Time Series Analysis of Spot Price Market Data

Joynul Muhboob
Student Number: 150320762
MSci Project MTH798U
Supervisor: Dr Wolfram Just
School of Mathematical Sciences

Abstract

In my dissertation, we will investigate whether extreme events, such as seasonal changes, have an impact on the autocorrelation function of a dataset. We observed electrical prices from 1999 to 2007, that were in an hourly interval. We conditioned the data on an hourly basis and tried to see whether the trend shown is what we would expect. The electricity prices were collected from a company called Nord Pool. When analysing the whole dataset and the data set conditioned hourly, we noticed some interesting extreme events. For example, at 2 am we noticed positive extremities that had somewhat of a periodic pattern, we then came up with a possible reason as to what caused such a periodic pattern. Furthermore, we investigated whether extreme events have an impact on how correlated the datasets are to itself, what we noticed is that the extreme events did not have an impact on the correlation. Finally, we analysed which dataset follows a Geometric Brownian Motion. The relevance of analysing which dataset follows a Geometric Brownian Motion is explained in section 6.

Table of Contents

1.0 Introduction

What is financial time series analysis? Well in simple terms a financial times series is the study of the prices of commodities or assets over some time. These studies are recorded and as a result, this leads to a huge amount of data being made available online. These data's have a huge role to play when it comes to predicting how future prices of commodities or assets will be impacted by certain events. The most iconic event that occurred which made a huge impact worldwide was the financial crisis that occurred in 2007. Many banks were impacted heavily, for instance, the Lehman Brothers filed for bankruptcy on September 15, 2008, and many others were expected to follow, such as Merrill Lynch, AIG, HBOS, Royal Bank of Scotland, Bradford and Bingley, Hypo and many more. Banks were not the only ones affected. The spending habits of the general public was affected immensely. The demand for goods had reduced due to economic uncertainty, as a result, firms witnessed a reduction in profit, this increased unemployment. Germany, France, Italy and Spain were among the countries that were hit the hardest by the financial crisis and took some time to recover from it. It is essential to be able to forecast how extreme events such as financial crisis would have affected prices of commodities or assets in the future, that way the government would be able to make strong financial decisions to cope with the damage done by these extremities. Another example which not only the UK but the whole world is going through is coronavirus. Around November 2019 coronavirus started spreading around China, by February the virus had spread throughout the world. With a lockdown announced in the UK and other parts of the world, several markets would have been hit extremely hard. One market that was affected was the property market, and with people stuck at home, we can expect electricity prices to be impacted as well. *The Guardian* (2020) has stated in an article that the UK economic output is set to plunge by 15% and unemployment is set to double. Furthermore, analysts at the Centre for Economics and Business Research (CEBR) has stated that *"The deepest recession since the financial crisis is now all but unavoidable".*

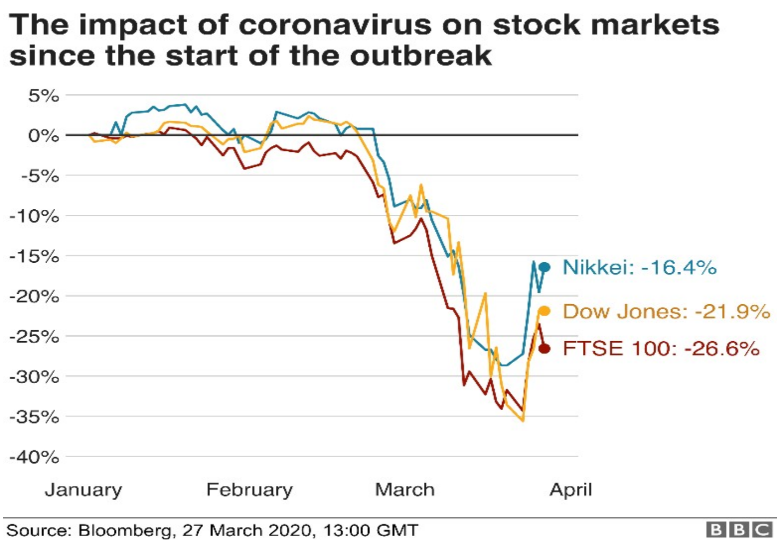


FIGURE 1:BLOOMBERG, 2020, "THE IMPACT OF CORONAVIRUS ON STOCK MARKETS", BBC NEWS [ONLINE}. AVAILABLE AT: HTTPS://WWW.BBC.CO.UK/NEWS/BUSINESS-51706225 [ACCESSED 30/03/2020]

Figure 1 illustrates the impact coronavirus has had on certain stock markets; this should give an idea about the impact that certain events have on the economy. It is essential to be able to see how prices are affected due to extreme events and then try to see whether future prices are correlated, that way predicting prices will become easier, thus the right political decisions could be made. In this thesis, we will analyse the behaviour of electricity prices in Nord Pool from 01/01/1999 to 26/01/2007. Firstly, we will take a look at the whole dataset, and then we will condition the data at an hourly rate and investigate further. But, before we get into that, we will give an overview of the financial market, the electricity market and Nord Pool, followed by the concept of stationarity and returns. And then we will finally analyse our dataset in detail and then conclude our findings.

2.0 Background

2.1 Financial Market

A Market is defined as a means by which good and services are exchanged between buyers and sellers, this can be done directly or through agents and institutions [BRITANNICA 2017]. These goods and services can be, but not limited to, bonds, equities, derivatives, or currencies. Markets can vary in size, for instance, the largest trading market is the New York Stock Exchange (NYSE) with a market cap of US$29.3 trillion. There are different types of markets. Below are 4 examples:

1. **Stock Market:** In a stock market ownership of public companies are traded. Each share is sold at a specific price, if the stock performs well in the market, then the investors make money. One advantage of the stock market is that it is easy to buy stocks, however, the difficulty is buying the right stock which will maximise investors earnings. Investors usually use various indices to monitor how the stock market is performing. Examples of these indices are the Dow Jones Industrial Average (DIJA) and the S&P 500.

2. **Bond Market:** In a bond market companies and government are offered the opportunity to secure money to finance a project or investment. How this works is that investors buy bonds from a company, and the company then returns the amount of the bonds within an agreed price, with interest added on top.

3. **Commodities Market:** This type of market involves the buying and selling of natural resources or commodities such as corn, oil, meat and gold. In this type of market prices are unpredictable. There is also a commodities future market where the price of items that are set to be delivered at a future date is already set and sealed.

