Technische Universiteit Eindhoven
University of Technology

Department of Mathematics And Computer Science
Statistics, Probability and Operations Research
Postbus 513, 5600 MB Eindhoven
The Netherlands
www.tue.nl

Author
Ignas Šablinskas [redacted]

Date
January 5, 2025

2MBS20 Modelling Assignment: Modelling The AEX Stock Index

Group 17

Ignas Šablinskas [redacted]

Where innovation starts

Technische Universiteit
Eindhoven
University of Technology

Table of contents

Title
2MBS20 Modelling Assignment:
Modelling The AEX Stock Index

1 Introduction

Within modern society, an economy is ubiquitous within its existence. The movement of goods and services begets the trade of capital amongst participants of an economy. It happens all the time and more so with larger entities such as companies. Their worth is measured in shares all of which are assigned financial value and can be obtained by members of the public. The study of this financial movement produces an ample amount of (publicly-available) data. To explore relationships with this data we use statistics; which when constructed with a well-defined mathematical framework, allows us to leverage formalised structure to consider and study aspects of the global economy.

In this report, we present an analysis that studies the financial history and properties of the AEX European stock market index. This index encompasses the information of a large number of traded securities on the Amsterdam Stock Exchange. The AEX is a principal tool used by the stock exchange group Euronext.

To defined our period of interest, we have based our report on approximately three decades' worth of data starting from January 3 1994 to December 29 2023. Our data is categorised in the form of a spreadsheet that describes the opening and closing prices on a per-date basis; the highs and lows of the market on these days; the volume of money exchanged; and finally, its change in terms of relative percentage. Within this set of data, we study 7679 entries that are each attributed to a single date with the corresponding information described above.

We begin our examination using techniques from Exploratory Data Analysis (EDA) where we aim to summarise, present, and quantify a premier look at our data. This will include visualising our raw data and functions derived from it, where we inspect the properties of this reorganised formulation alongside its relevance to societal adages on finance.

Our next section will continue with forms of statistical inference where we restrict to a subset of our dataset. We will use our work in our initial analysis and consider further defined functions on our data in an attempt to inspect the nature and properties of randomness within it. We make use of visualisation to supplement our endeavour in using frequentist methods to inspect independence and correlation of our data; all in an effort to consider whether we can model them through an appropriate distribution. This will include using our assumptions as a foundation in the process of constructing (parametric) estimators and applying non-parametric estimation models and evaluating their effectiveness on potential client-oriented concepts and our confidence in our results.

Finally, we will end with a concluding section reflecting on our approach and methodology.

2 Exploratory Data Analysis

Initial Realizations

In the initial stages of our examination, we conduct some elementary Exploratory Data Analysis to begin recognising potential lines of thought for further investigation. We summarise our data to obtain a perspective that broadly considers key aspects of central tendency and dispersion. While our dataset provides ample categories of consideration for this process, we find it pertinent to start by surveying the opening and closing prices of our stock index. Thus, let us visualise these categories against the dates within our dataset.

Figure 2.1: Opening and Closing Prices of the AEX between 1994 and 2023.

Certain factors are immediately of note. The price (in euros) has multiplicatively increased overall by a factor of 4. At the beginning of our data period, we start at 200 euros and end at 800. With reference to our legend, we observe that the open and closing price graphs do not differ widely and produce plots that vary proportionally along one another. In part, this can be attributed to the nature of a stock market, which, despite being volatile over a time period does not diverge relatively in local date ranges as that could indicate an outlier period of economic variance.

A significant set of observations are the peaks and troughs throughout the graph. An initial response is to begin connecting the economic data with the social and cultural events of the considered society.

Let us begin in the twentieth century. As we near the turn of the millennium, we find a drastic increase in the value of the stock. We attribute this increase to the cultural and economic phenomenon referred to as the dot-com bubble [Duignan, 2024] where the initial rise of internet companies and their share values resulted in an economic boom. Yet, as all bubbles do not last forever the economic bubble cycled into a downturn that we can observe to last for the first half of the decade before beginning to recover.

Secondly, we observe another drop in the price towards the end of the first decade of the twenty-first century. We correspond this detail to the 2008-2010 Global Financial crisis and the resulting Great Recession [Weinberg, 2013] caused by the burst of the housing bubble across countries. Accordingly, the following decade underwent a period of economic recovery faltered by the rising tension of Brexit [Buigut and Kapar, 2023] and the European sovereign debt crisis [Hobelsberger et al., 2023].

Finally, the recovery had spiked downward and was greatly slowed by the COVID-19 pandemic which saw a rise in unemployment and businesses having their change their *modus operandi* to cope with increasing norms of social restriction and health precautions. With the lull of the pandemic restriction, we notice a reflection in our plot as the remaining years observe a period of recovery once more with the peak of the price falling approximately in the year 2021.

Outside these historical circumstances, our plots can provide other valuable information. We seek to classify in mathematical terms what our data carries in meaning. Let us now restrict to a plot consisting of our closing values alone.

Figure 2.2: Closing Prices of the AEX between 1994 and 2023.

Our data points act as a sequence increasingly indexed by time. Thus, our closing points are receptive to being modelled as a stochastic process. To eventually undertake this task, we can begin hypothesizing on the properties that this sequence can exhibit.

It follows into the case of wondering whether our data sample can be termed as random. Then, as is the case with any random vector and its realization, we want to start thinking about the distribution and independence of the random variables that could model our data. We will argue with reference to our plot of closing prices alone.

Considering the plot above, we immediately note that the closing prices will depend on the previous days. The principal reason is that the closing prices will depend on the opening prices, this is due to the closing price reflecting the price at the end of a specific date which would require a starting point that being the opening price. However, it is vital to realise that the opening price of a working day will certainly find a relationship with the closing price of the previous day. Suppose not, then the market will consist of non-deterministic jumps in price that have no relation, resulting in an economy that holds no sense of continuity in the movement of money alongside the transfer of goods. Thus, the closing price will inherently depend on other closing prices and their corresponding random variables have a greater likelihood of expressing the same behaviour

Our plot further encourages this assertion as we observe that the closing prices grow and decrease in size with one another which can suggest a possible dependence.

