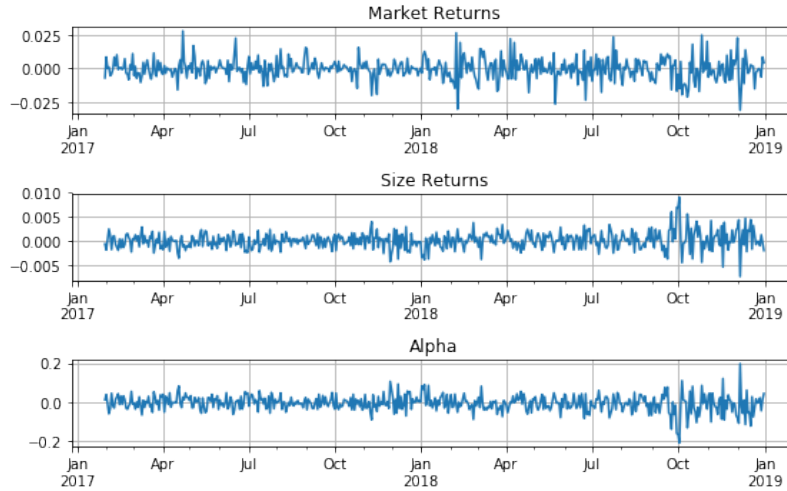


Fig. 17 *Estimated Market Factors*



### 2.4.5.c. Specific Returns Analysis

Factors based return decomposition is done on a spatial level (across stocks), while this section looks at the temporal characteristics. Specifically, the correlation over time between the company returns and specific returns  $\epsilon_i$  are calculated, which are plotted in the figures below. Since  $\epsilon_i$  is defined as the difference between the real returns and the returns explainable by the factors, the high magnitude of these correlations (averaging about 81%) shows that the 2 factors are not sufficient for explaining the given returns (otherwise the correlation would be 0). The correlation is also shown in the scatter plot as the points are centered around the  $y = x$  black line.

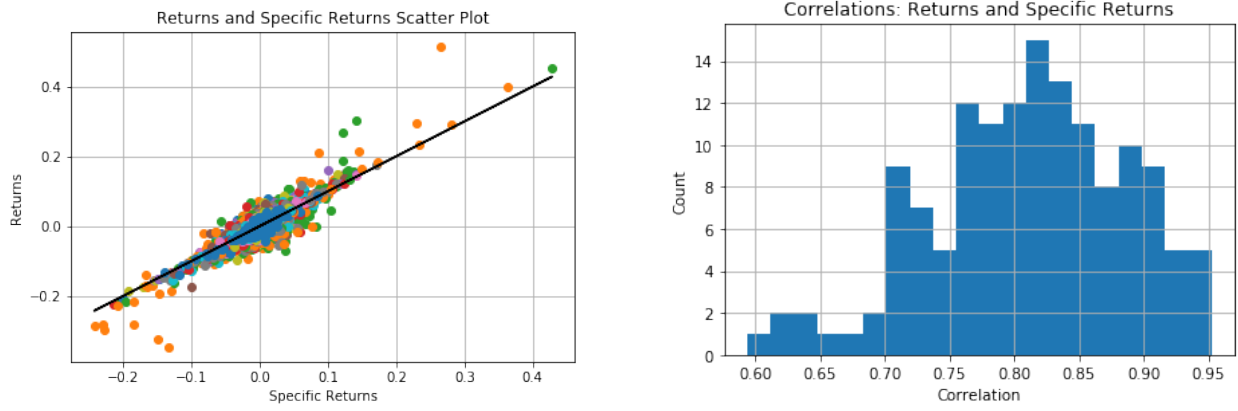


Fig. 18 Returns and Specific Returns Correlation

### 2.4.5.d. Factor Returns Analysis

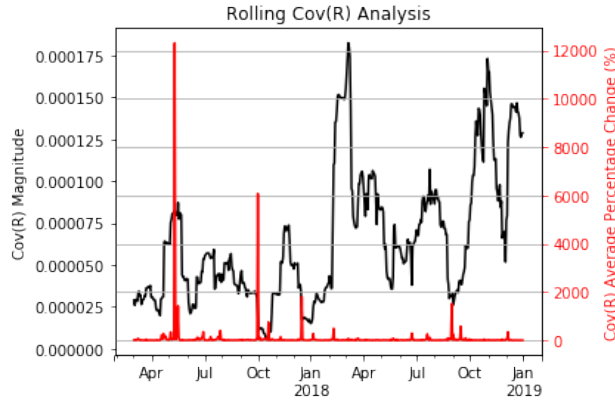


Fig. 19 Rolling Covariance Matrix Characteristics

This section explores the magnitude and the stability of covariance matrices  $cov(\mathbf{R}) \in \mathbb{R}^{2 \times 2}$  estimated using a rolling window of 22 days on the factor return matrix  $\mathbf{R}$ , which is defined as follows:

$$\mathbf{R} = \begin{bmatrix} Rm_1 & Rs_1 \\ Rm_2 & Rs_2 \\ \dots & \dots \\ Rm_{500} & Rs_{500} \end{bmatrix} \quad (45)$$

The magnitude of the rolling covariance matrices  $cov(\mathbf{R})$  is plotted in the figure above in black, and it is clearly a highly fluctuating signal, which suggests instability. In addition, the average percentage change between each of the 4 elements in the covariance matrix is calculated for successive time steps, which is plotted in the figure in red. The covariance values can change drastically from one day to another, which also highlights the unstable nature of the covariance matrix. This demonstrates that past estimations of covariances tend to be obsolete for future uses.

#### 2.4.5.e. PCA of Specific Returns

This section explores the application of PCA on the covariance matrix of the specific return matrix  $\mathbf{E}$ . Specifically, from the 2 factors model mentioned in the previous section, the specific return is computed as the difference between the actual and estimated company returns for every companies every day, which are collected in the following Matrix:

$$\mathbf{E} = \begin{bmatrix} \epsilon_{i=1,t=1} & \dots & \epsilon_{i=N,t=1} \\ \epsilon_{i=1,t=2} & \dots & \epsilon_{i=N,t=2} \\ \dots & \dots & \dots \\ \epsilon_{i=1,t=T} & \dots & \epsilon_{i=N,t=T} \end{bmatrix} \quad (46)$$

The PCA is a dimensionality reduction technique which captures the direction of highest variation in data by projecting the data into a feature space, which offers a quantitative method for identifying factors that explain the variances in the returns. This is done by carrying out the eigenvalue analysis on the covariance matrix, where the eigenvectors represent the directions of most variation (as a linear combination of given features), while the eigenvalues represent how much variation there is for its corresponding eigenvector.

The figure below shows the eigenvalues arranged in descending order. As it is clear from the figure, the first components contains most of the information as the explained variances decrease rapidly. The first component in particular explains about 7.37% of the total variances.



Fig. 20 Explained Variance

The market return is often the primary driver of stock returns, which is why the first principal component is usually loosely equal to the market return when applying the PCA analysis on the return matrix instead. In the following graph, the blue line is the cumulative return of the portfolio which uses the first principal component as weights (the principal component is scaled to have a sum of 1), while the red line is the market return computed previously.

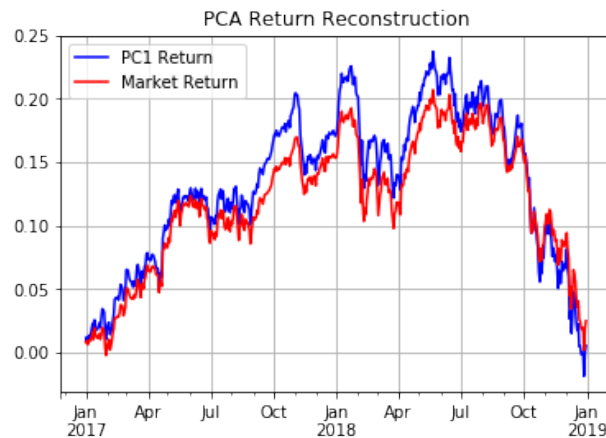


Fig. 21 *PC1 of Returns and the Market Factor*

However, since the market factor and the size factor are already taken into account in the above APT model, when applying the PCA to the specific returns matrix, the first principal component is not the market return. As it is shown in the figure below, the returns are computed by using the non-normalized first principal component as weights of specific returns, which is clearly different from the market return. The first component in this case can indicate some other factors relating to the momentum premium or the value vs growth premium. Finally, given that the first component explains only 7.37% of the variances, it highlights the chaotic nature of the stock market where multiple factors must be taken into account to explain most of the price movements.



Fig. 22 *PC1 of Specific Returns*

### 3 Portfolio Optimization

#### 3.1 Adaptive Minimum-Variance Portfolio Optimization

##### 3.1.1 Lagrangian Formulation

This section will explore the minimum variance portfolio optimization problem. Specifically, given a set of weights  $\mathbf{w}$  in a vector form, the return and the variance of the portfolio at a given time are:

$$r_p = \mathbf{w}^T \mathbf{r} \quad (47)$$

$$\sigma_p^2 = \mathbf{w}^T \mathbf{C} \mathbf{w} \quad (48)$$

where  $\mathbf{r}$  is a vector containing the returns of the assets and  $\mathbf{C}$  is the covariance matrix which each  $(i, j)$  element represents the covariance between  $i^{th}$  and  $j^{th}$  assets  $\sigma_{i,j} = \sigma_i \sigma_j \rho$ .

The minimum-variance portfolio can therefore be formulated as the following constrained optimization problem:

$$\mathbf{w}_{opt} = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{2} \mathbf{w}^T \mathbf{C} \mathbf{w} \text{ s.t. } \mathbf{w}^T \mathbf{1} = 1 \quad (49)$$

which has the following Lagrangian:

$$L = \frac{1}{2} \mathbf{w}^T \mathbf{C} \mathbf{w} - \lambda (\mathbf{w}^T \mathbf{1} - 1) \quad (50)$$

By differentiating w.r.t.  $\mathbf{w}$  and  $\lambda$  and assuming that  $\mathbf{C}$  is invertible, the following optimality conditions are obtained:

$$\frac{dL}{d\mathbf{w}} = \mathbf{C} \mathbf{w} - \lambda \mathbf{1} = 0 \rightarrow \mathbf{w} = \lambda \mathbf{C}^{-1} \mathbf{1} \quad (51)$$

$$\frac{dL}{d\lambda} = \mathbf{w}^T \mathbf{1} - 1 = 0 \rightarrow \mathbf{w}^T \mathbf{1} = 1 \quad (52)$$

The set of optimal weights  $\mathbf{w}_{opt}$  can be found by solving for  $\lambda$  and substituting:

$$1 = \mathbf{w}^T \mathbf{1} = (\lambda \mathbf{C}^{-1} \mathbf{1})^T \mathbf{1} = \lambda \mathbf{1}^T \mathbf{C}^{-1} \mathbf{1} \rightarrow \lambda = \frac{1}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} \quad (53)$$

$$\mathbf{w}_{opt} = \lambda \mathbf{C}^{-1} \mathbf{1} = \frac{1}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} \mathbf{C}^{-1} \mathbf{1} \quad (54)$$

The variance of the optimal portfolio is therefore:

$$\sigma_p^2 = \mathbf{w}_{opt}^T \mathbf{C} \mathbf{w}_{opt} = \frac{1}{(\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1})^2} (\mathbf{C}^{-1} \mathbf{1})^T \mathbf{C} (\mathbf{C}^{-1} \mathbf{1}) = \frac{1}{(\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1})^2} \mathbf{1}^T \mathbf{C}^{-1} \mathbf{1} = \frac{1}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} \quad (55)$$

Note that the above optimization problems doesn't put a constraint on the individual weights  $w_i$  which can be positive or negative. The negative number means that shorting is allowed, which means that it is possible to borrow a stock and sell it, which effectively increases the cash value of the portfolio which can be used to buy additional stocks.