4. **Derivative Market:** A derivative is a financial contract whose value depends on the value of some underlying asset [*M. Phillips 2020].* In this type of market derivatives or contracts whose value is based on the market value of the asset being traded is involved. Derivative products can be broken down into two classes, namely “lock products” and “option products”. Lock products such as swaps, futures or forwards bind the respective parties from the outset to the agreed-upon terms over the lifetime of the contract. Option products such as interest rate swaps give the buyer the right but not the obligation to adhere to the contract under the agreed terms.

2.2 European Electricity Market

Overtime the electricity market has changed in the way it operates, these changes have occurred due to regulations placed on the amount of greenhouse gasses emitted per country. For instance, on 1st October 2012 in Australia a new legislation came into effect, which was known as the Greenhouse and Energy Minimum Standards Act (GEMS), this legislation created a national framework for product energy efficiency in Australia. Even though this legislation was specific to Australia and New Zealand, other countries had some sort of ruling opposed that put a limit onto how much greenhouse gasses were allowed to be emitted. As a result, this bought about changes to the electricity market. The biggest change that was bought about was the fact that electricity generated from renewable sources has been on the rise, electricity generated from renewable sources has surpassed electricity generated from gas. The Department for Business, Energy and Industrial Strategy has stated that "*In 2010, the UK used 40.23 tonnes of coal to produce power with 345.69 terawatts per hour (TWh) of gas used. However, by 2018, just 6.64 tonnes of coal were used – a dramatic shift.*" These changes also bought about changes in, for example, transport. These come in the form of an increase in hybrid/ electric vehicles.

The Electricity System

The picture below gives a brief overview on how the electricity system works.

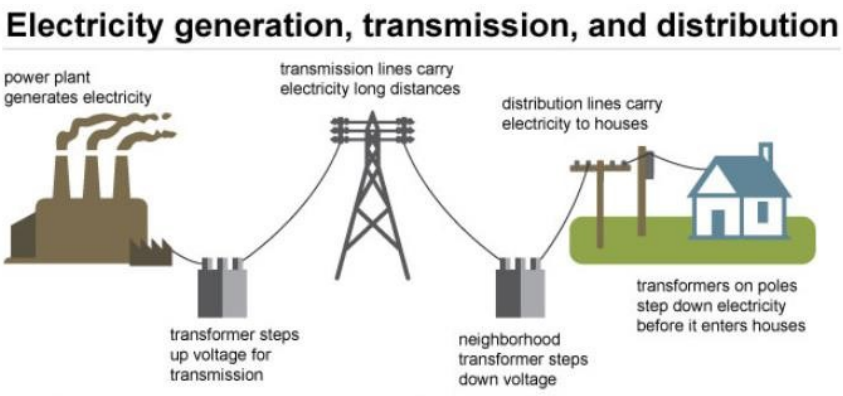


FIGURE 2: ADAPTED FROM NATIONAL ENERGY EDUCATION DEVELOPMENT PROJECT [ONLINE] AVAILABLE AT:HTTPS://WWW.EIA.GOV/ENERGYEXPLAINED/ELECTRICITY/DELIVERY-TO-CONSUMERS.PHP [ACCESSED 31/03/2020]

The whole process starts at a power plant. This is where electricity is generated. Next, the transformer then steps up the voltage for transmission. The electricity after being stepped up is then passed along transmission lines. The role of the transmission is to send the high voltage electricity across long distances. The electricity is then passed from the transmission lines to the neighbourhood transformer. The role of the transformer is to step down the voltage, so it is safe for households to use electricity. Finally, the distribution lines carry electricity from the neighbourhood transformer to houses. But before electricity is passed onto the houses, the transformers on poles step down the electricity, which allows electricity to enter houses. We have spoken about how electricity is transmitted to houses, but how does the market work? We shall answer this in the next section.

The UK Electricity Market

As mentioned above, the power plants are where electricity is generated, but that is not the only role of power plants. In a power plant, the electricity generators sell the electricity on the wholesale market. These generators can also buy electricity if enough electricity has not been produced to meet demand. It is the electricity suppliers who buy electricity from the wholesale market to supply it to their customers. If suppliers buy more electricity than required, then they can sell it back into the wholesale market. We also have traders; their job is to ensure customers get the best price for electricity. The price at which electricity is supplied to customers depends on demand and supply.

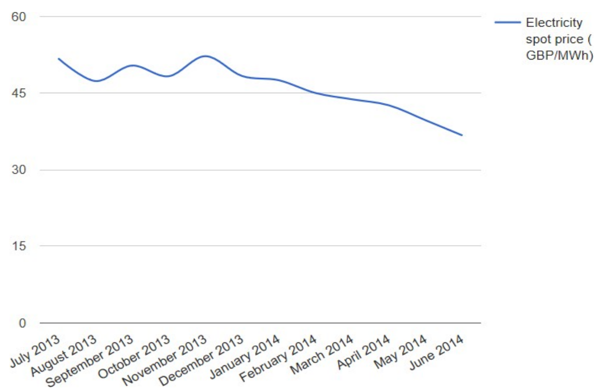


FIGURE 3: APX SPOT PRICES [ONLINE] AVAILABLE AT: HTTPS://THESWITCH.CO.UK/ENERGY/GUIDES/ELECTRICITY-MARKET [ACCESSED 31/03/2020]

Figure 3 shows us how electricity spot prices change every month. Demand for electricity is higher on certain months, this in return causes the price to be higher, this could be due to seasons.

Now we keep mentioning that electricity is traded in the wholesale market. But how is electricity traded in the wholesale market? Well, this is done in 3 different ways.

1) **Bilateral trading:** This refers to contracts being formed between generators and suppliers for the purchase of electricity. This is obtained through some sort of master contract for a set period, this contract establishes trading conditions. Individual trading contracts then set the amounts of electricity to be traded and the trading price.
2) **Market Trading:** Electricity supply and demand is mostly matched on two exchanges when it comes to electricity delivered on the same day or the next day. Next, an auction process matches the offers from the generators and the bids from suppliers or large consumers.
3) **Long-term Trading:** This is achieved through electricity brokers. Examples of these brokers are: BGC Partners, Evolution Markets Ltd, GFI Group Inc, ICAP plc, Marex Spectron Group Ltd, PVM Oil Associates, Tradition Financial Services Ltd, Tullet Prebon Inc, Griffin Market Limited. When it comes to long-term trading, electricity trading prices are established less according to supply and demand, and more on forecast market development.