Therefore, we do not posit that the terms of this sequence can be independent, making a model where these values are considered i.i.d to be unviable. On the other hand, these random variables may exhibit relationships in terms of the identity of their distribution and perhaps we can consider this as a line of reasoning to evaluate in our statistical inference.

A Different Perspective On Our Data

Despite closing price alone not being a feasible measure, we believe it is potentially useful to view functions of our data as a measure for evaluation. Accordingly, we support the idea of considering the daily returns of the index. In essence, the notion of capital return if one were to buy stock at the beginning of a day and sell at the end. We define this as the difference between the opening and the closing prices over the opening value.

Let $x_t^{(opening)}$ denote the opening price on day t and $x_t^{(closing)}$ to be the closing price on the same day. Then the daily return rate r_t for day t is given by the following formula:

$$r_t := \frac{x_t^{(closing)} - x_t^{(opening)}}{x_t^{(opening)}}.$$

An important assumption here for our function to be defined is that the opening prices are non-equal to zero. This will be highlighted in subsequent sections as we will emphasize studying the daily returns.

We begin to analyze this quantity by computing and plotting its values for the given data.

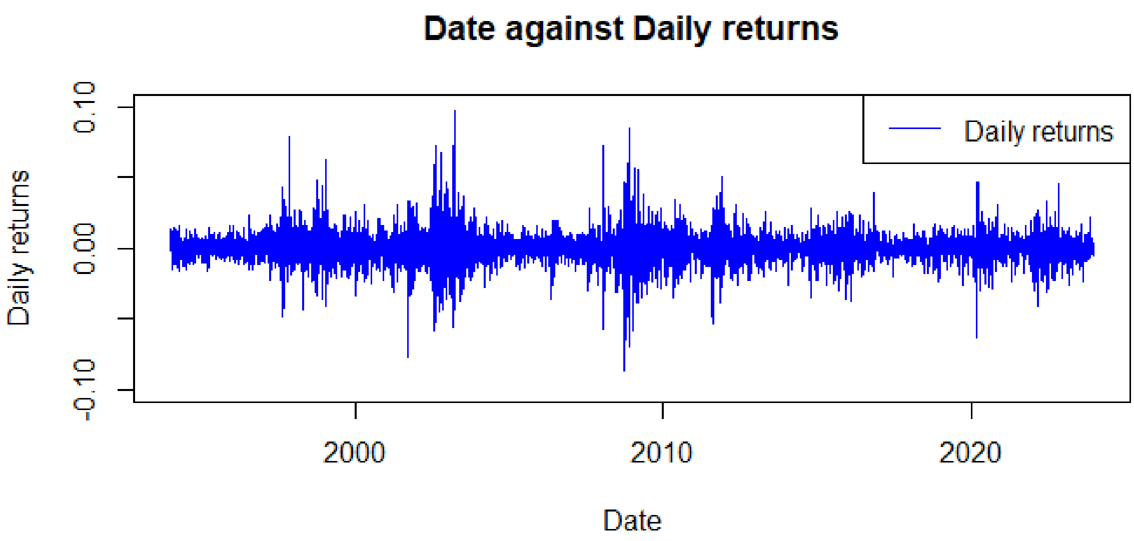


Figure 2.3: Daily returns through the years

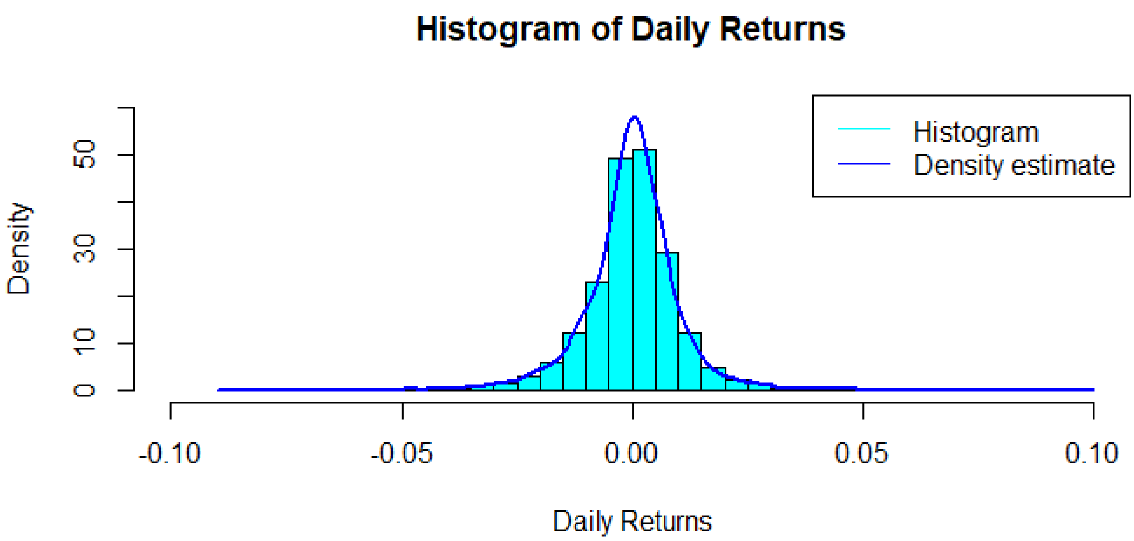


Figure 2.4: Histogram of Daily returns

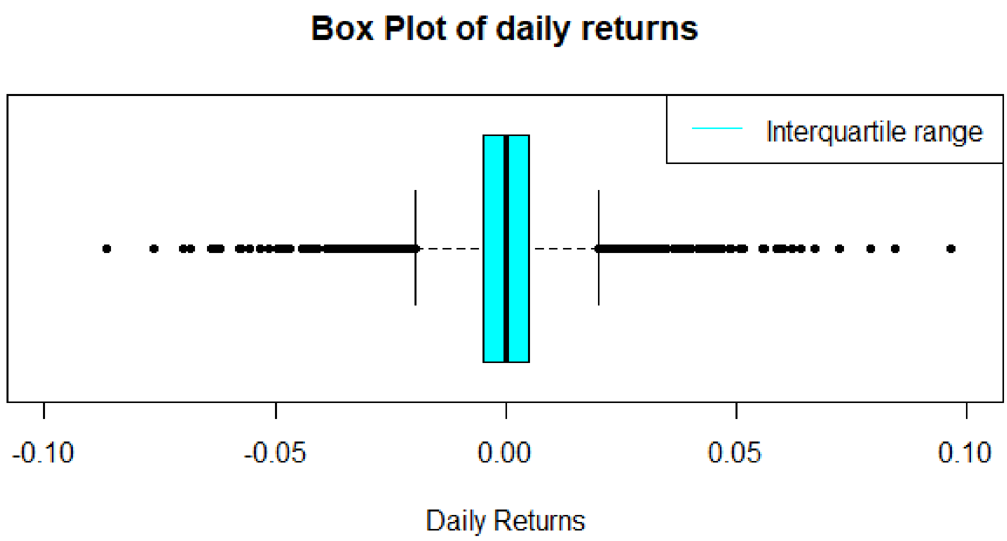


Figure 2.5: Box plot of Daily returns

The first plot provides an excellent visualization of the daily returns from January 3, 1994, to December 29, 2023. We observe that all values range from –0.10 to 0.10. However, in the majority of dates, the daily returns are between –0.5 and 0.5. There are only a few positive and negative spikes that, as one can see, may correspond to the previously mentioned historical events. The lack of large spikes at all points in these values describes a realistic scenario where the economy is not necessarily extremely volatile at all times. Moreover, the subsequent histogram and box plot affirm our descriptions of the daily returns' values, especially concerning frequency.

To proceed with our analysis, we will calculate some of the key statistical summaries.

Statistic	**Value**
Minimum	$-8.658 \cdot 10^{-2}$
1st Quartile	$-4.753 \cdot 10^{-3}$
Median	$2.296 \cdot 10^{-4}$
Mean	$-9.002 \cdot 10^{-5}$