### 3.1.2 Static Minimum-Variance Portfolio

This section will compute the minimum-variance portfolio using the returns of 10 stocks from 2017 to 2018, and test its performance by applying those optimal weights on the same stocks from 2018 to 2019. The figures below plot the results of this portfolio, showing the variance and the cumulative return. This is also compared to the equal weights portfolio where  $w_i = \frac{1}{10}$  for  $i = 1, 2, \dots, 10$ .

#### 3.1.2.a Equal Weights Portfolio

The figure below shows the returns and the cumulative returns of the equal-weights portfolio for the period 2018-2019. The performance of this portfolio is summarized in the table below. The performance of this portfolio is undesirable both in terms of returns and volatility.

Mean Return	-0.000473
Cumulative Return	-0.123503
Variance	0.000079
Sharpe	-5.964545

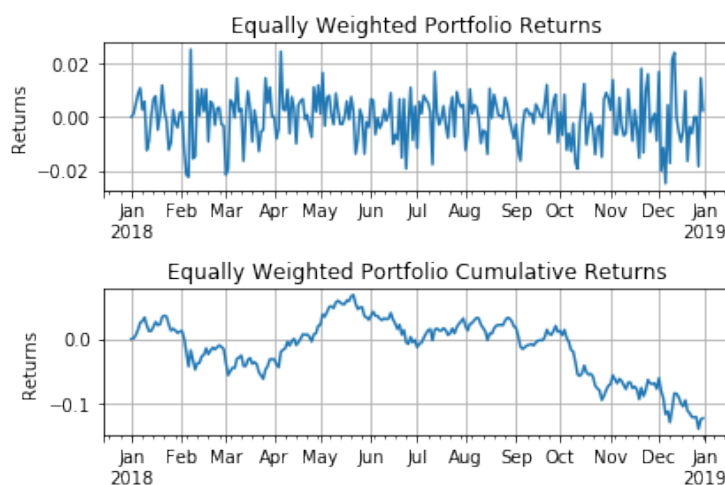


Fig. 23 *Equal Weights Portfolio (Test Period)*

#### 3.1.2.b Static Minimum-Variance Portfolio

The figure below shows the minimum variance portfolio using weights computed for the period 2017 to 2018. These weights are first applied to the training period, which shows very small variance as expected which is shown in the table below. Overall, the return is high for this portfolio, which suggests a premium associated with low volatility stocks. Finally, the theoretical variance of the portfolio matches the real one exactly, which is not surprising since the optimal weights are applied to the training data.

Mean Return	0.000958
Cumulative Return	0.249062
Variance	0.000029
Theoretical Variance	0.000029
Sharpe	33.475039

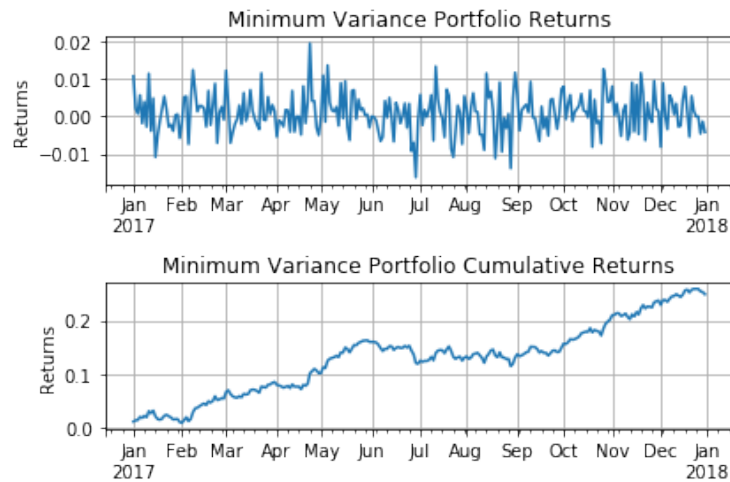


Fig. 24 *Minimum-Variance Portfolio (Train Period)*

However, it is not realistic to apply the weights to the period in which the optimal weights are computed from, for the simple reason that this would require future data to be included in the calculation.

The figure below shows the performance of a portfolio which uses the same weights from before applied to the following year from 2018 to 2019. As shown in the table below, the performance is worse compared to both the equal weights portfolio and the optimal weights portfolio in the training period. The worsening performance in terms of volatility suggests that the variance might have some mean-reversion characteristics. This suggests the use of an adaptive method that continuously updates the optimal weights.

Mean Return	-0.000451
Cumulative Return	-0.117582
Variance	0.000082
Sharpe	-5.499329

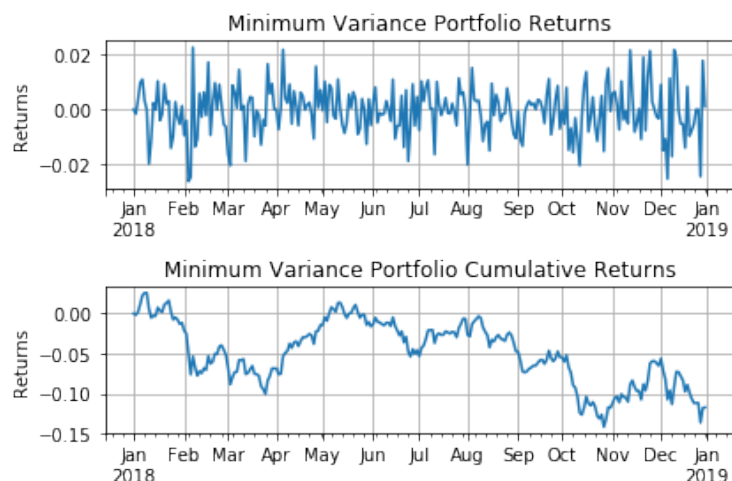


Fig. 25 Minimum-Variance Portfolio (Test Period)

### 3.1.3 Adaptive Minimum-Variance Portfolio

This section uses the minimum-variance optimization method as discussed in previous sections in a recursive manner, updating the weights day by day using a rolling window of length  $M$ . The figure and the table below shows the performance of the adaptive minimum-variance portfolio for 252 days (roughly 1 trading year) as well as the changing weights for each of the stocks.

Overall, the adaptive portfolio shows satisfactory performance in terms of variance, which is lower than both the static minimum-variance portfolio and the equal weights one. However, the cumulative return of the portfolio is not ideal, although this is expected since the optimization did not include any return forecasting.

Mean Return	-0.000436
Cumulative Return	-0.113804
Variance	0.000066
Sharpe	-6.572814

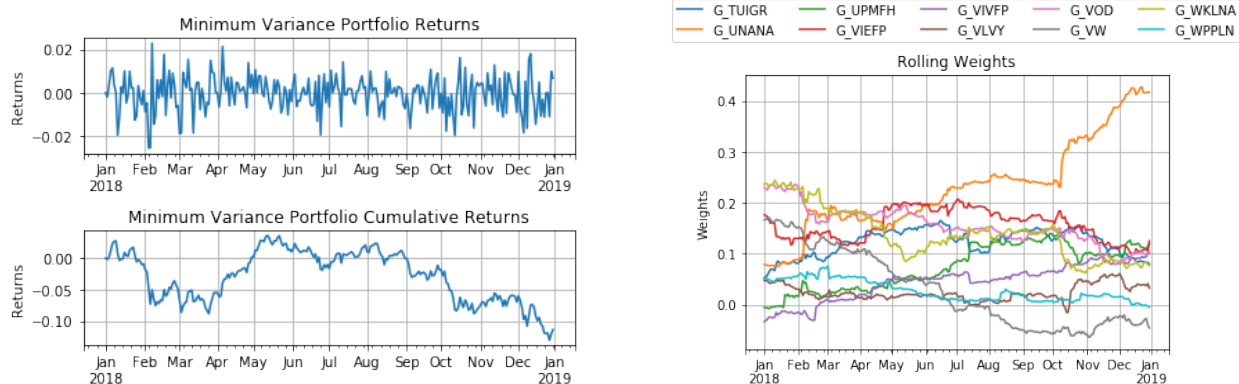


Fig. 26 Adaptive Minimum-Variance Portfolio

Finally, the performance of this adaptive portfolio is highly dependent on the estimation window. The figure below shows the cumulative return and the variance of the portfolio using different rolling windows, from 1 month to 12 months.

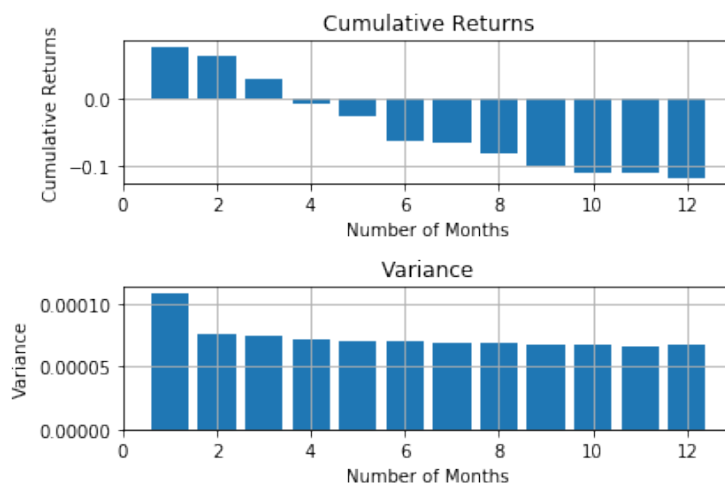


Fig. 27 Estimation Window vs Performance

Overall, the adaptive portfolio performs better than both the static minimum variance portfolio and the equal weights portfolio in terms of minimizing the variance, obtaining an average and median variance of 0.000073 and 0.000069 for estimation windows of 1 to 12 months. Not surprisingly, the performance in terms of returns isn't great for the adaptive portfolio, which produced negative results most of the time. This however is to be expected as the optimization didn't take into account the forecasted returns but only variance. In addition, repeatedly changing the weights of the portfolio incurs high transaction costs that worsens the returns further in a practical scenario.

Finally, the covariance matrix is estimated by using the returns of the past  $M$  day in this example, but there are other methods of estimating it. For instance, it is possible to change the weights for each day in an exponentially weighted manner such that the recent days have a stronger influence on the estimation of the covariance than the older days, although this method introduces additional tunable parameters (exponential weight decay).



## 4 Robust Statistics and Non-Linear Methods

### 4.1 Data Import and Exploratory Data Analysis

#### 4.1.1 Descriptive Statistics

This section will explore the use of different descriptive statistics on 3 stocks (AAPL, IBM, JPM) and 1 index (DJI). Descriptive statistics are summary statistics that are quantitative measures that summarize a collection of data. In particular, the descriptive statistics explored in this section are mean, median (measures of central tendency), standard deviation, median absolute deviation, interquartile range (measure of dispersion), skew (measure of asymmetry) and kurtosis (measure of distribution tails). Each of the statistics has different characteristics and behave differently in the presence of outliers, which will be explored in detail in the following sections.