2.3 Nord Pool

The origin of Nord Pool can be traced back to 1932, where it was known as the *Coordination Association.* It was formed by eastern Norwegian electricity companies, and soon it encompassed all the electricity companies in eastern Norway. In 1971 the coordination association merged with many electricity companies in other parts of Norway, and it became known as the *Coordination of power stations.* In 1988 it consisted of 118 power companies as members. In 1991 the Norwegian parliament decided to deregulate the market for power trading, this allowed the establishment of Statnett Marked AS in 1993. In 1996 the Swedish electricity market was also deregulated, thus Statnett Marked AS was replaced by Nord Pool ASA, which was owned equally by the Swedish and Norwegian Transmission System Operators (TSO's), Svenska Kraftnat and Statnett. In 1998, 1999 and 2000 Finland, Western Denmark and Eastern Denmark respectively joined Nord Pool ASA. On December 27, 2001, Nord Pool ASA's spot market activities were spun off into a new company called Nord Pool AS. The remaining parts of Nord Pool ASA were later acquired by Nasdaq, which is currently known as Nasdaq Commodities. On 12th January 2010, Nord Pool AS in cooperation with Nasdaq Commodities launched the N2EX power market in the United Kingdom. On 20th January 2016 Nord Pool AS was rebranded to Nord Pool. On 27th August 2019 Nord Pool started trading in France, Germany, Luxembourg, Belgium, Austria and the Netherlands. On 5th December 2019 Euronext announced that it would acquire 66% of Nord Pool, this was officialised on 15th January 2020.

Nordic Electricity Market

The Nordic power system consists of different sources, such as hydro, nuclear and wind power as its main sources. Many industries in the Nordic region rely heavily on energy, furthermore, the Nordic region has a large share of electricity heated houses. As a result, the electricity consumption and the electricity's share of total power use is higher than in the rest of the EU. Longer winters and shorter summers are reasons why the electricity consumption is greater in Nordic countries than in the EU. The weather has a major influence on electricity consumption, so during the winter, there is a higher electricity consumption and higher demand, whereas during the summer there is a lower electricity demand as well as lower electricity consumption. Over half of the electricity production is generated from hydropower, 20% from nuclear, 15% from fossil fuels and the rest from other sources. Furthermore, the pumped storage capacities available in Norway means that Scandinavia has considerable reservoirs at its dispose. All of these factors mean that electricity prices in the Nordic countries are far lower than in other parts of the world.

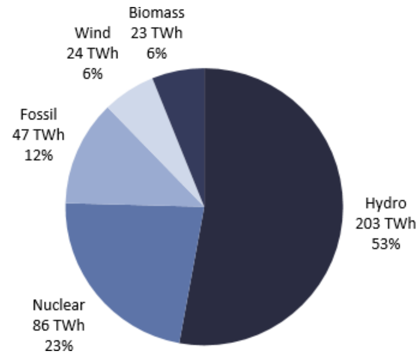


FIGURE 4: SOURCE: ENTSO-E [ONLINE}, AVAILABLE AT: HTTP://WWW.NORDICENERGYREGULATORS.ORG/WP-CONTENT/UPLOADS/2014/06/NORDIC-MARKET-REPORT-2014.PDF
[ACCESSED ON 31/03/2020]

Figure 4 highlights how much different sources contribute to electricity consumption in Nordic companies.

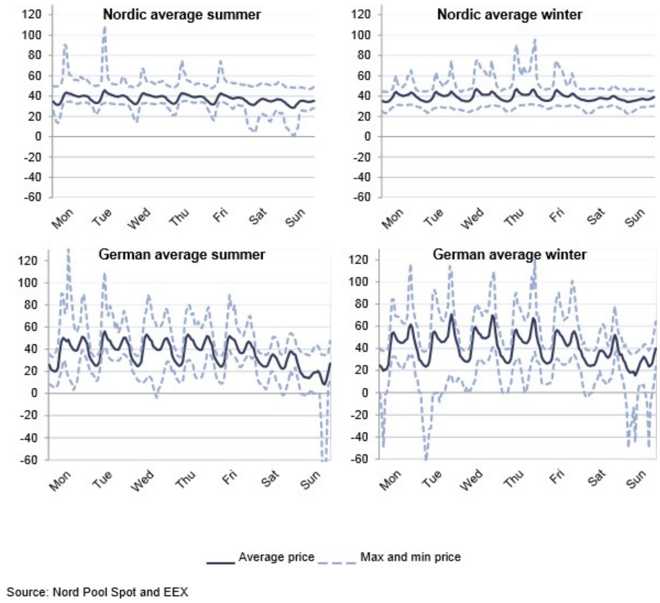


FIGURE 5:NORDIC SYSTEM PRICE AND GERMAN WHOLESALE PRICE - AVERAGE, MAXIMUM AND MINIMUM HOURLY PRICES DURING THE SUMMER WEEKS (14-39, 2013), EUR/MWH AND AVERAGE, MAXIMUM AND MINIMUM HOURLY PRICES DURING THE WINTER WEEKS (40-13, 2013), EUR/MW. AVAILABLE AT HTTP://WWW.NORDICENERGYREGULATORS.ORG/WP- CONTENT/UPLOADS/2014/06/NORDIC-MARKET-REPORT-2014.PDF [ACCESSED ON 31/03/2020]

Figure 5 above highlights electricity prices in Nordic countries compared to that of Germany, as we can see the average price of electricity is far lower in Nordic countries than in Germany.

Overall, the Nordic electricity market has been a huge success, however, the Danish Energy Agreement, established in 2012, set a few primary targets by the end of 2020. These primary targets are "Reduce emissions by 34%, Increase renewable energy penetration by 35% and improve energy efficiency by 7.6%". The success of the Nordic electricity market stems from the fact that not only are prices lower, but a lot of the electricity produced is from renewable sources.

2.4 Stationarity

Stationarity is an important aspect of time series analysis. A stationary time series is one whose statistical properties such as mean, variance, autocorrelation (which we will define later on in the thesis) etc. is constant over time. The concept of stationarity is important because many statistical tests and models rely on it. This form of stationarity is known as a weak form of stationarity or covariance/mean stationarity (*Liverani, 2020).* A more formal definition of weak stationarity is given below.

Definition 1 (W Yoo): We say that $\{X_t\}$ is weakly stationary if

1) *$\mu(t)$ is independent of t. Where μ(t) is the mean function.*
2) *$\gamma(t+h, t)$ is independent of t for each h. The function $\gamma(.)$ is the autocovariance function.*
3) *The autocorrelation function of $\{X_t\}$ at lag h is ρ(h)=$\frac{\Upsilon(h)}{\Upsilon(0)}$=Corr($X_{t+h}, X_t$)*

The above tells us that the covariance function does not depend on shifts in time, therefore we can rewrite 2) as

$$\gamma(t+h, t) = \gamma(h, 0) =: \gamma(h)$$

For stationary time series the autocorrelation function drops to zero relatively quickly, whereas, for non-stationary time series, the autocorrelation function decreases slowly.

As previously mentioned, the above is known as the autocovariance function. We will give a more rigorous definition of the autocovariance function later on.