3rd Quartile	$5.182 \cdot 10^{-3}$
Maximum	$9.677 \cdot 10^{-2}$
Standard deviation	$1.076 \cdot 10^{-2}$

Figure 2.6: Statistical summaries

Investigating the table, we can make a few general comments about the daily returns:

- Even though the mean is negative, the value is close to zero. This means that the daily return of a buyer resulted in an experience of a minor loss.
- The minimum represents the lowest daily return throughout the period. Similarly, the maximum represents the highest daily return. These extremes indicate that the stock

market has provided the opportunity for a buyer to experience both relatively significant daily losses and gains.

Similar to our consideration of the mathematical interpretation of closing prices, we aim to study whether the daily returns can act as a sequence whose terms are a random sample from a distribution. Looking at Figure 2.3, we notice that the values do not follow a noteworthy global pattern, but rather, local trends. From a broad perspective, they vary randomly from one another and tend to be unpredictable. Combining these observations alongside the lack of a plausible relationship as was described for the closing prices, we posit that the daily returns can be modelled as independent and identically distributed random variables. This model can be reasonable as the price changes depend on other economic factors and unexpected events.

Connecting Data To Society

We have discovered possible lines of thought through our initial examination. Of course, it can be beneficial to consider the connection of the stock market to society. The stock market involves publicly traded securities and a popular adage related to their sale is as follows:

"Sell in May and go away, but remember to come back in September."

In the midst of studying returns as a function, we lack the trends that are provided across years due to their argued independence. However, this does not preclude us from analysing whether we can find relationships amongst periods within years themselves.

A first approximation to vaguely discretize or separate a single year into periods is to use the aspect of seasons. This can be backed up by the quarterly seasonal weather found within Western Europe whose countries' markets are the major component of the index we are studying. This will enable us to critically question whether the claim of the adage that returns reduce during summer is true.

We begin with a few visuals:

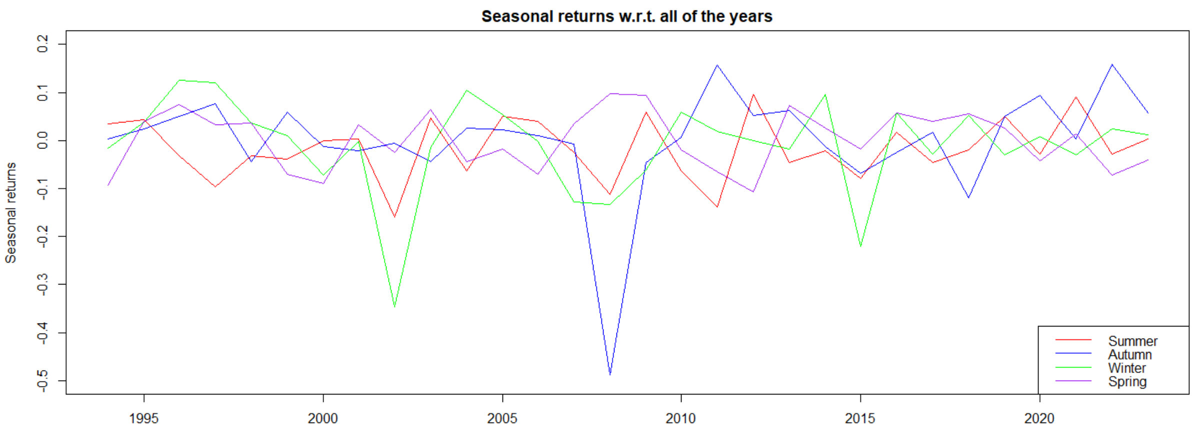


Figure 2.7: Seasonal Returns With Respect To All Years.