Overall, the measures of central tendency (mean and median) tend to be similar for all assets, while the measures of dispersion (standard deviation and median absolute deviation) can have very different estimates, which could result in drastic performance differences (section 4.4.).

##### AAPL

	Open	High	Low	Close	Adj Close	Volume
Mean	187.687	189.562	185.824	187.712	186.174	3.27048e+07
Median	186.29	187.4	184.94	186.12	184.352	2.9184e+07
StdDev	22.1456	22.2816	22.0088	22.1607	21.9047	1.41797e+07
MAD	15.89	15.61	15.92	15.94	15.476	7.5739e+06
IQR	36	36.34	36.06	36.755	35.6854	1.63117e+07
Skew	0.259917	0.300385	0.220489	0.263849	0.29077	1.74332
Kurtosis	-0.912594	-0.924602	-0.917632	-0.932425	-0.928017	4.35318

##### IBM

	Open	High	Low	Close	Adj Close	Volume
Mean	138.454	139.492	137.329	138.363	134.903	5.19894e+06
Median	142.81	143.99	142.06	142.71	138.566	4.2379e+06
StdDev	12.1143	11.9131	12.2046	12.0281	10.6716	3.32896e+06
MAD	5.27	5.31	5.19	5.23001	4.49352	920700
IQR	15.38	14.72	16.34	15.505	14.1039	1.95295e+06
Skew	-0.676024	-0.622707	-0.713446	-0.682246	-0.811222	3.1929
Kurtosis	-0.585272	-0.623607	-0.561975	-0.584037	-0.420852	11.7969

##### JPM

	Open	High	Low	Close	Adj Close	Volume
Mean	108.708	109.652	107.683	108.607	107.263	1.47007e+07
Median	109.18	110.53	107.79	109.02	107.219	1.3633e+07
StdDev	5.35908	5.20287	5.43254	5.30048	4.83332	5.34977e+06
MAD	4.47	4.31	4.24	4.35001	3.45016	3.0354e+06
IQR	8.81001	8.845	8.845	8.835	7.22244	6.2336e+06
Skew	-0.420811	-0.376221	-0.377517	-0.374853	-0.344491	1.69346
Kurtosis	-0.322536	-0.544163	-0.2657	-0.396579	-0.105437	4.4302

##### DJI

	Open	High	Low	Close	Adj Close	Volume
Mean	25001.3	25142	24846	24999.2	24999.2	3.32889e+08
Median	25025.6	25124.1	24883	25044.3	25044.3	3.1379e+08
StdDev	858.835	815.204	903.302	859.132	859.132	9.4078e+07
MAD	543.541	537.619	601.568	590.721	590.721	5.046e+07
IQR	1109.43	1077.82	1204.42	1158.16	1158.16	1.0893e+08
Skew	-0.372127	-0.239367	-0.456447	-0.380147	-0.380147	1.73956
Kurtosis	0.485736	0.118153	0.557592	0.400668	0.400668	5.85758

### 4.1.2 Histograms and PDF

The figure below shows the empirical probability density function (red) and the histogram of the distributions (blue) of the adjusted closing prices and of 1-day returns for each of the tickers mentioned previously. As discussed in the part 1 of the coursework, the returns tend to follow a more normal distribution than the raw prices, which can be seen in the PDFs below. This means that most of the classical statistics which assumes an underlying Gaussian distribution work better with returns rather than prices. This suggests the use of statistics such as median and median absolute deviation when working with prices instead of mean and standard deviation.

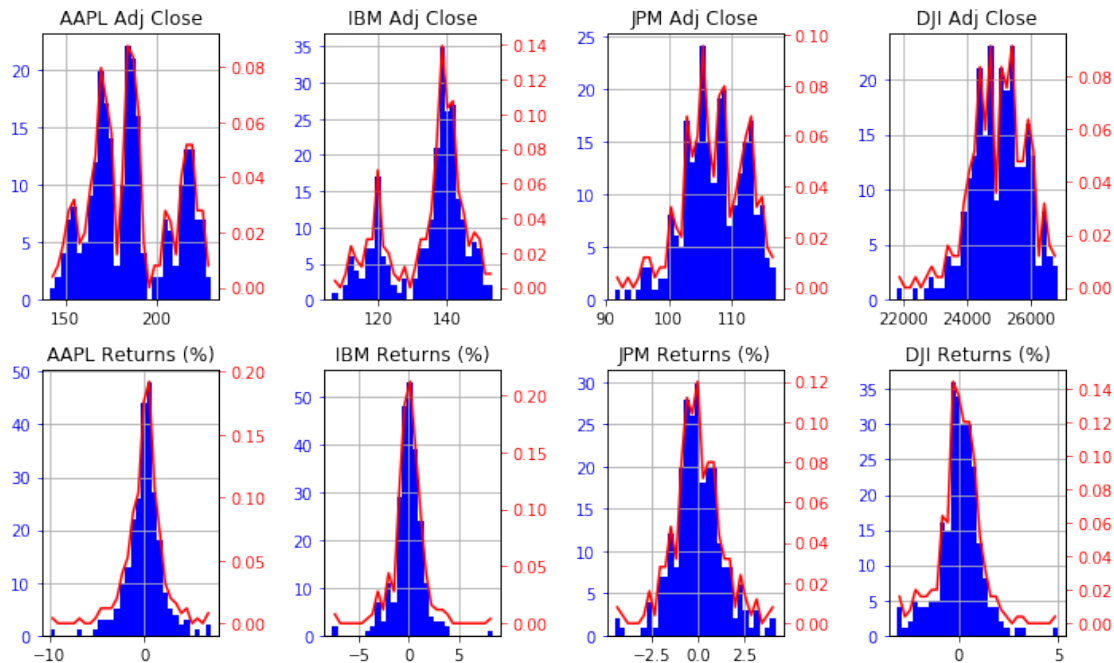


Fig. 28 Histograms and PDFs

### 4.1.3 Mean vs Median Z-Score Outlier Detection

This section uses 2 Z-score based methods for outlier detection. The first one classifies all prices levels outside of the rolling price mean  $\pm 1.5 \times$  standard deviations as outliers, while the second one uses the rolling price median  $\pm 1.5 \times$  median absolute deviation.

As shown in the figures below, the blue area that denotes the Z-score range changes more smoothly and is less tight for method 1 than method 2. This is because the mean based estimators are affected more by the presence of outliers than median based estimators, which makes the mean based estimators more susceptible to changes. For the same reason, the number of outliers detected in method 1 is much less than method 2 since the measure of deviation is impacted more in method 1, which is shown in the table below. This demonstrates the robustness of median as estimator compared to the mean in the sense that it is less effected by the contaminated data.

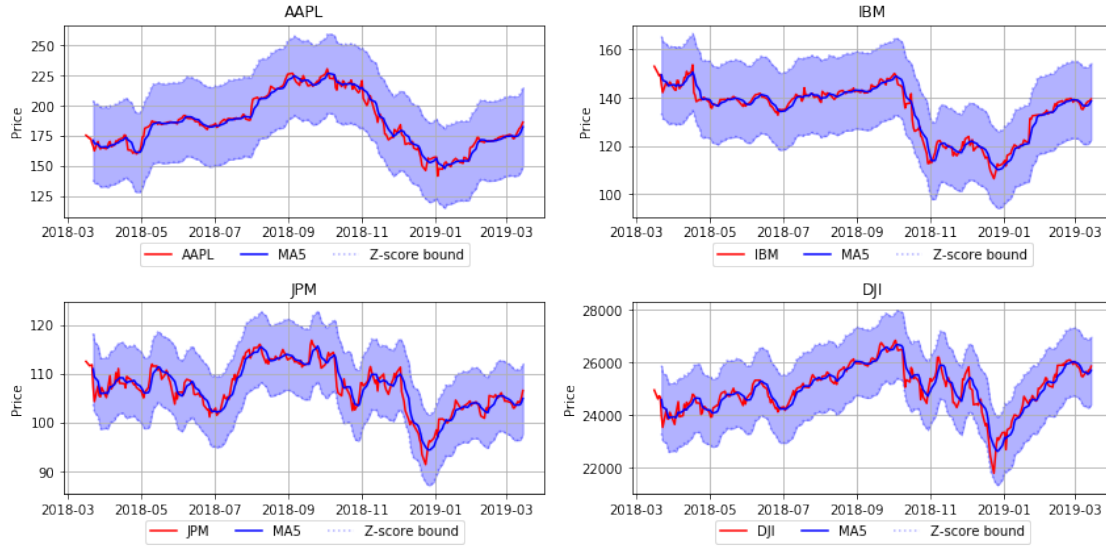


Fig. 29 Moving Average Outlier Detection

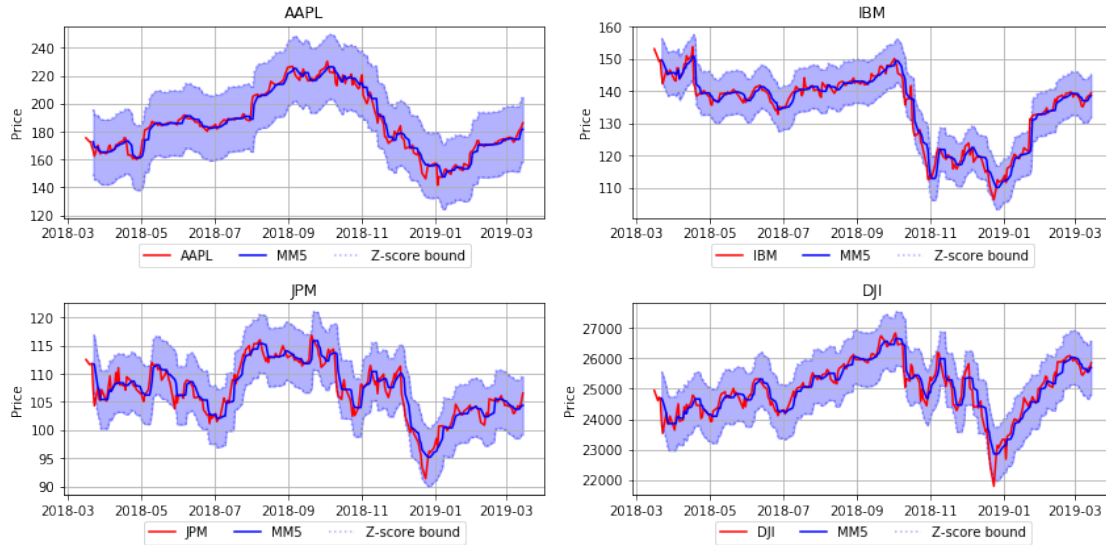


Fig. 30 Moving Median Outlier Detection

In reality, it is impossible to use the above method since the computation of standard deviation and median absolute deviation requires the knowledge of the future data. An alternative approach would be to calculate the z-score using the rolling estimation of these values, which is shown in the figure below. Although the computation is on a rolling basis, the same arguments apply and the results are similar. Overall, the rolling method detects more outliers than the static method, which is expected since the price range tends to be bounded within a small area in the short term window.