Another type of stationarity is strict stationarity, below we have the formal definition for strict stationary.

Definition 2: A time series $\{X_t\}$ is called strictly stationary if for all sets of indices $\{t_1, t_2, t_3, \ldots, t_k\}$ and all integers *h,* the joint distribution of $(X_{t_1}, X_{t_2}, X_{t_3}, \ldots, X_{t_k})$ is the same as the joint distribution of $(X_{t_{1+h}}, X_{t_{2+h}}, X_{t_{3+h}}, \ldots, X_{t_{k+h}})$. It can be written as:

$$\left(X_{t_1}, X_{t_2}, X_{t_3}, \ldots X_{t_k}\right) =^d \left(X_{t_{1+h}}, X_{t_{2+h}}, X_{t_{3+h}}, \ldots X_{t_{k+h}}\right)$$

Similar to how for stationarity we need the mean, variance and covariance to be constant over time, we also need the joint distribution $(X_{t_1}, X_{t_2}, X_{t_3}, \ldots, X_{t_k})$ to be constant over time.

2.5 Returns

When analysing and studying a financial data set, such as stock price of a company etc, it is often a better idea to analyse the log-returns of a dataset. This is because log-returns often has several benefits to it. One benefit is normalisation. What this means is that all variables are measured in a comparable metric, this allows us to evaluate analytic relationships amongst two or more variables. The concept of normalisation is a requirement for many statistical analysis and machine learning techniques. Another benefit is the concept of additive. What this means is that with log returns you can add the values, whereas you can't add simple returns, they need to be compounded first. These are just two benefits out of many. Next, we will define the returns, there are two definitions of returns.

Definition 3 (W Yoo 2019): Simple return is defined as

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}}$$

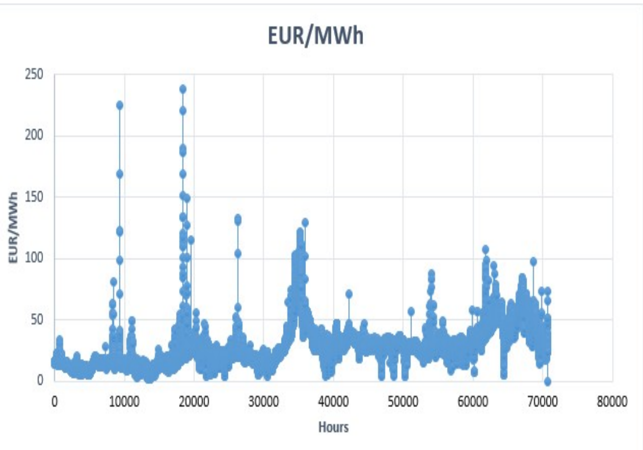
Where P_t is the price of an asset at time t, and P_{t-1} is the price of the asset at time t-1.

Definition 4 (W Yoo): Log-return is defined as

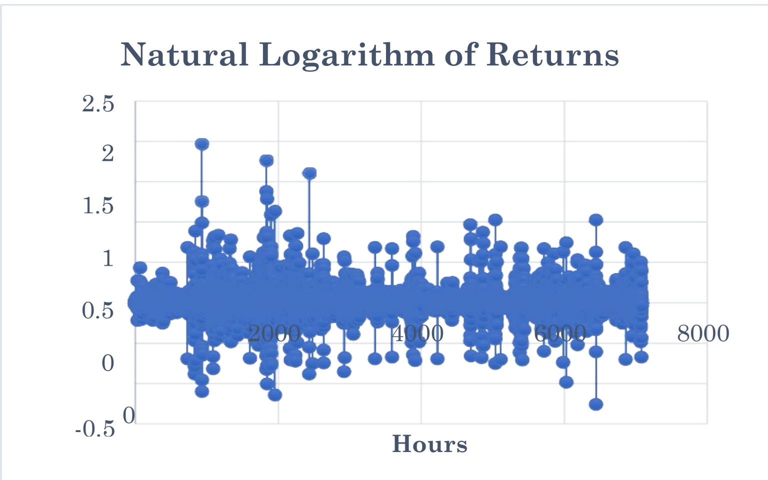
$$r_t = \ln(1 + R_t) = \ln\left(\frac{X_t}{X_{t-1}}\right) = \ln(X_t) - \ln(X_{t-1})$$

Where $1 + R_t = \frac{X_t}{X_{t-1}}$ (also we note that $\ln(1 + R_t) \sim R_t$ for small R_t therefore $r_t \sim R_t$).

The graph below is a plot of the electricity prices per hour from 1st January 1999 to 26th January 2007 (roughly 72000 data points) from Nord Pool. From the data below, we can see that prices tend to fluctuate to a relatively large extent. Here the x-axis represents consecutive hours and the y-axis represents electrical prices.



We next look at the logarithm return



From the logarithm plot above we see very large fluctuations, which consists of high peaks and low peaks. These peaks represent extreme price fluctuations, i.e. high peak represents an upward price fluctuation, whereas a low peak represents a downward price fluctuation. By analysing the natural logarithm of returns we can see price fluctuations, some extremely high, others extremely low. Further down the report, we shall see when these extremities occurred, more precisely what hour these extremities occurred in, and the impact these extremities had on the electricity prices. To do this we need to condition the full data set per hour. To do this we use a combination of the "=OFFSET" and "=ROW" functions.

F3 =OFFSET(B2,(ROW(A2)-1)*24,0)

	A	B	C	D	E	F	G
1	Hours	EUR/MWh	Log-Returns		DataPoints	12am	Log Return 12am
2	1	15.29	0		1	15.29	0
3	2	15.14	-0.009858772		2	15.28	0.009204536
4	3	14.9	-0.015979035		3	15.29	0.002619517
5	4	14.73	-0.011474982		4	15.24	-0.001966569
6	5	14.68	-0.003400207		5	14.96	-0.034814682
7	6	14.72	0.00272109		6	14.84	-0.031179972
8	7	14.82	0.006770507		7	15.17	0.000659413
9	8	14.84	0.001348618		8	15.92	0.015190165
10	9	14.91	0.004705891		9	16.17	0.016209831
11	10	15.15	0.015968403		10	16.45	0.008547061
12	11	15.39	0.015717416		11	16.83	0.026491615
13	12	15.52	0.008411567		12	16.59	-0.015550553
14	13	15.59	0.004500168		13	16.48	-0.004842624
15	14	15.6	0.000641231		14	16.28	-0.006734032
16	15	15.75	0.009569451		15	15.72	-0.028223887
17	16	15.9	0.009478744		16	15.49	-0.004508864
18	17	16.07	0.010635071		17	14.83	-0.031200672
19	18	16.07	0		18	14.94	0.013477293
20	19	15.96	-0.006868588		19	14.72	-0.026149663
21	20	15.85	-0.006916092		20	14.01	-0.056880796
22	21	15.7	-0.009508788		21	13.98	-0.0226318
23	22	15.77	0.004448689		22	13.8	-0.017241806
24	23	15.5	-0.017269377		23	14.08	0.054738217
25	24	15.14	-0.023499776		24	13.84	-0.005763705