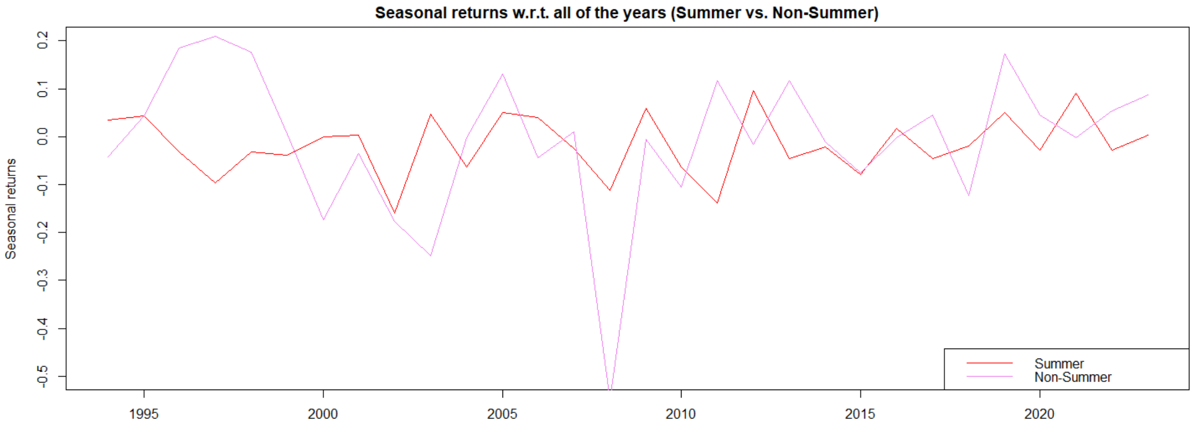


Figure 2.8: Seasonal Returns With Respect To All Years (Summer vs. Non-Summer).

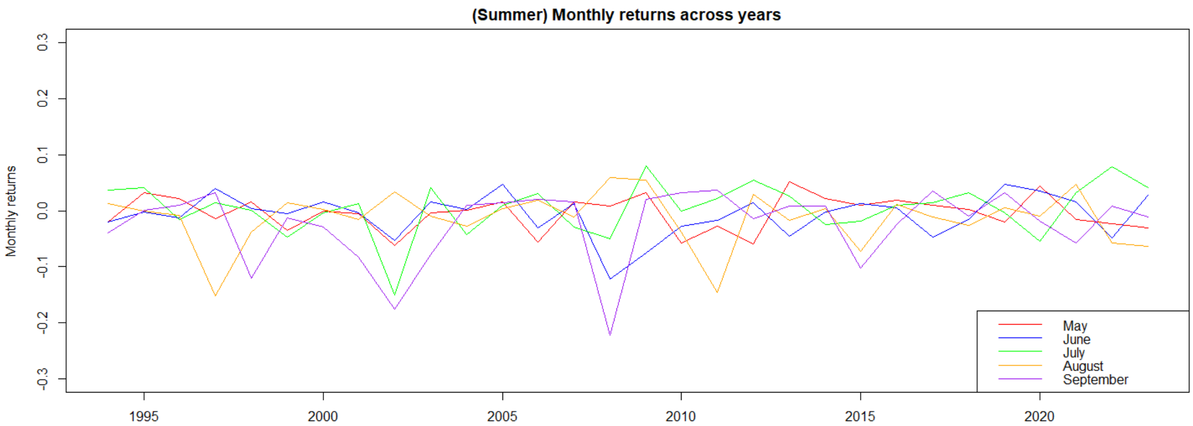


Figure 2.9: Summer Monthly Returns Across All Years.

Now, we will split our interpretation into two sections. The first will consider the summer returns against other seasons, next, we will consider the returns against all non-summer months alongside the summer months individually. In the end, we will combine our perspectives to conclude our thoughts.

Primarily, we note with reference to Figure 2.7 that the summer months exhibit returns that are non-significant in difference relative to the other seasons. They have no prerequisite to follow a specific pattern and aside, from outliers in specific years (that are possibly correlated with historical and societal events) the daily returns exhibit a consistent pattern.

Secondly, in Figure 2.8 we note that the combination of the remaining months does not produce a significant pattern that conflicts substantially from the returns during the summer. Of importance, is to note that this pattern stands true outside the influence of visually-prominent outliers. Furthermore, Figure 2.9 does not argue that the summer months alone have particular changes in returns that would cause a divergence in returns either lower or greater.

Therefore, after collecting our perspective from our visual results, we posit that the adage does not hold to be true.

3 Statistical Modelling and Basic Inference

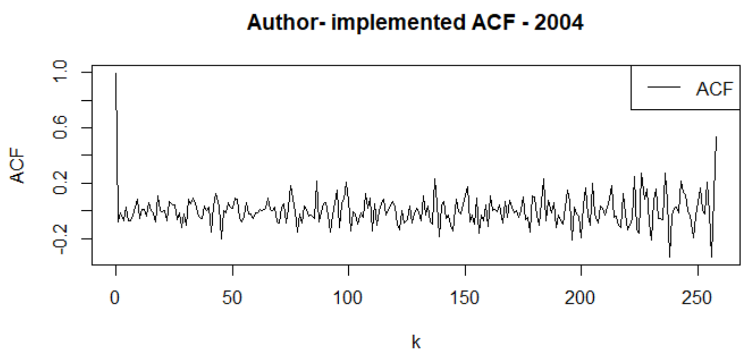
Correlation

So far, We have now defined potential lines of thought for investigation through our EDA. From the perspective of inference, we will simplify our study by considering two periods of time from our dataset: the individual calendar years 2004 and 2019. In this section, our main goal will be to design and investigate a reasonable model for the daily returns.

First and foremost, we will check if we can model the values for 2004 and 2019 as independent and identically distributed random variables. A qualitative way to do this is to look at the sample autocorrelation function. We will begin with our own implementation of an Autocorrelation function, following which, we will compare its performance and results with a built-in variant from the statistical software package R.

For this component of our analysis, we make the assumption that the daily returns have positive and finite variance.

Autocorrelation plots are usually used for checking randomness in a set of data. The values of ACF range from –1 to 1. The autocorrelation at lag zero always equals 1 since it measures the correlation of each term with itself. If the data can be modelled as an i.i.d. sample from an unknown distribution, then all lags (except the zero one) should have autocorrelation close to zero.