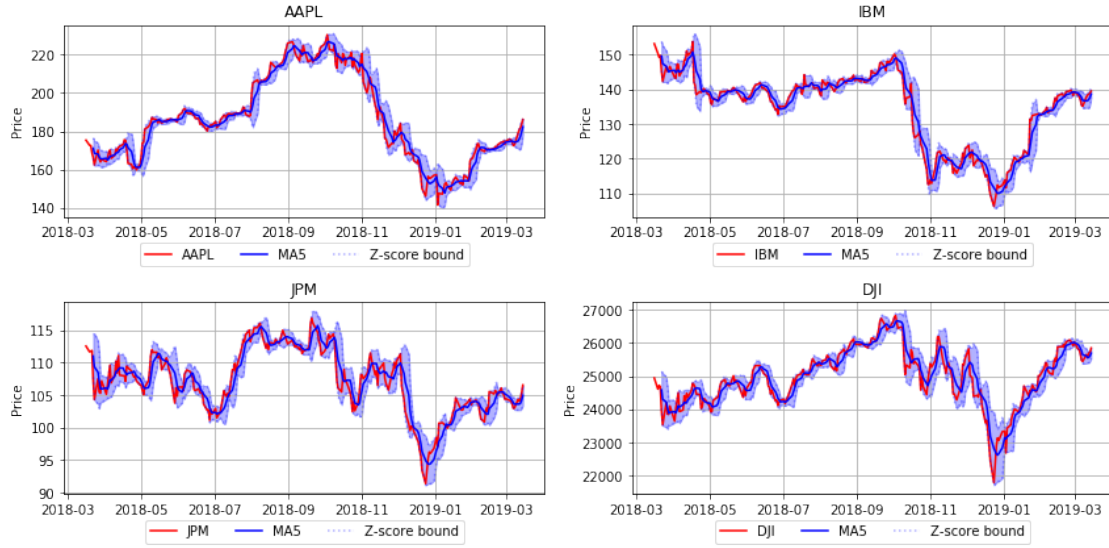


Fig. 31 Moving Average Outlier Detection (Rolling)

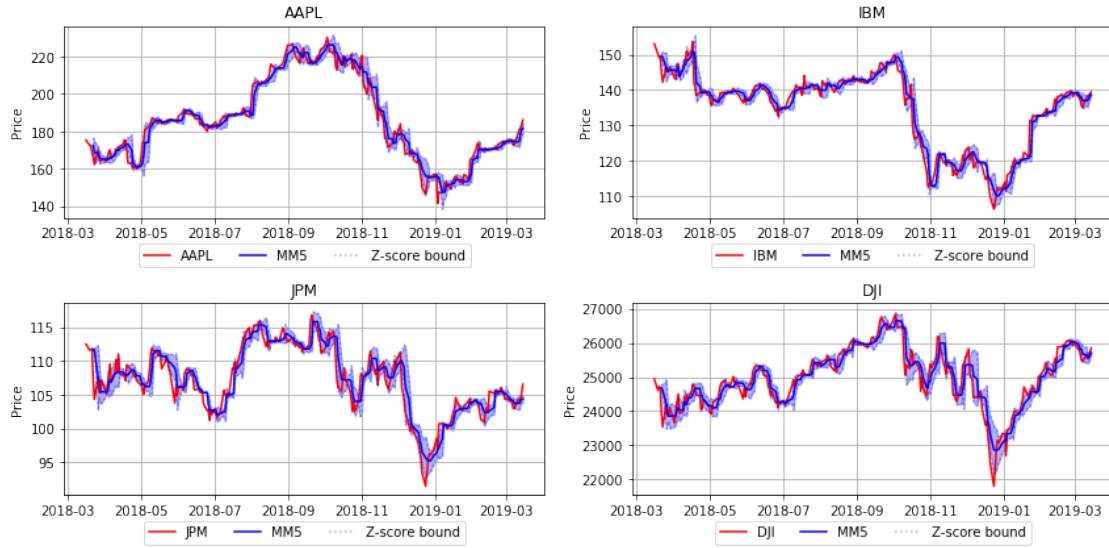


Fig. 32 Moving Median Outlier Detection (Rolling)

	Mean Method Outliers Count	Median Method Outliers Count	Mean Method (Rolling) Outliers Count	Median Method (Rolling) Outliers Count
AAPL	0	0	30	103
IBM	0	8	31	94
JPM	0	3	33	105
DJI	0	3	30	97

#### 4.1.4 Impact of Outliers

For this section, outliers are manufactured and introduced to four different dates. The outliers have the value equivalent to 1.2 times the all time maximum for each of the assets. Due to the same reasons discussed before, it is not surprising to see in the figures below that the both the Z-score range and the mean is highly affected by the presence of outliers in method 1, while the second method didn't change at all.

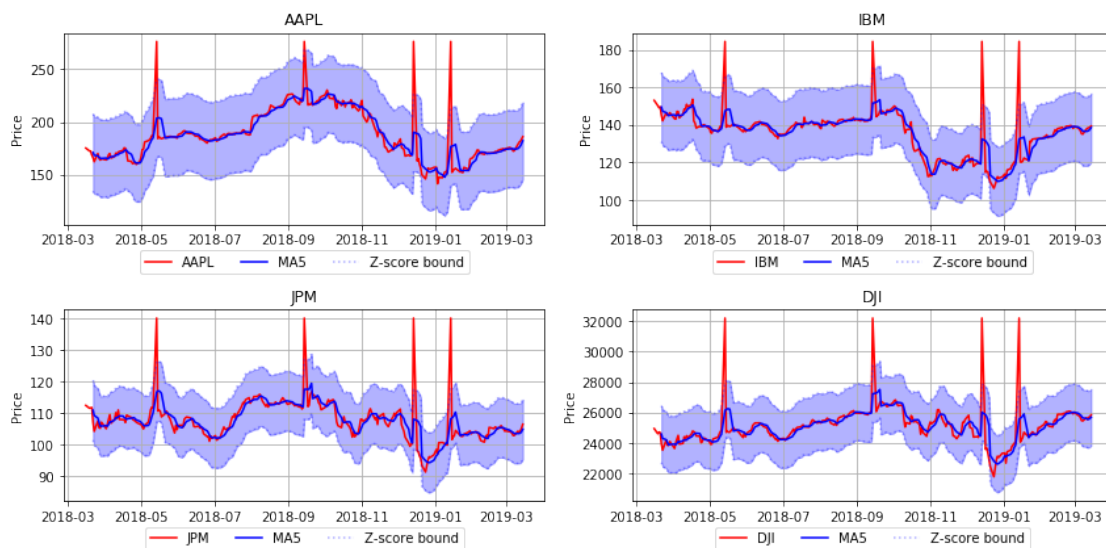


Fig. 33 Impact of Artificial Outliers (Moving Average)

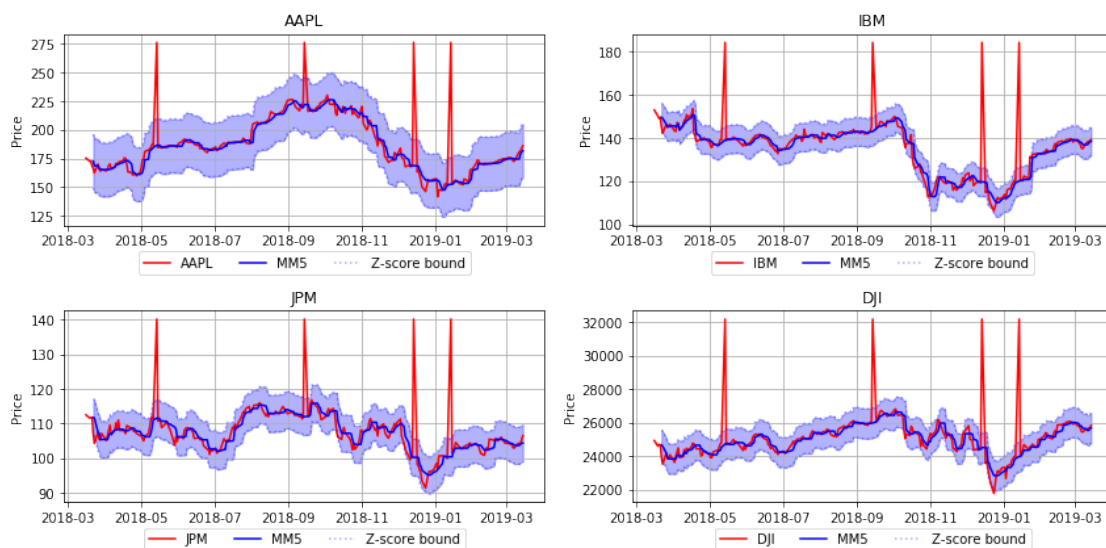


Fig. 34 Impact of Artificial Outliers (Moving Median)

The figures below show the same performance but applied to the rolling method, which further confirms the robustness of the median method.

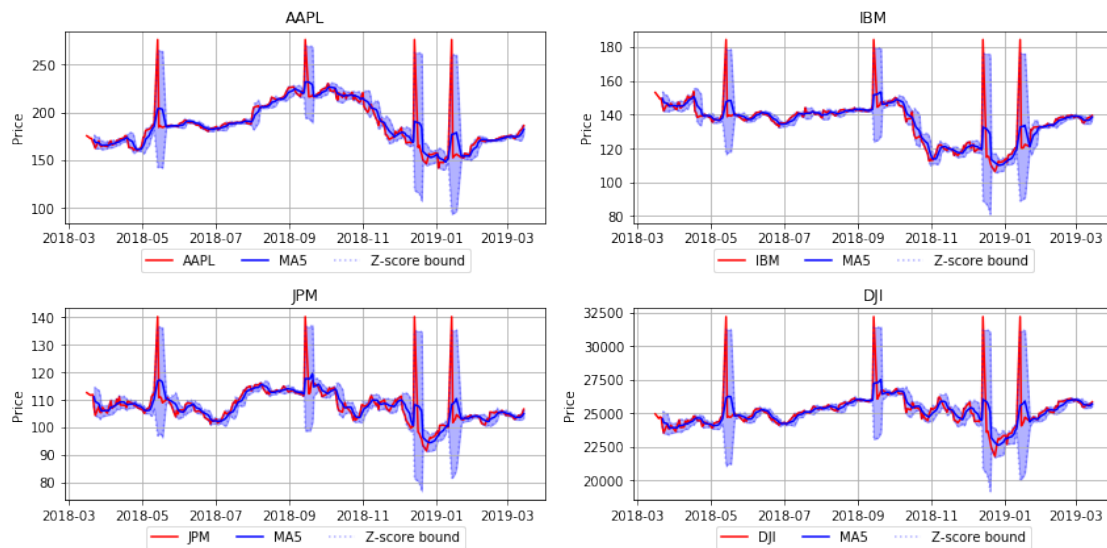


Fig. 35 Impact of Artificial Outliers (Moving Average, Rolling)