Sheet1

Column A up until column C represents the consecutive hours, electrical prices and the log return for the full data set respectively. Cell A2 has value "1" which represents midnight on January 1st, 1999. Now we need to filter out every 12 am data (from 1999 to 2007) from the original data. The first column of the original data and the first column for the 12 am data points are the same. To get the value for January 1st, 2000 (12am), we use "=OFFSET (B2, (ROW (A2)-1)*24, 0)". What the offset function is doing here is, the position of cell B2 is fixed, and the "ROW" function is moving 24 positions down from cell B2 and returning the value that is in that specific cell i.e. cell B26, which is the value 15.28. Now use autofill to complete all of column F. For column G repeat the same process, using "=OFFSET (C2, (ROW (A2)-1)*24, 0)". This filters out all the data corresponding to midnight. Repeat the same process for 1 am up until 11 pm.

3.0 Analysis

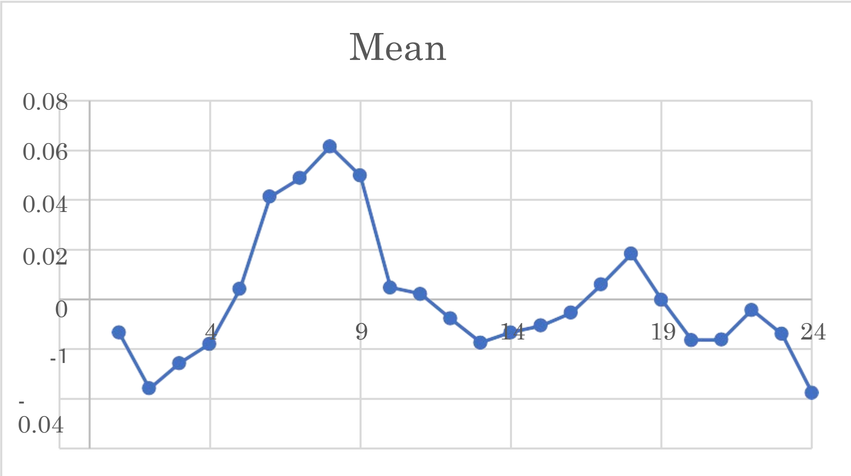
3.1 Mean/Variance and conditioned Time Series Plots

We will now be analysing the mean and variance of the log-returns of the whole data set as well as the conditioned data set. I have mentioned above how to filter out the conditioned data from the whole data set using the "=OFFSET" function. We will discuss the mean and variance of the conditioned data set in more detail later on. Now to get the mean value of the whole data set we use the function "=AVERAGE", and to get the variance we use the function "=VAR". The mean of the whole data set is 0.00000669257, which is very small, therefore can be considered as zero. What this means is that on average the rate of return of the electricity prices for the whole data set was zero.

Time	12am	1am	2am	3am	4am
Mean	-0.013418417	-0.035791617	-0.02568097	-0.01788281	0.004151762
5am	6am	7am	8am	9am	10am
0.041234304	0.048697344	0.061423478	0.049650698	0.004674251	0.002181059
11am	12pm	1pm	2pm	3pm	4pm
-0.007824629	-0.01743224	-0.013289305	-0.01060853	-0.005467997	0.005864162
5pm	6pm	7pm	8pm	9pm	10pm
0.018271985	-0.000148934	-0.016335204	-0.016300567	-0.004282033	-0.013909881
11 pm					
-0.037615285					

Now let's look at the mean of the conditioned data set, the table above shows us the mean value of the data set conditioned per hour. Now straight away what we notice is that the mean values of the data set conditioned per hour is a lot larger than the mean value of the whole data (which was considered to be zero). The smallest average return was -0.037615285, which corresponded to 11 pm, and the highest average return was 0.061423478, which corresponded to 7 am. In simple terms, negative average values reflect a reduction in price and a positive average value reflects an increase in price. When analysing the data, we see that between 12 am-3 am we have negative mean values, which is expected as most people would not be awake during that time, so the demand for electricity during those times will be low, hence price would go down. Now from 4 am-10 am the mean values are positive, this comes across as somewhat surprising, because between 4 am-6 am we would expect the mean to be negative since most people would still be asleep, hence not much electricity is being used. From 11 am-3 pm we also have negative means, which is also a bit of surprise since most people would be either at work or school, therefore we would be utilising a lot of electricity. From 4 pm to 5 pm the average values are positive which is expected, however from 6 pm- 8 pm the average values were negative, which is shocking because around 6 pm it would be rush hour, furthermore, it would be dark outside, so street lights, home lights etc. would be switched on. From 9 pm-11 pm the average mean values were negative, which seems sensible, as most people would be asleep at that time.

The next step would be to plot the mean values, and to observe if there are any obvious patterns.



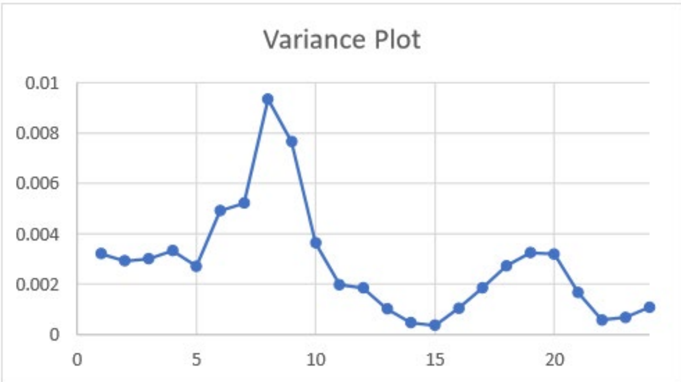
What we notice is that the mean is changing constantly, furthermore, what we see is that there is a fluctuating pattern, the fluctuations start of fairly large, but gets smaller throughout the day. Next, we will analyse the variance of the whole data set and the variance of the data set conditioned hourly. The variance of the whole data set was 0.003491829, as we can see the variance is very small which means that generally speaking the log-return values are very close to the mean value i.e. the data points do not deviate away from the mean.