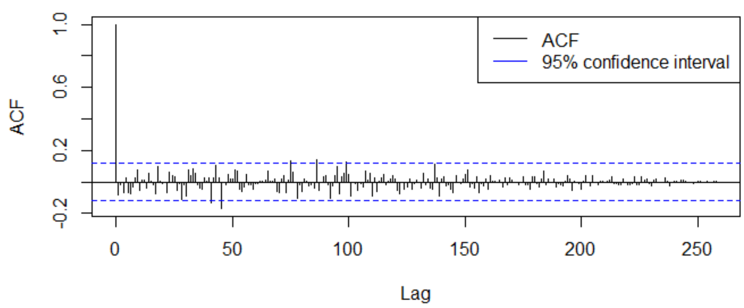


Figure 3.1: Built-in ACF - 2004

The plot for 2004 shows that almost all ACF values are in between –0.4 and 0.4. Towards the end, the values start to fluctuate slightly more, but they remain relatively close to zero.

This is an indicator that the daily returns for 2004 can be modeled as i.i.d. random variables. To ensure that we will look at the results from the built-in function. The blue lines represent the 95% confidence interval. Since the built-in ACF values (except a few outliers) are in this interval, we can conclude once again that the data for 2004 is an i.i.d sample from an unknown distribution.

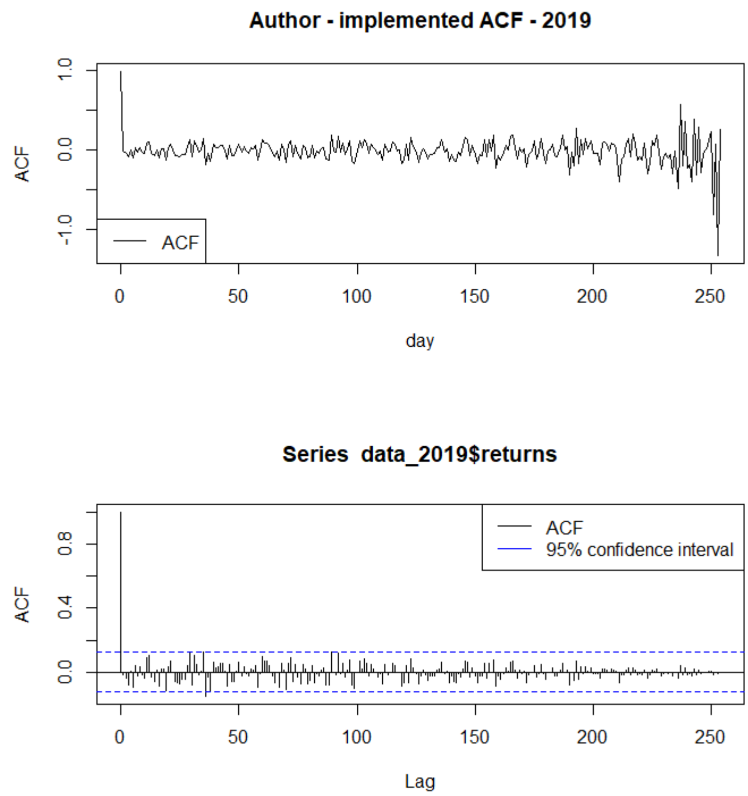


Figure 3.2: Built-in ACF - 2019

Observing the plot for 2019, we notice that most of the values range from –0.3 to 0.3, which is a reasonable interval for concluding independence. However, there is a single value that goes below –1. This anomaly is hard to explain. It may be due to a roundoff error, since we are working with small numbers. Nevertheless, we are going to proceed with the analysis of the ACF results. Since all other values are relatively close to zero, we can conclude that the daily returns for 2019 are i.i.d. random variables. To check if our intuition is correct, we will investigate the plot from the built-in ACF. Again, almost all values are in between the 95% confidence interval. Then the random sample model is suitable for the data from 2019.

In the next step of our report, we aim to use visualization tools to further explore the properties of the data from the years 2004 and 2019.

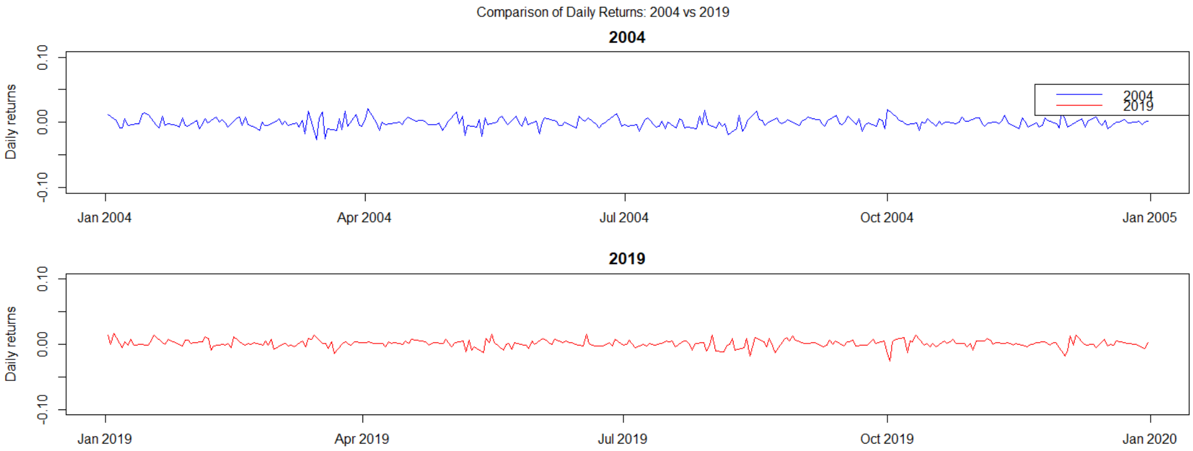


Figure 3.3: Comparison Of Daily Returns 2004 vs. 2019.