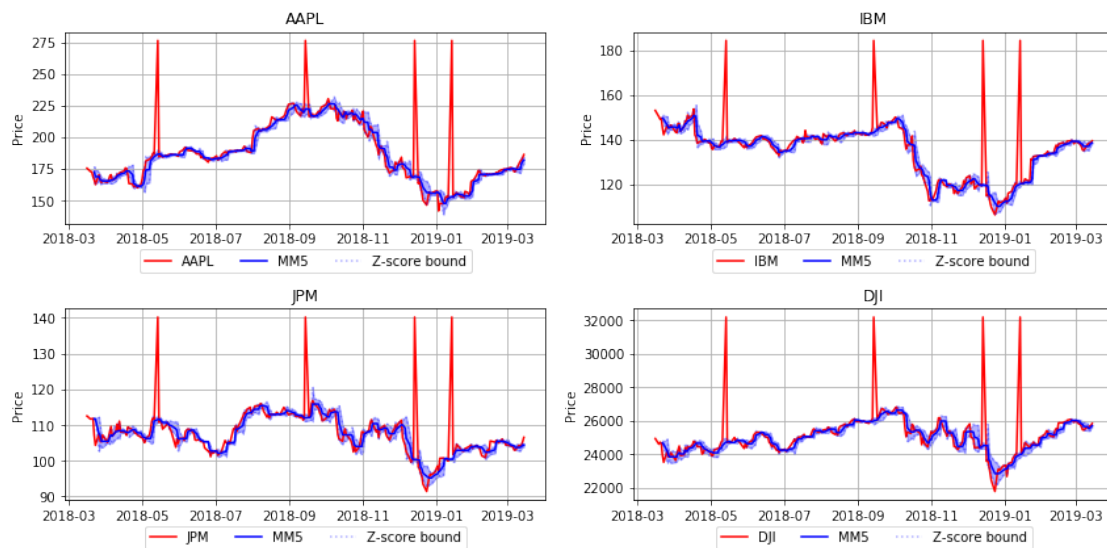


Fig. 36 Impact of Artificial Outliers (Moving Median, Rolling)

#### 4.1.5 Box Plots

Another useful tool for analyzing the data without making parametric assumptions about the distribution is to plot the box plots, which are plotted below for each of the assets. The box plot summarizes the data graphically by analyzing the dispersion of the data, where the orange line denotes the median of the data, the length of the box represents the interquartile range IQR (from Q1 where 25% of the data can be found to Q3 where 75% of the data can be found), and the whiskers extends to the total range of the data. Finally, the outliers are plotted as points, which are defined as points which are outside  $1.5 \times$  the IQR.

Overall, for the following box plots, it can be seen that the data tends to be skewed with asymmetric quartiles and has multiple outlier points. This means that the classical gaussian models are unfit to describe the price distributions.

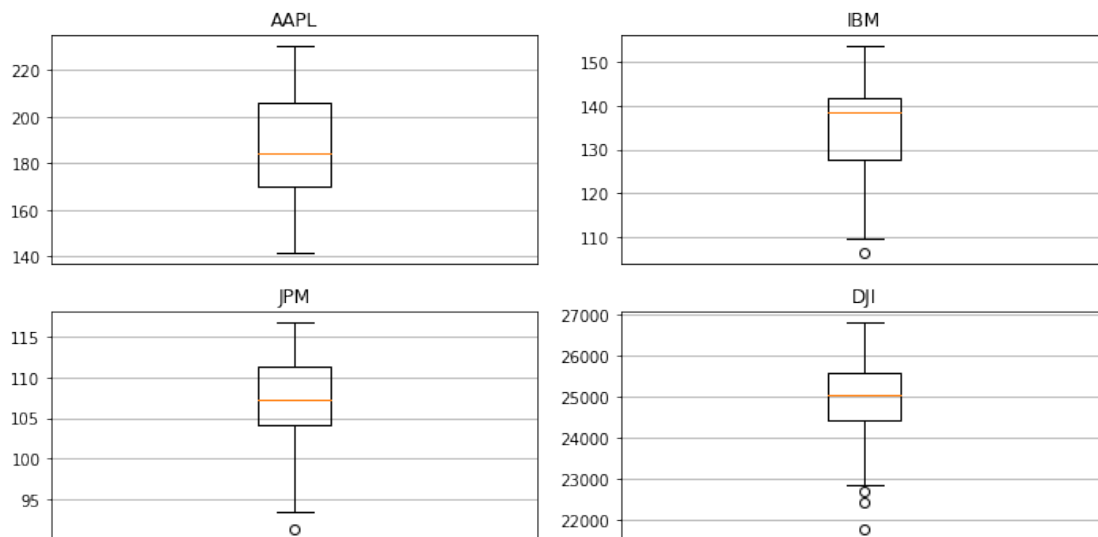


Fig. 37 Box Plots

## 4.2 Robust Estimators

### 4.2.1 Custom Robust Estimators

This section explores the computational complexity of robust estimators such as median, interquartile range (IQR) and the median absolute deviation (MAD). These robust estimators are non-linear methods since operations such as sorting is needed. The following custom function performs the calculation of the mentioned estimators:

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

def CustomIQR (ts):
    tmp = ts.sort_values() # sorting
    p25 = tmp[round(len(tmp)/4)] # find the 25th percentile value
    p75 = tmp[round(len(tmp)/4*3)] # find the 75th percentile value
    iqr = p75-p25
    return iqr