Time	12am	1am	2am	3am	4am
Variance	0.003212052	0.002927647	0.003010217	0.00333136	0.002718644
5am	6am	7am	8am	9am	10am
0.00491896	0.00522026	0.009349064	0.007659255	0.003642484	0.001984682
11am	12pm	1pm	2pm	3pm	4pm
0.001857343	0.001020176	0.000471443	0.000367755	0.001046861	0.001863455
5pm	6pm	7pm	8pm	9pm	10pm
0.002734446	0.003248869	0.003205587	0.001671708	0.000580473	0.00068256
11 pm					
0.001094724					

The variance is a measure of volatility, in our instance, we are measuring how volatile the price of electricity is. The smaller the variance is then that means the data points are more evenly distributed and there are a minimal number of outlier's present. The outliers are what leads to fluctuations in the data set. Variance of zero means that the data is completely evenly distributed, and no outliers exist. From the table above what we see is that the variance ranges from 0.000367755 (smallest variance corresponding to 2 pm) to 0.009349064 (largest variance corresponding to 7 am). This means that price tends to fluctuate the least at 2 pm and the most at 7 am.

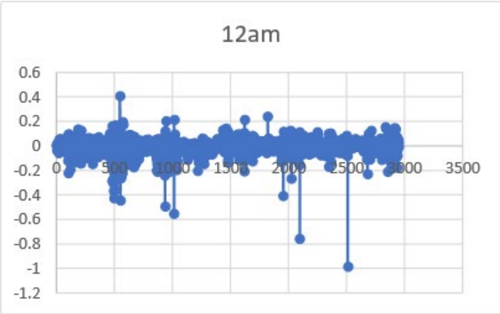
Previously I had mentioned how the average electrical price from 6 pm-8 pm were negative, which came as a surprise since most electricity would be used around that time, but now with the variance we can somewhat explain why that may be the case. The variance from 6 pm-8 pm, were 0.003, 0.003 and 0.002 respectively, which is fairly high when compared to the variance of the conditioned time. As previously mentioned, a higher variance means more spread in the data, hence the chances of outliers being present is very high. So, this means one possible reason for a negative mean around 6 pm-8 pm could be the fact that there were some outliers present that were highly negative, which consequently pulled the mean down to a negative value.

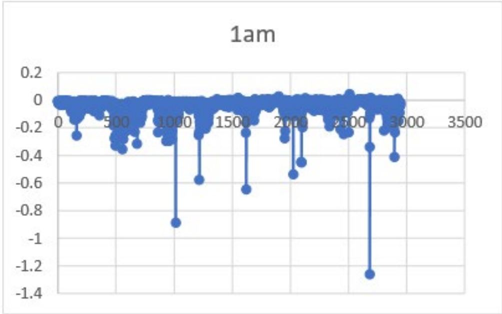
I will now plot the variance to see if there are any visible trends.



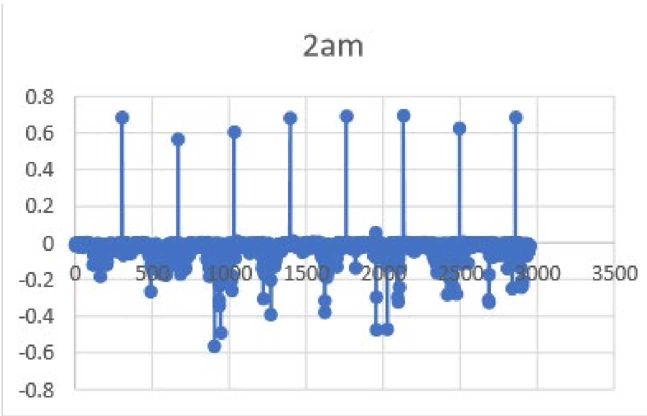
From the variance plot above what we can see is that between 1 am and 4 am the fluctuations are at a minimum, which means there isn't much of a change with the variance. However, from 5 am to 11 am there is a huge fluctuation in terms of the variance. Finally, from 11 am onwards the variance is almost periodic. Similar to the mean plot, the variance is not always constant, there is a degree of fluctuation present within the data set.

We will now plot the hourly data set for each time, here we will analyse in more detail how the mean and variance change from each time, and we will look at any pattern within the data, i.e. whether there are any positive or negative extreme events are present or not.

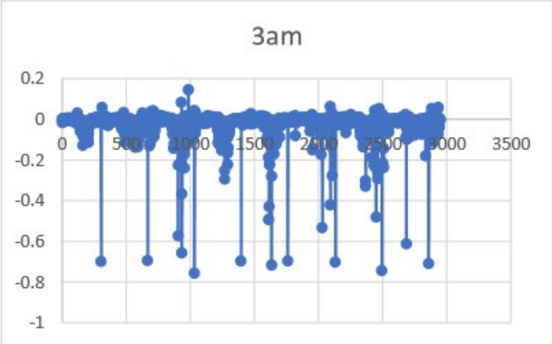


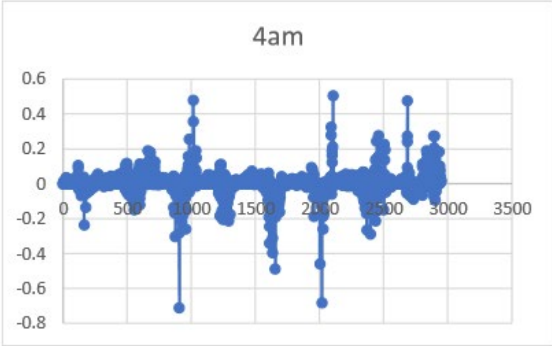


In the 12 am plot there does not seem to be any set pattern in terms of the extremities, however towards the end of the 12 am plot we notice that the negative extremities become a lot more noticeable. The mean at 12 am was - 0.013418417, which meant that between 11 pm and 12 am the average electricity prices went down, which as previously mentioned makes sense, as most people would be asleep. The variance was 0.003212052, which was higher than the variance at 11 pm (0.001094724, this means that the data set at 12 am is more disperse than the data set at 11 pm. At 1 am straight away what we notice is that there are negative extremities which are visible. The mean 1 am is -0.035791617, which means that from 12 am till 1 am the average priced reduced. The rate at which the electrical price decreased from 12-1 am is greater than the rate at which the electrical price decreased from 11 pm-12 am. The variance at 1 am was 0.002927647, which compared to 12 am tells us that the data is less dispersed, which is evident from the graph, because most of the data have a negative spike, and there are far fewer fluctuations present.