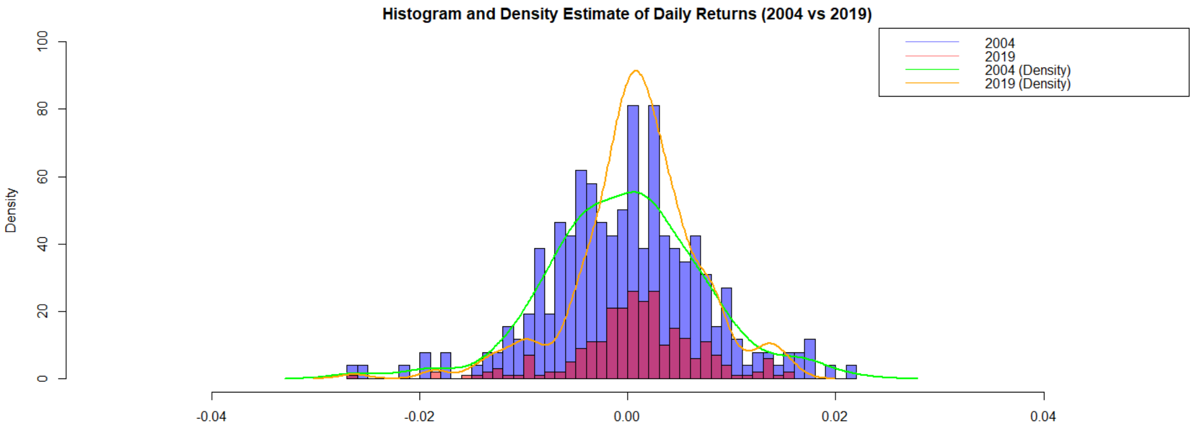


Figure 3.4: Histogram and Density Estimates of Daily Returns (2004 vs. 2019).

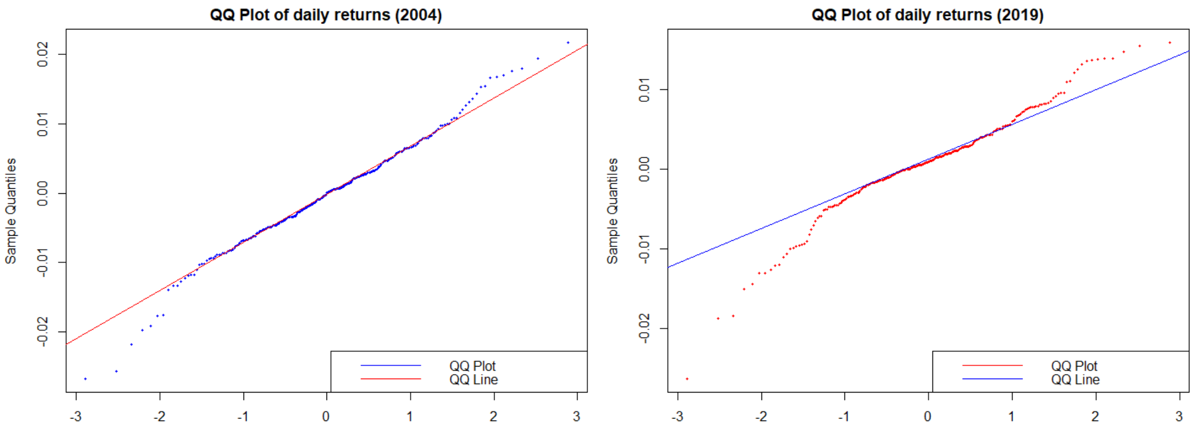


Figure 3.5: Normal QQ Plots for the years 2004 and 2019.

In Figure 3.3 we note that the daily returns for both years are largely consistent around the zero value with a specific range of variation. We observe that 2004 is a bit erratic in April with

2019 finding relative dips in October. Overall, they are largely similar with 2019 facing more negative returns relative to its year and 2004 having a slight lean to the positive side.

Our following plot, Figure 3.4 substantiates our case as we indeed see that 2004 is more haphazard in its trends and thus has a greater visual variance relative to 2019 which displays a conspicuous peak at the centre. Both remain to be visually centred around zero and would have a mean that displays a negligible difference to it. Moreover, in terms of the histogram frequency density in area and the density estimates calculated through Kernel Density Estimation (KDE) both exhibit a shape that could potentially approximate a Gaussian. Thus, turning our attention to using a Normal QQ plot.

Subsequently, we visualise both data groups in normal QQ plots (Figure 3.5) where we remark that both have a good portion of the points going through the QQ line. Interestingly, the line is more uniform for 2004 despite its more turbulent distribution of daily returns and in 2019 we note the opposite where we find a greater number of points outside the line despite its uniformity. We do however contend that the discrepancy in points is not extremely significant to exclude the possibility of the data from 2019 being sampled from a normal distribution.

Collectively, analysing these plots brings us back to our proposed review of daily returns as a measure can be modelled as a random (i.e. i.i.d) sample from a distribution. There are several methods for checking the validity of this model. In the next paragraphs, we are going to examine this aspect of our subset of data.

If the data can be modelled as a sample from a continuous distribution, then the histogram can be used as an estimate of the corresponding underlying probability density function. In our case, as noted, both histograms have bell-like shapes. Moreover, we observe that both curves are almost symmetric around the corresponding mean values. As a result, the normal distribution seems to act as a suitable model for the values obtained in 2004 and 2019.

To further endorse this position we need only consider the QQ plots. As we have commented for 2004 we observe that the points form almost a straight line. This is an indicator that the proposed model is satisfactory. For 2019 the situation is slightly different in that we observe the existence of some outliers. Despite this case, most of the points predominantly lie on the QQ line. Therefore, a normal distribution can perform as an appropriate model for the data from 2019.

On a final comment, the histograms present that one is much steeper than the other. Then we can expect the standard deviation for 2019 to be smaller than the standard deviation for 2004 since it is a measure of dispersion.

Estimator Construction and Evaluation

By our argument, we can model the given data using a normal distribution. We will now construct estimators for the corresponding parameters i.e., the mean μ and the variance ν. We assume that the variance is strictly positive and finite.