def CustomMAD (ts):
    median = CustomMedian (ts) # compute the median first
    deviations = abs(ts - median) # compute the deviations from the median
    mad = CustomMedian(deviations) # compute the median of deviations
    return mad
```

The logic behind each of the functions above are as follows:

- Median:  
The given array is sorted in terms of value first, then the middle number is extracted as the median.
- IQR:  
The values are sorted, then the 25th and the 75th percentile numbers are found (values below which 25% and 75% of the data can be found respectively). Then, the interquartile range is found as the absolute difference between the two values.
- MAD:  
For the median absolute deviation, the median is found first. Then, the absolute deviation from the median is computed for all the numbers available. Finally, the MAD is computed as the median of all absolute deviations.

### 4.2.2 Complexity Analysis

To find the median, the array must be sorted first, which has complexity  $O(n\log(n))$  using merge sort algorithm where  $n$  is the size of the array. Then, the middle value is simply the value with index  $\text{int}(\frac{n}{2})$ , which takes complexity  $O(1)$ .



For the IQR, there is first a sorting operation of complexity  $O(n \log(n))$ , then it finds the values with index  $\text{int}(\frac{n}{4})$  and  $\text{int}(3\frac{n}{4})$  respectively and computes their differences, which are 3 operations of complexity  $O(1)$ . This is hence slightly more expensive than the median estimator.

Finally, for the MAD, it needs to find the median first which has complexity  $O(n \log(n))$ . Then, it computes the deviation from the median for all  $n$  data which has complexity  $O(n)$ . Finally, the median of all deviations is found to be the MAD which is another operation of complexity  $O(n \log(n))$ . Overall, MAD is the estimator with the highest computational cost.

The time required for the robust estimators mentioned above is much more than the classic methods such as mean and standard deviation, which have complexity of  $O(n)$ .

### 4.2.3 Breakdown Points

The breakdown point of an estimator measures its robustness. It measures a level above which the estimator can arbitrarily deteriorate if the sample data were contaminated. Formally, the finite sample breakdown point is the fraction of the data that can be contaminated without making the estimator bad for a sample of data of size  $n$ . The breakdown point instead refers usually to the asymptotic breakdown point, which is the finite sample breakdown point as  $n$  tends to infinity.

Since the median estimator is the middle value of a sorted array, up to  $\frac{(0.5n-1)}{n}$  can have arbitrary values without affecting the estimate, which makes the breakdown point of the median 0.5 as  $n$  approaches infinity. Similarly, the IQR estimator has a breakdown point of 0.25 since it splits the data array into 4 parts instead of 2 parts (the case for the median). Finally, the MAD has the same breakdown point as the median of 0.5, since it computes the median and then finds the median of absolute deviation from the median.

### 4.3 Robust and OLS Regression

#### 4.3.1 OLS Regression

This section will compare the ordinary least squares (OLS) regression method against a more robust method such as Huber regression by regressing the returns of each stock against the DJI index. Overall, the regression problem has the following set up:

$$\mathbf{r} = \mathbf{X}\mathbf{b} + \mathbf{e} \quad (56)$$

where  $\mathbf{r}$  is the returns vector containing the returns of the stock,  $\mathbf{X}$  is the matrix with a column of ones and a column of DJI returns, and  $\mathbf{e}$  is the error vector. The OLS method aims to minimize the squared error  $\|\mathbf{e}\|^2 = \|\mathbf{r} - \mathbf{X}\mathbf{b}\|^2$  and has solution  $\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{r}$ . The table and the figure below show the results of the OLS regression, as well as the given parameters.

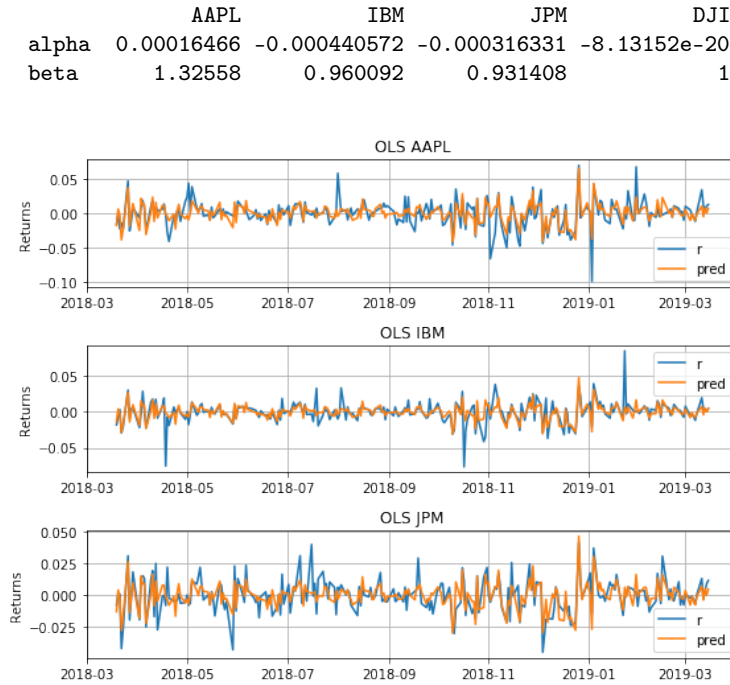


Fig. 38 OLS Estimation

#### 4.3.2 Robust Regression (Huber Regression)

The OLS regression method aims to minimize the squared errors which is a measure of squared deviations. Therefore, this regression method is highly sensitive to outliers in the data for reasons similar to why standard deviation is a non-robust estimate. Therefore, the OLS model assumes that there is a constant underlying distribution with constant variance, and its performance degrades under heteroscedasticity or in presence of outliers.

The Huber regression instead aims to provide a more robust version of OLS method by minimizing the squared error for  $\frac{\|\mathbf{r} - \mathbf{X}\mathbf{b}\|}{\sigma} < \epsilon$  and the absolute error for  $\frac{\|\mathbf{r} - \mathbf{X}\mathbf{b}\|}{\sigma} > \epsilon$ . By making the error

asymmetric such that it is quadratic for small values but linear for bigger ones, it effectively reduces the impact of outliers on the estimation of the parameters. Therefore, by changing the value of  $\epsilon$ , it is possible to control the point at which the errors are considered outliers. The table and the figure below shows the results of this regression and the estimated parameters respectively.

	AAPL	IBM	JPM	DJI
alpha	-0.000130371	-0.000509435	-0.000800961	-2.47219e-11
beta	1.27021	0.973562	0.919662	1



Fig. 39 *Huber Estimation*

### 4.3.3 Regression and Outliers

Overall, the two regression methods produced fairly similar estimates of the coefficients as shown in the tables below. It is possible to quantify the impact of outliers by using the same method of section 4.1.4., where artificial outliers with values equivalent to the 1.2x the all time maximum are introduced. As shown in the table below, the new parameters of the OLS method changed by 81% from the previous values while the Huber method only changed by 54%, which empirically demonstrates the robustness of Huber regression.

	AAPL	IBM	JPM	DJI
alpha	0.000350419	-6.28721e-05	-0.000467133	2.1684e-19
beta	1.35047	1.22204	0.906468	1