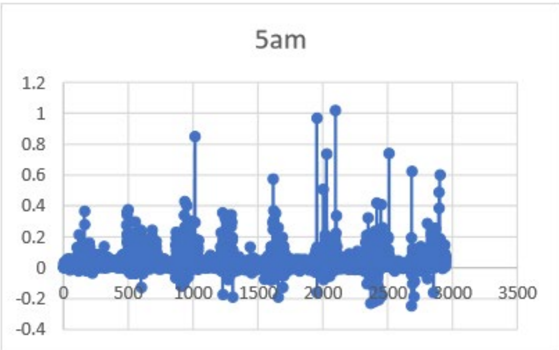


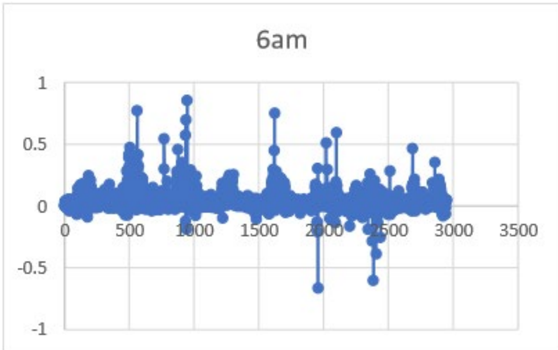
The plot of 2 am is somewhat interesting, because immediately what we notice are extreme positive values, which seem to be periodic. What this can imply is that at a specific day, every year there seems to be a huge price increase. One possible explanation could be due to a computer update, for example, there might have been power network updates, which makes sense because at 2 am most people would be asleep, so it would be logical to have a power upgrade, without causing any disruption at 2 am. The mean at 2 am was -0.026 (3dp), which means on average price of electricity goes down from 1 am to 2 am, and the variance is 0.003 (3dp).



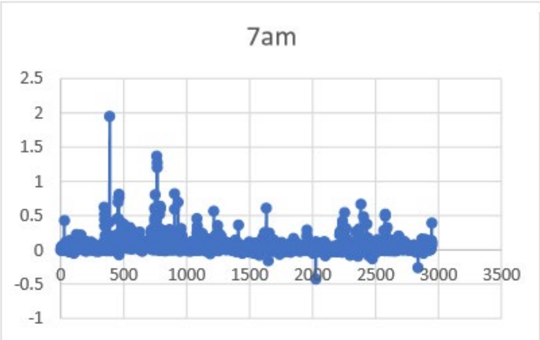


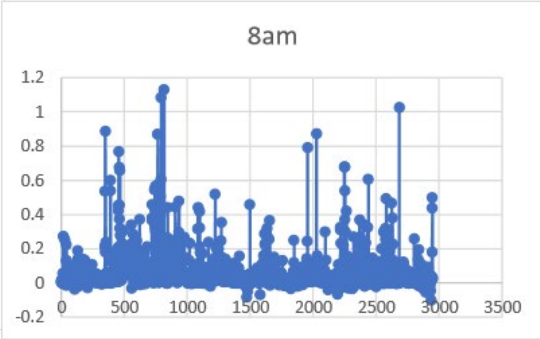
In the 3 am plot we get negative extreme points, which means that from 2 am to 3 am there seems to be a price decrease. Unlike the 2 am plot, there is no visible periodic pattern present. The mean of the 3 am plot is -0.018 (3dp), and the variance is the same as that of 2 am to 3dp, i.e. the variance is 0.003, so the dispersion is constant from 1 am to 3 am. When looking at the 3 am plot we notice that the negative peaks in the 3 am data is similar to that of the positive peaks of 2 am, in terms of the modulus both data points have roughly a value of 0.7, thus having a constant variance from 1 am to 2 am, makes somewhat of a sense. At 4 am we get a mixture of positive and negative fluctuations; the negative extreme values seem to be greater than the positive extreme values. The mean at 4 am was 0.004 (3dp), which means on average there was a slight increase in electricity prices from 3 am to 4 am, however, the variance remained constant (3dp), i.e. the variance was 0.003, this means that price fluctuations were constant between 3 am and 4 am.



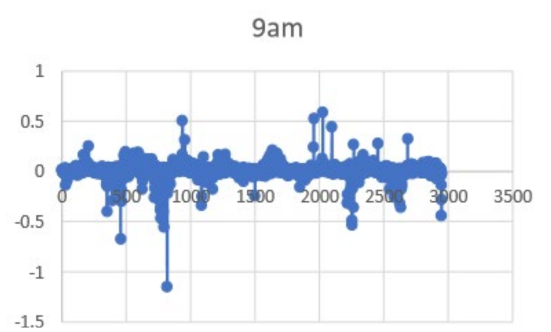


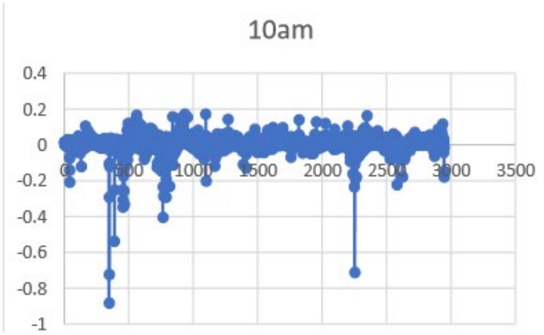
At 5 am the positive spikes seem to be more noticeable, so generally there seems to be some sort of a price increase, however, the mean (0.041 to 3dp), shows that the price remained fairly constant from 4 am to 5 am, however, the variance at 5 am was 0.005 (3dp), which is larger than that of 4 am, so the data seems to be more dispersed than the data points at 4 am. At 6 am we have more positive extreme events, which indicates that there is a price increase from 5 am to 6 am, which is highlighted from the fact that the mean at 6 am (0.05 3dp) is higher than that of 5 am (0.04 3dp). The variance to 3dp remained constant from 5 am to 6 am (0.005).