Suppose that $\{X_i\}_{i=1}^n$ is the data for one calendar year. Then $\{X_i\}_{i=1}^n$ are independent and identically distributed random variables, such that $X_i \sim \mathcal{N}(\mu, \nu)$. Let $\theta = (\mu, \nu) \in \Theta \subseteq \mathbb{R}^2$, where Θ is the parameter space.

To construct an estimator $\hat{\theta}$ of the unknown parameter θ we are going to use the maximum likelihood estimation method. Then the estimator $\hat{\theta}$ is defined as

$$\hat{\theta} := \arg\max_{\theta' \in \Theta} \prod_{i=1}^{n} f_{\theta'}(X_i).$$

Then the likelihood function is given by

$$L(\theta, X) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\nu}} e^{-\frac{(X_i-\mu)^2}{2\nu}}.$$

We observe that the likelihood function is positive and differentiable. As a result we can introduce the log-likelihood function

$$l(\theta, X) := \log L(\theta, X) = -\frac{n}{2}\log(2\pi\nu) - \frac{1}{2\nu}\sum_{i=1}^{n}(X_i-\mu)^2$$

and the score function

$$\dot{l}(\theta, X) := \begin{pmatrix} \frac{\partial}{\partial\mu} l(\theta, X) \\ \frac{\partial}{\partial\nu} l(\theta, X) \end{pmatrix} = \begin{pmatrix} \frac{1}{\nu}\sum_{i=1}^{n}(X_i-\mu) \\ \frac{1}{2\nu}\left(\frac{1}{\nu}\sum_{i=1}^{n}(X_i-\mu)^2 - n\right) \end{pmatrix}.$$

Now we will identify the local extremes of the likelihood function by equating the score function to zero. Since $\hat{\theta}$ maximises the likelihood function we are interested only in the extreme corresponding to the global maximum. In our case we obtain the following estimator

$$\hat{\theta} = \begin{pmatrix} \hat{\mu} \\ \hat{\nu} \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} X_i \\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j)^2 \end{pmatrix}.$$

To explain our choice of estimators, we are going to investigate the properties of $\hat{\mu}$ and $\hat{\nu}$. First, we are going to compute the bias and the variance of these estimators:

- $b(\hat{\mu}) = \mathbb{E}[\hat{\mu}] - \mu = \mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} X_i] - \mu = \mu - \mu = 0$;
- $b(\hat{\nu}) = \mathbb{E}[\hat{\nu}] - \nu = \mathbb{E}[\frac{1}{n}\sum_{i=1}^{n}(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j)^2] - \nu = -\frac{\nu}{n}$;
- $\mathrm{Var}(\hat{\mu}) = \mathbb{E}[\hat{\mu}^2] - \mathbb{E}[\hat{\mu}]^2 = \frac{\nu}{n}$;
- $\mathrm{Var}(\hat{\nu}) = \mathbb{E}[\hat{\nu}^2] - \mathbb{E}[\hat{\nu}]^2 = \frac{2(n-1)\nu^2}{n^2}$.

Having these results, we want to make some conclusions about the mean squared error of the estimators. Furthermore, we want to check if $\hat{\mu}$ and $\hat{\nu}$ achieve asymptotically the Cramér–Rao bound.

- The estimator $\hat{\mu}$ is unbiased and $\mathrm{Var}(\hat{\mu}) \to 0$ as $n \to 0$. Then with reference to [Röttger, 2024], we can conclude that the estimator is MSE consistent.
- Since $b(\hat{\nu}) \to 0$ as $n \to 0$ the estimator $\hat{\nu}$ is asymptotically unbiased. We also have that $\mathrm{Var}(\hat{\nu}) \to 0$ as $n \to 0$. Then with reference to [Röttger, 2024], we can conclude that the estimator is MSE consistent.

As a result, both estimators satisfy the conditions of the Asymptotic Normality of MLE Theorem provided in the lecture slides. Then the estimators $\hat{\mu}$ and $\hat{\nu}$ achieve the Cramér–Rao bound, meaning that they are asymptotically the best estimators.

To visually prove our intuition we plotted the density function of the normal distribution over the histograms for the daily returns from 2004 and 2019. As we can see, the graphs of the density functions acceptably follow the shape of the histograms, implying that this is a suitable model for our data.

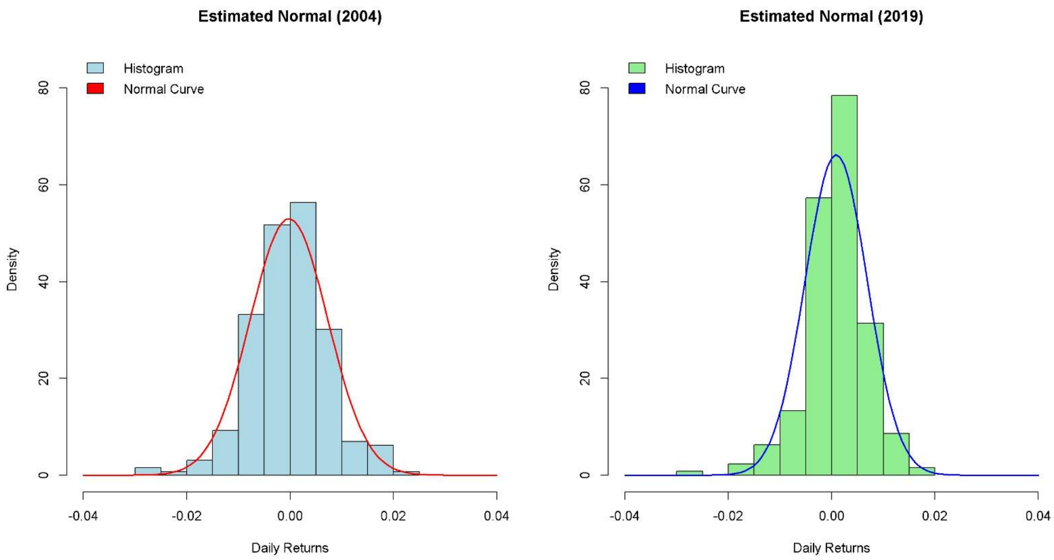


Figure 3.6: Density function vs Histogram

Note: for the following exercises we use the following estimates for the mean and the variance based on the formulas above, but for a realized random vector $\bar{x}$ instead:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \nu = \frac{1}{n}\sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j)^2$$

The Value-at-Risk for some p can be computed (in R programming language) using the following formula:

$$\text{VaR}_p = -\left(\mu + \nu \cdot qnorm(p)\right)$$

Where:

- μ is the mean of the returns,
- ν is the standard deviation of the returns,
- $qnorm(p)$ is the quantile function for the standard normal distribution at probability p, in our case $p = 0.01$.