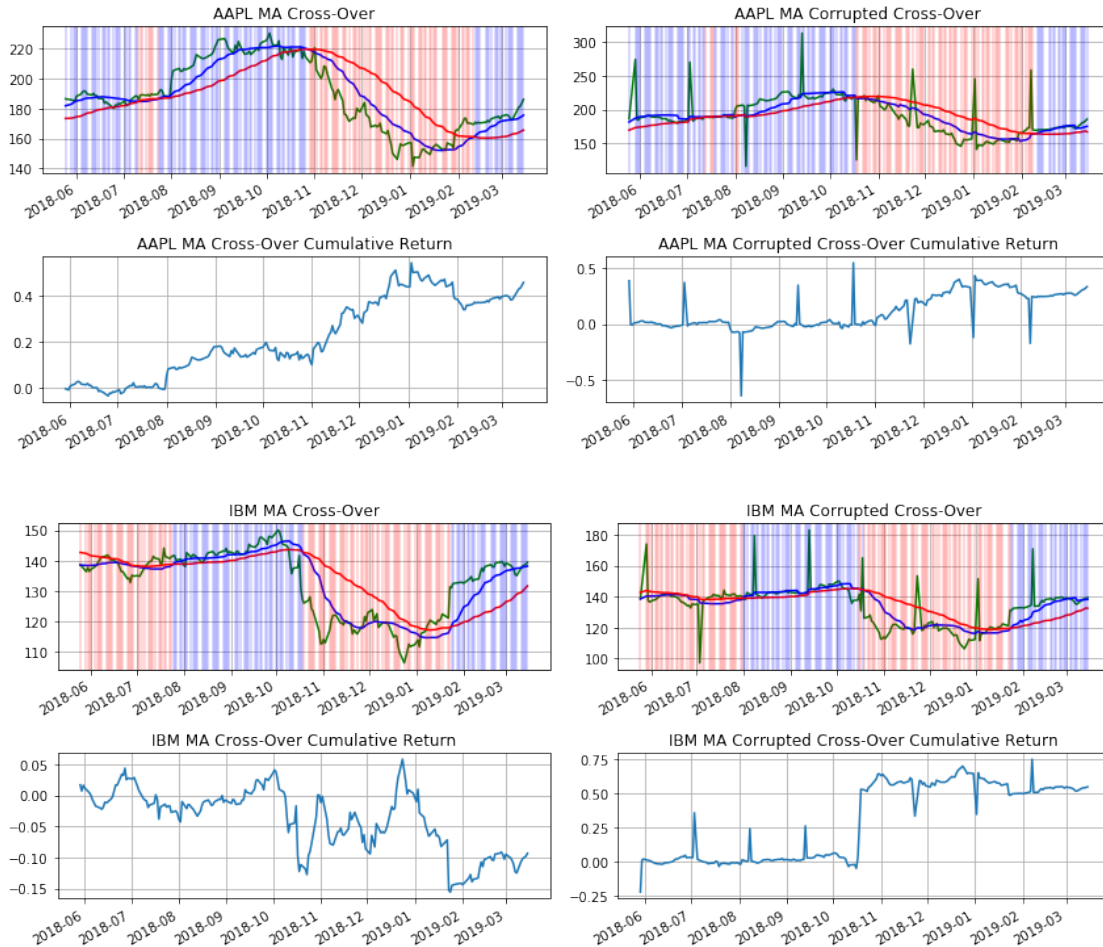
	AAPL	IBM	JPM	DJI
alpha	0.000122268	-0.00033875	-0.000933215	1.72253e-11
beta	1.33077	1.06664	0.890359	1

## 4.4 Robust Trading Strategies

### 4.4.1 Simple Moving Average-Cross Over

This section explores the moving average cross-over strategy which decides to buy if the 20-day moving average is above the 50-day moving average and to sell if it is below. The performance of this strategy is plotted in the figures below for each of the assets, where the left figures show the standard performance while the right ones show the same time-series but corrupted with spikes in random directions. Following the discussion of previous sections, it is not surprising to see that the cross-over regions (blue for long and red for short) are affected heavily by the presence of outliers, which led to significant change in performance. The amount of overlap in long-short decisions between the 2 cases are listed in the table below.

AAPL	region overlap (%): 92.08
IBM	region overlap (%): 96.53
JPM	region overlap (%): 84.65
DJI	region overlap (%): 96.53



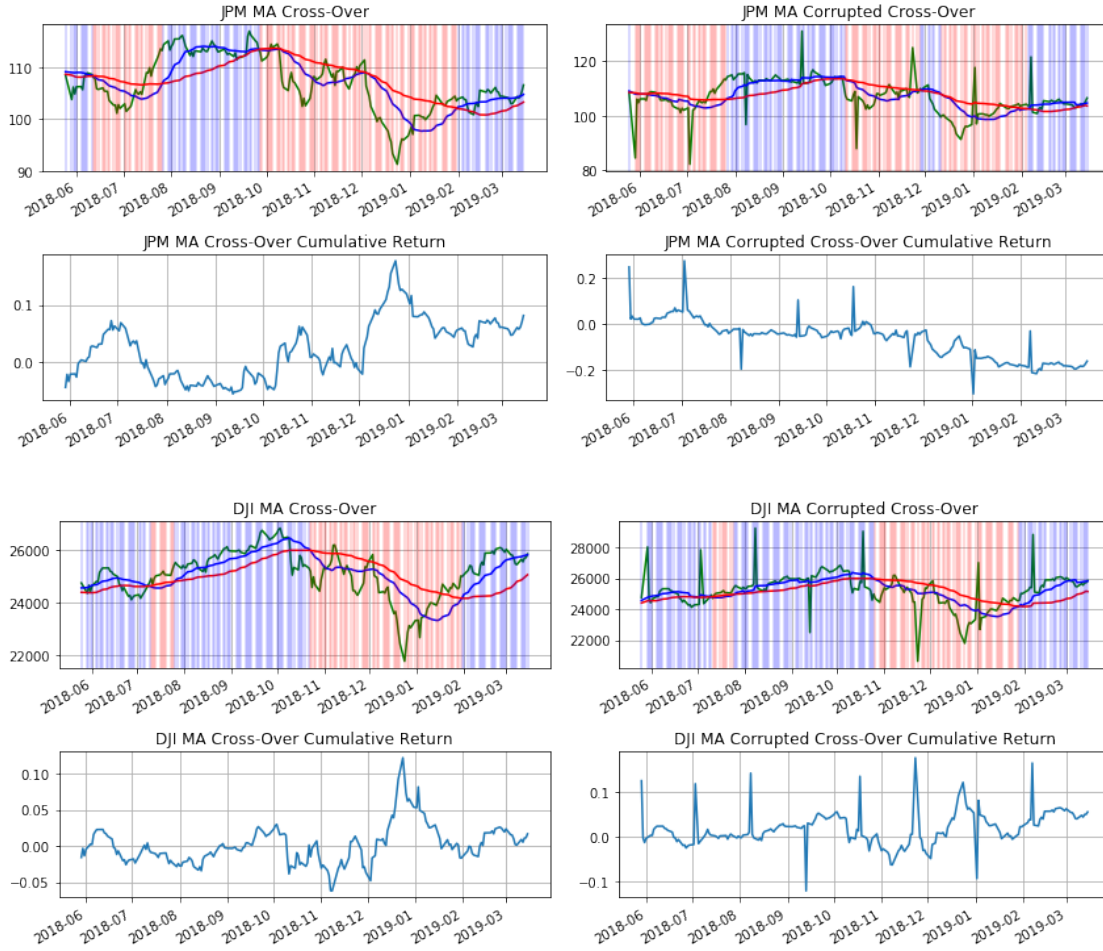
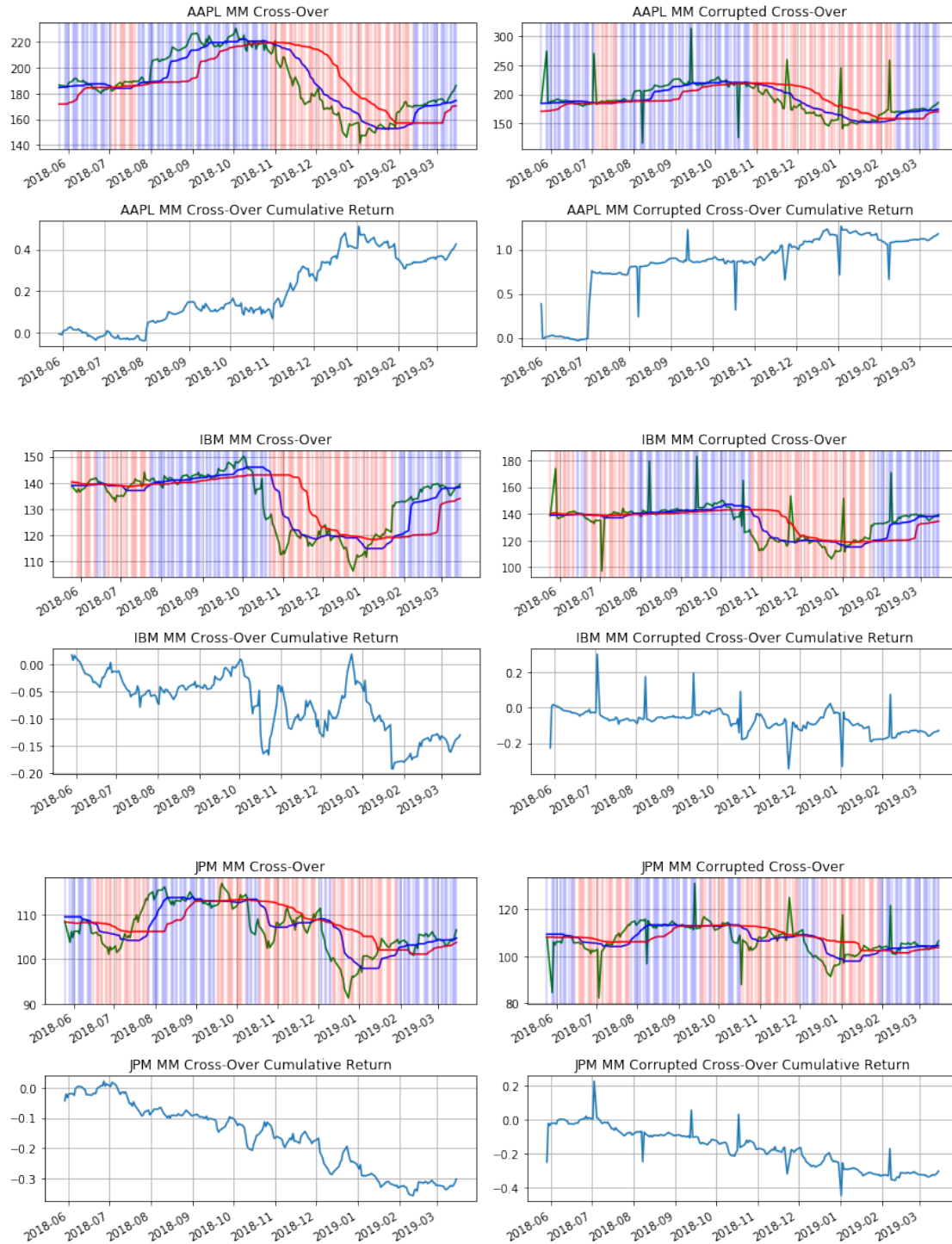


Fig. 40 Simple Moving Average Cross-Over Strategy

#### 4.4.2 Robust Moving Median Cross Over

This section explores the moving median cross-over strategy instead using the same windows as the previous section and the same logic. The performance of this strategy is plotted in the figures below for each of the assets, where the original and the disturbed times serieses are the same as the ones in the moving average strategy to make the comparison sensible. As expected, the introduction of the outliers in this strategy makes very little impact due to the robustness of the median estimate. The overlap of long-short decision points for the corrupted and non-corrupted scenarios average about 98%, which is higher than the 92% overlap in the moving average case.

AAPL	region overlap (%):	99.01
IBM	region overlap (%):	98.02
JPM	region overlap (%):	96.53
DJI	region overlap (%):	98.51



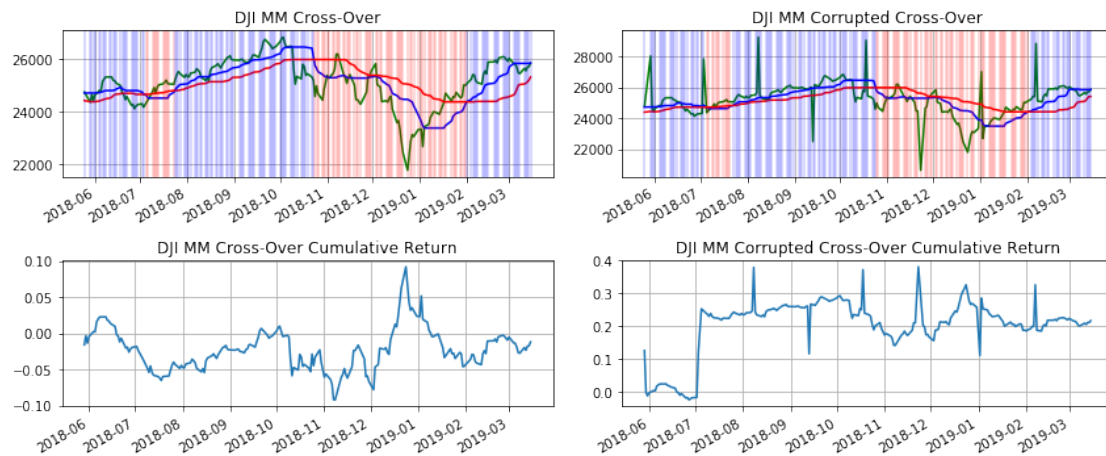


Fig. 41 Simple Moving Median Cross-Over Strategy

## 5 Graphs in Finance

### 5.1 S&P Stocks Selection

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

This section will explore the use of graphical methods to infer relationships among the given assets. Specifically, this section will analyse the relationship of the 10 stocks listed in the table above. These 10 stocks are the New York based financial companies which are biggest in their respective GICS Sub Industry by market capitalization.

This particular screening was chosen due to the intricate relationship that exists among the financial institutions, which are often affected by the same regulations and macroeconomic factors, such as change in interest rates or the yield curve inversion. In addition, by limiting the location to New York, all the geographical factors such as population and local laws are filtered out which will aid the ease of analysis. Finally, by excluding competitors from the same GICS sub-industries, the graphical method should highlight the underlying macro characteristics of the financial network instead of aspects relating to firm-wise competition. However, this is a generalization that will not hold true completely since there will be some overlap in the operations of these firms.



## 5.2 Correlation Based Graphs

	BLK	AXP	JPM	SPGI	MMC	MS	MET	L	AIG	C	Mean
BLK	1.00	0.52	0.72	0.65	0.64	0.73	0.64	0.66	0.55	0.72	0.68
AXP	0.52	1.00	0.58	0.46	0.47	0.58	0.50	0.50	0.43	0.57	0.56
JPM	0.72	0.58	1.00	0.61	0.63	0.87	0.75	0.70	0.64	0.88	0.74
SPGI	0.65	0.46	0.61	1.00	0.58	0.61	0.50	0.53	0.45	0.59	0.60
MMC	0.64	0.46	0.62	0.58	1.00	0.59	0.54	0.63	0.53	0.58	0.62
MS	0.73	0.57	0.86	0.61	0.59	1.00	0.74	0.68	0.61	0.85	0.73
MET	0.64	0.49	0.75	0.50	0.54	0.74	1.00	0.66	0.65	0.75	0.68
L	0.66	0.50	0.70	0.53	0.63	0.68	0.66	1.00	0.62	0.66	0.67
AIG	0.56	0.43	0.64	0.45	0.53	0.61	0.65	0.62	1.00	0.63	0.62
C	0.72	0.57	0.88	0.59	0.58	0.85	0.75	0.66	0.63	1.00	0.73

This section will explore the construction of a network based on the correlation matrix of the returns of the 10 stocks mentioned previously (table above). The network groups each node (representing the stocks) spatially in relation to its correlation value with every other node. In the network below specifically, the edges (connections) are drawn only for pairs of nodes with a correlation value greater than 0.5 in magnitude, and all the self-correlation of value 1 are omitted from the graph.

The correlation matrix can be interpreted as a network since it shows how each node is connected to one another. This is because the correlation of each pair represents the strength of connection that there is between the two points. This strength can then be in turn represented spatially by placing highly connected nodes closer together and loosely connected nodes further apart. This allows the formation and visualization of local groups that demonstrate some underlying relationship between the connections.

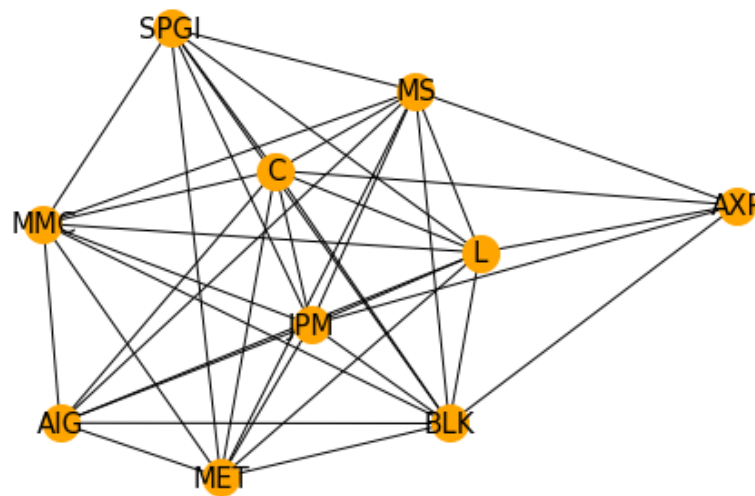


Fig. 42 Returns Correlation Graph

### 5.3 Correlation Graph Analysis

As shown in the network drawn in figure 42, it can be seen that most of the stocks are concentrated around a main group with JPM at its center, while there are two outliers AXP and SPGI which are placed slightly far away from the rest. This is reflective of their correlation value since JPM has the highest correlation with every other node on average (74%) compared to the rest of the stocks. Similarly, AXP and SPGI have the least amount of correlation (56% and 60% respectively).

This spatial relationship is also reflective of the business nature of these stocks. For instance, the main business line of the American Express Co. (AXP) is the deposit business which is centered around their debit and credit cards, and not any retail banking or investment banking activities such as the majority of the stocks in consideration. For this reason, it is placed far away from the central group, as it is relatively less correlated with the rest.

Similarly, the main business lines of S&P Global (SPGI) are data, analytics and rating provision, which are different business operations from banking and brokerage. However, since its services are used across multiple financial institutions, it's still connected with the remaining nodes.

For the opposite reason, the central position of banks with diversified operations such as JP Morgan Chase & Co. (JPM) is due to the diversified nature of its businesses. Since it operates in a variety of businesses such as asset management, brokerage, retail and investment banking and many more, they have branches that overlap in operation with most of the stocks present in the list, making them the assets with the return most correlated with the rest.

Finally, the correlation is a measure of similarity between realization of two random variables that is unaffected by the re-ordering of the values or the indices. Although the graph will not be the same every time, the overall arrangement and local grouping will remain constant for any type of re-shuffling as long as the operation is applied equally across all realizations. This is shown in the figure below where the network is draw for returns that are randomly re-shuffled in time.

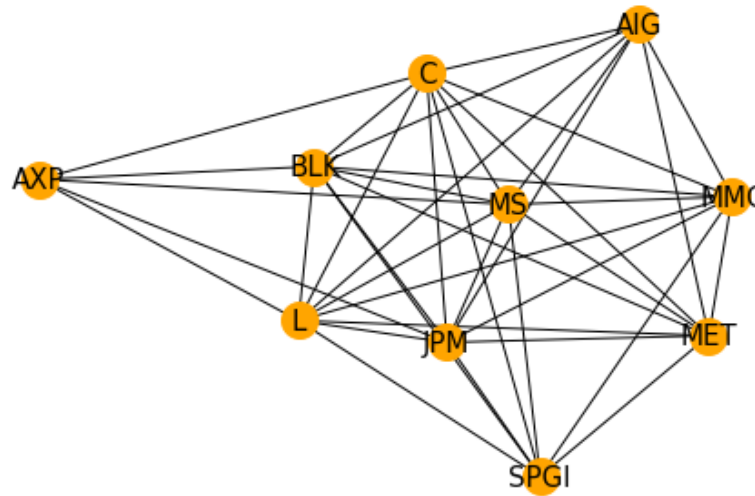


Fig. 43 Shuffled Returns Correlation Graph

## 5.4 Spectral Similarity Graphs

This section explores the construction of a network based on the power spectrum characteristics. Specifically, the power characteristics with associated frequencies are computed for all the given stocks using the Fourier transform. Then, the correlation is computed for each pair as a measure of similarity between the power spectrums. Both the spectral correlation matrix and the associated network are plotted below, where the edges are drawn only for pairs that have correlation greater than 0.3 and self-correlations are omitted.

The main justification for this measure of similarity is based on the fact that firms undergo different business and economic cycles that can affect the returns. Therefore, by using this metric, it is possible to construct a visual representation of which firms are affected by similar business or economic cycles.

	BLK	AXP	JPM	SPGI	MMC	MS	MET	L	AIG	C	Mean
BLK	1.00	0.14	0.46	0.40	0.41	0.51	0.33	0.42	0.25	0.49	0.44
AXP	0.14	1.00	0.31	0.12	0.15	0.26	0.18	0.21	0.19	0.27	0.29
JPM	0.46	0.30	1.00	0.35	0.40	0.71	0.51	0.44	0.41	0.81	0.54
SPGI	0.40	0.12	0.35	1.00	0.35	0.35	0.21	0.27	0.19	0.32	0.36
MMC	0.41	0.14	0.40	0.34	1.00	0.34	0.26	0.42	0.29	0.35	0.40
MS	0.51	0.26	0.72	0.35	0.34	1.00	0.52	0.40	0.38	0.72	0.52
MET	0.33	0.18	0.51	0.21	0.27	0.51	1.00	0.38	0.48	0.55	0.46
L	0.42	0.21	0.44	0.27	0.42	0.40	0.38	1.00	0.34	0.39	0.43
AIG	0.25	0.19	0.41	0.19	0.29	0.38	0.48	0.34	1.00	0.42	0.40
C	0.49	0.27	0.81	0.32	0.35	0.72	0.55	0.39	0.42	1.00	0.54

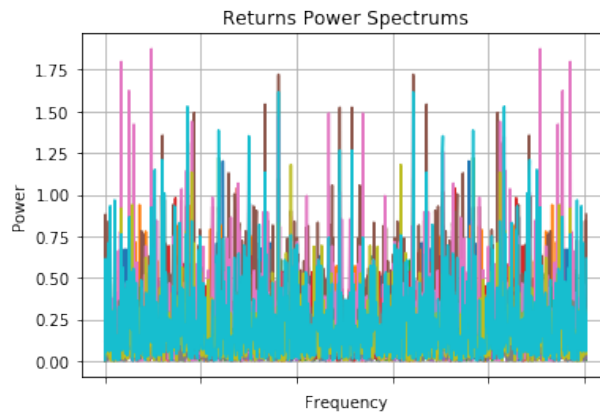


Fig. 44 Returns Power Spectrum

The figure below shows the network constructed using the power spectrum correlation matrix. Firstly, the American Express (EXP) is placed far away from the rest, which highlights the same conclusion about the nature of its business as discussed in the previous section.

Secondly and more interestingly, this graph shows a strong symmetry which highlights the relationship between different business cycles. For instance, the main groups that lay on the line of symmetry are JPM, MS and C, which are all banks that offer multiple services that overlap with other stocks, hence explaining their central positions as their business cycles are tied with the performance of every other stock.

On the right side of the line of symmetry there are AIG, MET, and L, which are all insurance companies, and hence subject to similar cycles. The insurance business cycles doesn't affect as much the firms operating in asset management, brokerage or data provision, which explains why BLK, MMC and SPGI are placed on the opposite side.

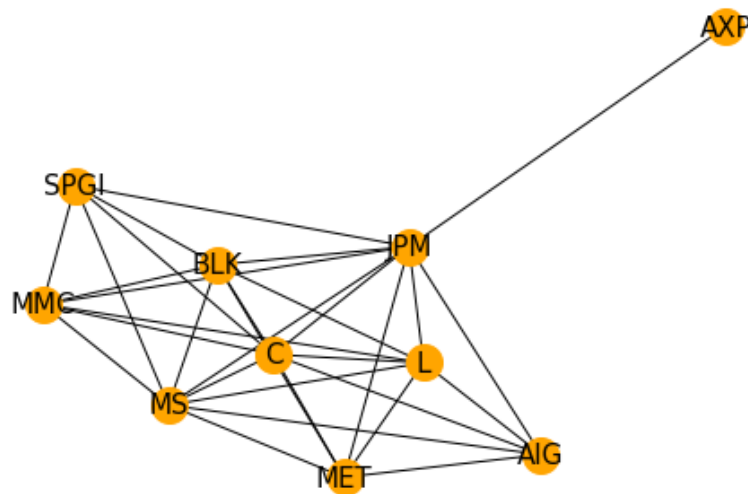


Fig. 45 Returns Spectral Correlation Graph

Finally, unlike to the returns of correlation matrix, the re-ordering of the returns will affect the network topology. This is because the re-ordering of the returns will affect the estimation of the power associated with each frequency component, which will in turn change the correlation matrix and hence the network. Figures below show the effect of randomly shuffling the return series time-wise.

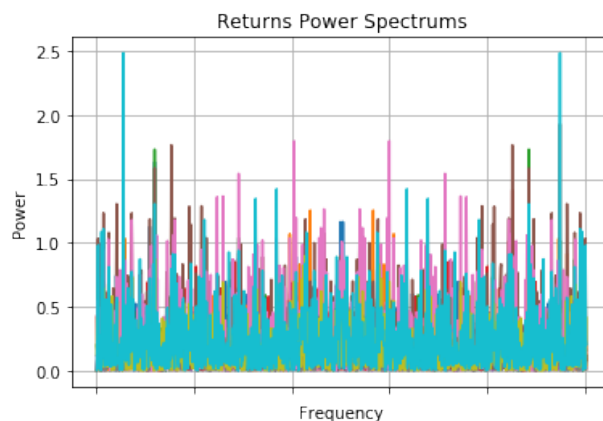


Fig. 46 Shuffled Returns Power Spectrum

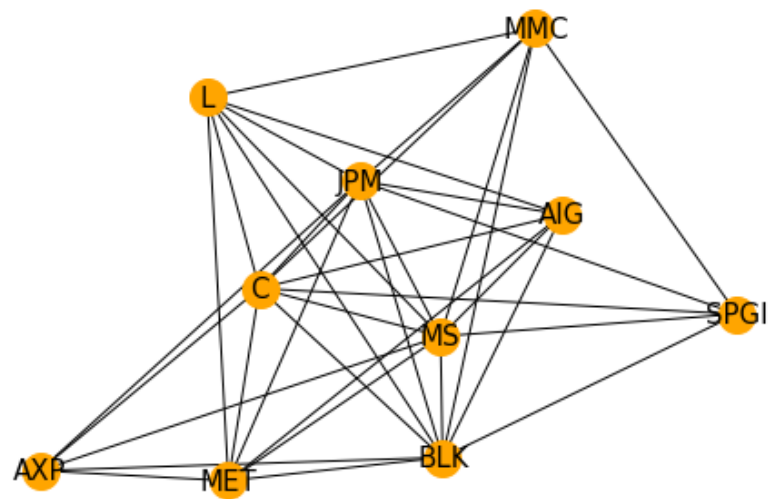


Fig. 47 *Shuffled Returns Spectral Correlation Graph*

## 5.5 Price vs Return Graphs

It is important to highlight that the above networks work only with returns and not with prices. For instance, if the network were drawn from the price correlation matrix, it would highlight the price trend relationships between each pairs of stocks. The price correlation network is drawn in the figure below.

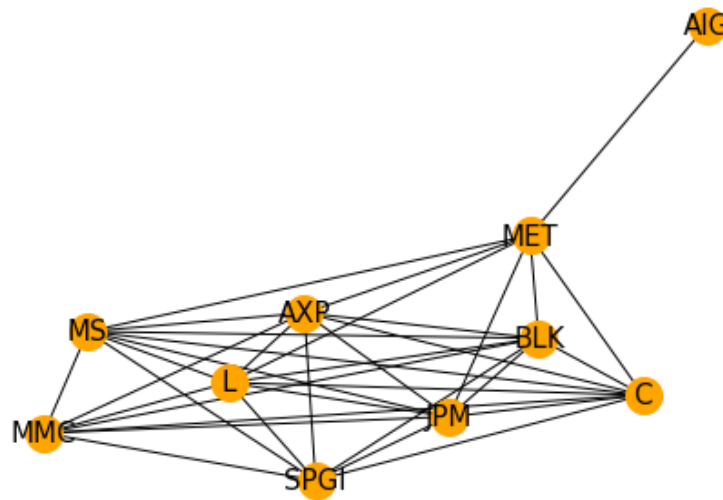


Fig. 48 Prices Correlation Graph

As it is clear from the network, AIG is an outlier and it is placed far away from all other stocks. This is congruent with the fact that the AIG is the worst performing stock out of all, as it is shown in the graph below where the prices are normalized to have an initial price of one.

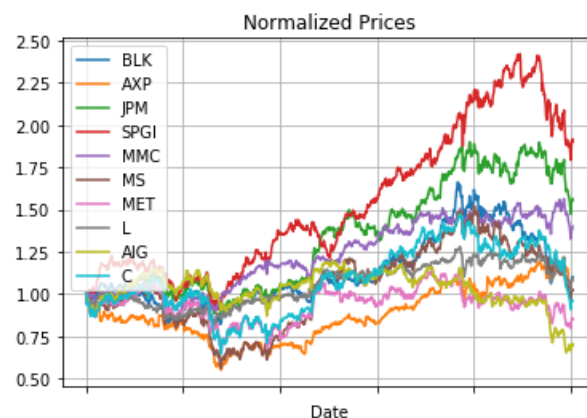


Fig. 49 Normalized Prices

Finally, given that the price series is not a zero-mean random variable, the power spectrum will have a strong bias component which will contain most of the power, which will make this analysis completely obsolete. This is shown not only in the power spectrum below but also in the network where every stock is well-connected with each other since the DC component is present in all stocks.

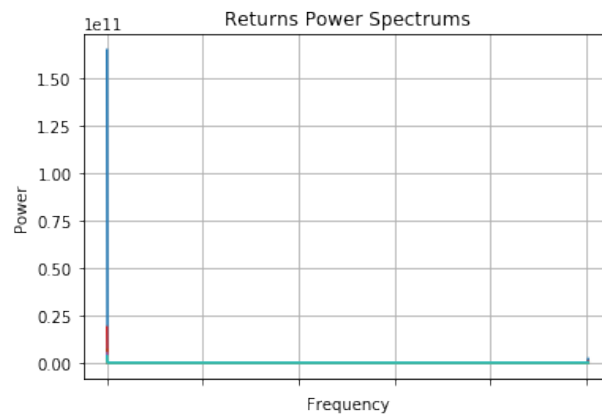


Fig. 49 *Power Spectrum of Prices*

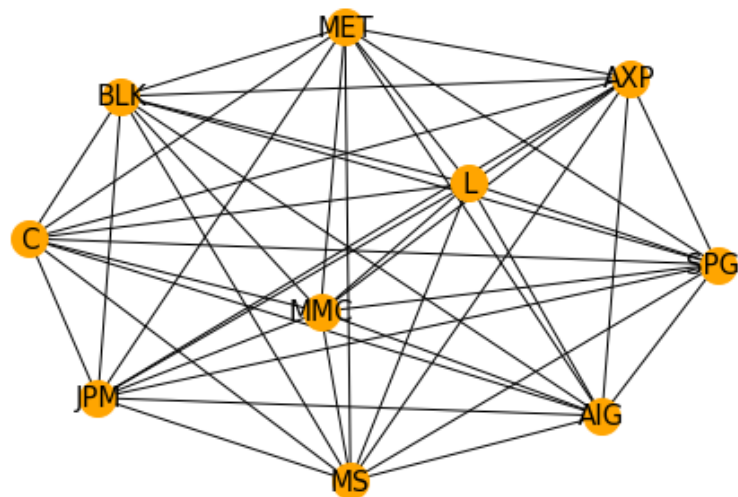


Fig. 50 *Prices Spectral Correlation Graph*

## References

- [1] D. Mandic. Lecture Notes: "Signal Processing and Machine Learning for Finance", April 2019.