This formula works, since we can normalize

$$\mathbb{P}(R \leq x) = \mathbb{P}(\frac{R-\mu}{\nu} \leq \frac{x-\mu}{\nu}) = \mathbb{P}(Z \leq \frac{x-\mu}{\nu}) \geq 0.01$$

where Z follows standard normal distribution and find that

$$x = \mu + \Phi^{-1}(0.01) \cdot \nu$$

as desired. For years 2004 and 2019 we get the following Value-at-Risk (VaR) values respectively: 0.01779959 and 0.01317642.

The empirical $100p\%$ quantile for some p can be computed (in R programming language) using the following formula:

$$\text{Quantile}_p = quantile(data,\ p)$$

After computing them, the non-parametric estimates for the years 2014 and 2019 for the extent of assets companies need to have to cover possible losses turns out to be 0.02063069 and 0.01661635 respectively.

Now, parametric estimators are more robust as they do not depend on any assumptions made by the statisticians and are concerned with the actual empirical data. Furthermore, non-parametric estimators are very sensitive to outliers and can distort the data, although, in our case, where we are trying to help companies to determine the extent to which they should prepare for big losses, considering the outliers might be most useful thing to do as this is where the companies would suffer the most. We can see that for both years, the non-parametric estimations are higher, meaning, they capture the more extreme cases and thus, in some sense, makes the non-parametric estimation more useful, but at the same time, if the current economy is stable, an estimate which pushes companies to dedicate more assets for preparation of an economic crash can turn out to be counterproductive, making the parametric estimators more practical.

For estimating the mean, we use the formula from above (estimator exercise) to get the following result: $\mu_{2004} = -0.0002468,\ \mu_{2019} = 0.0008694$. This suggests that on average the market generated slight profits in 2004 and modest losses in 2019.
For confidence interval calculation we used the following formula ([Röttger, 2024]):

$$[\mu - z_{0.025}\frac{\sqrt{\nu}}{\sqrt{n}}, \mu + z_{0.025}\frac{\sqrt{\nu}}{\sqrt{n}}]$$

where μ is the mean estimate, n is the sample size (number of days in the respective years) and ν is the variance estimate and obtained the following results:

$$c_{2004} = [-0.00116, 0.00067] \quad c_{2019} = [0.00013, 0.00161]$$

Since we were considering $n_{2004} = 255$ and $n_{2019} = 259$ days for years 2004 and 2019 (greater than 50, which is the threshold for trusting normal distribution, [Röttger, 2024]) we can expect to get accurate confidence intervals. Confidence interval for true mean during 2004 does contain the 0 value, which indicates that the market conditions in 2004 may not have been different from neutral returns. This can not be said about 2019, since the confidence interval shows us that the true mean is more likely to be positive, indicating overall positive returns.

4 Conclusion

We recollect the progress that we have made so far. We began with a brief analysis and a recounting of the provided data. In this venture, we focused mainly on the opening and the closing prices as categories within our dataset. To understand the societal background of our modelling inquiry, we commented on some of the key events that have influenced the stock market and our data throughout the years. Armed with an initial grasp of the environment we introduced one of the crucial concepts in this report, that being the daily returns. This quantity played an important role in our investigation as we discovered that it can be nicely modelled in a befitting manner as an i.i.d. sample from some continuous distribution.

We continued by reducing the complexity of our setting. To effectively begin the process of drawing statistically significant conclusions, the rest of the report was dedicated to the examination of the data for the years 2004 and 2019. Initially, we used the auto-correlation function to understand whether the data can be modelled as an i.i.d. sample. Consequently, we investigated the normality behaviour of the given sample and proceeded to study our data using the Gaussian distribution. We devised estimators for the parameters of the distribution which were later used in the corresponding estimation of them and in the enhancement of investor-aligned risk management strategies. To culminate, we constructed confidence intervals that further reinforced our conviction in our method and the handling of data.

In summary, our work within this course has supplied us with a chance to grow in the skill of applying statistical mathematics and seeing theory play out in real-life scenarios. The difficulty in setting up adequate experiments and inference through direct proof implores the use of computational tools and estimates derived from them. We recognise that the use of further numerical simulations provides roads to challenge our model and test results that allow us to bridge the gap between the mathematical ideal and a pragmatic study of our question at hand. In future projects, we hope to delve further into the use of these tools.

Despite the pitfalls of our proposed model, we believe that our model provides an effective start to the mathematical modelling cycle and are confident in our ability to further question, research and collaborate to improve on what we have developed. All things considered, mathematical models are iterative by design; it is natural for them to evolve.

Bibliography

S. Buigut and B. Kapar. How did brexit impact eu trade? evidence from real data. *The World Economy*, 46(6):1566–1581, Apr 2023. doi: 10.1111/twec.13419.

B. Duignan. dot-com bubble, Sep 2024. URL `https://www.britannica.com/money/dot-com-bubble`.

K. Hobelsberger, C. Kok, and F. Paolo Mongelli. A tale of three crises: synergies between ecb tasks, Jun 2023.

F. Röttger. *Lecture Slides 2MBS20/2WS30 Foundations Of Statistics/Mathematical Statistics*. Technische Universiteit Eindhoven, 2024.

J. Weinberg. The great recession and its aftermath, Nov 2013. URL `https://www.federalreservehistory.org/essays/great-recession-and-its-aftermath`.