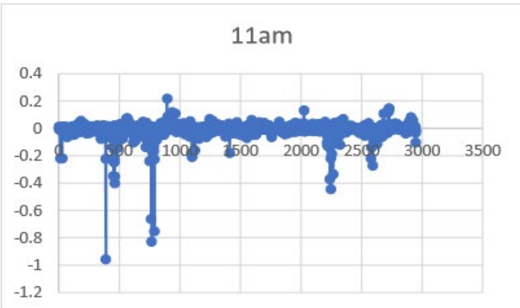


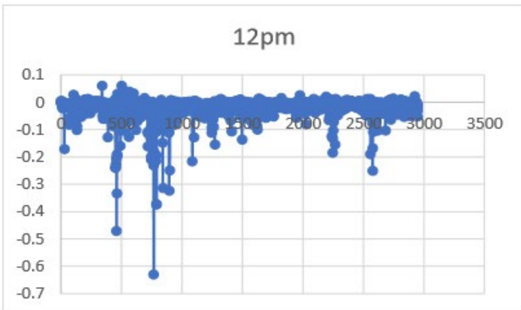


At 7 am we notice positive events taking place, and a few positive extreme events, but when comparing with the 6 am data we notice that the values of the extreme positive events is higher than that of 6 am, upon analysing we see that the extreme positive events are above 1, whereas at 6 am all the values are below 1. This may explain why the variance at 7 am is higher than that of 6 am. The variance at 7 am is 0.009. The mean at 7 am is higher than that of 6 am, this means that the price of electricity increased from 6 am to 7 am. By observing the 8 am data set we notice that there seems to be far more positive events compared to any other data set. We would expect more positive events to occur (as the graph shows), because at 8 am most people would be getting ready for work or getting ready for school, thus more electricity would be utilised. It would seem that the fluctuations of the data points at 8 am would be far greater than that of 7 am, however that is not the case, in reality, the variance at 7 am is far greater than that of 8 am. The variance at 8 am was 0.007 (3dp). Just as the previous case, one reason the fluctuations may be small is because the extreme events at 8 am have a value smaller than 1.2, whereas at 7 am some values are larger than 1.2, so at 7 am the data points are more spread out compared to 8 am. The mean at 8 am is 0.050 (3dp), which is smaller than that of 7 am, this means that in general, the average price of electricity reduces from 7 am to 8 am, which comes as somewhat of a surprise, since most people would be up at 8, thus utilising more electricity, so we would expect electricity price to be higher at8.



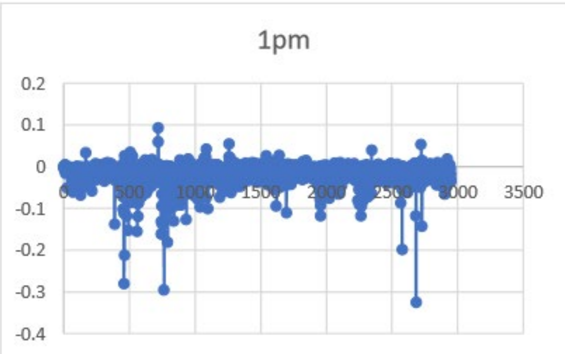


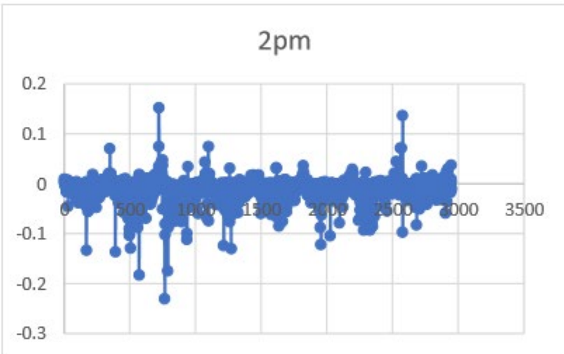




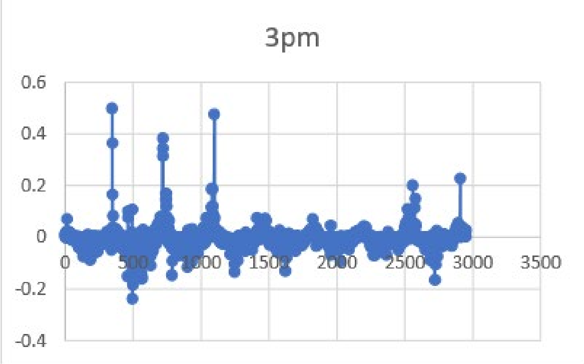
From 9 am till 12 pm there seems to be more negative events occurring than that of positive events, thus indicating that we should expect some sort of price decrease occurring. At 9 am the mean is 0.005 (3dp), which is considerably smaller than the mean at 8 am (which was 0.05), thus indicating that there was a price decrease from 8 am to 9 am. The variance at 8 am was 0.007 (3dp), which is smaller than the variance at 7 am, thus there are fewer price fluctuations at 8 am than there are at 7 am. At 10 am the negative events seem to be more apparent than that of 9 am. Thus, we would expect there to be a price decrease. This is supported by the fact that the mean is smaller at 10 am, the mean is 0.002. This means from 9 am to 10 am there is a price decrease. The variance is smaller at 10 am than 9 am. The variance at 10 am is0.001.

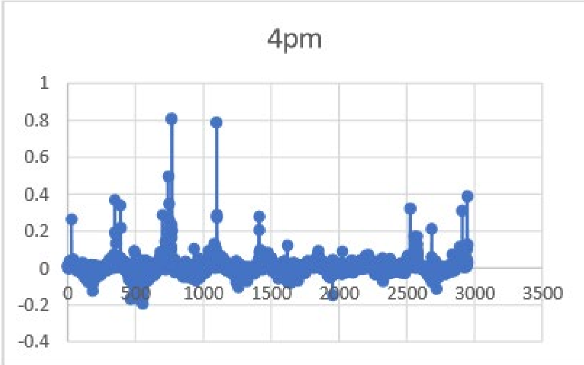
Similarly, to the 9 am and 10 am plot we notice negative events occurring, the mean at 11 am is -0.008, thus there is a price decrease from 10 am till 11 am. Just by observing the 11 am plot, we would expect the variance to be fairly small, since there seems to be a few fluctuations present which is indeed the case. The variance at 11 am is 0.001 (3dp), which is fairly small, furthermore, the variance is the same (to 3dp) as the variance at 10 am. This means that from 10 am to 11 am price fluctuations remain fairly constant. At 12 pm we also notice negative events occurring. The mean at 12 pm is -0.017 (3dp), which means overall there is a price decrease, and when comparing to the mean at 11 am price decreases from 11 am till 12 pm. Just like the 10 am and 11 am plots, the variance is 0.001 (3dp), which means that price remained fairly constant from 10 am till 12 pm.



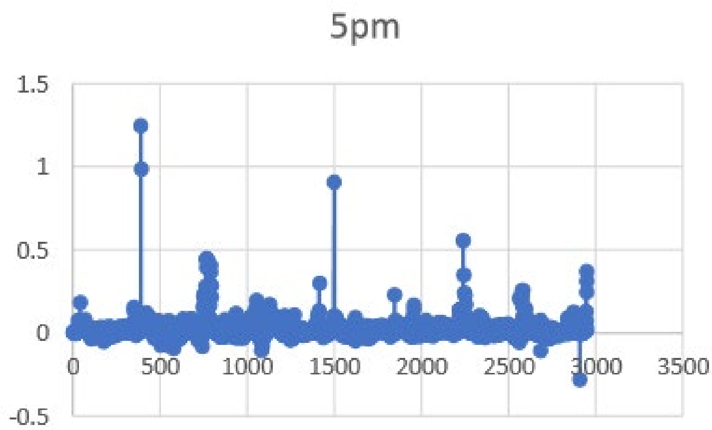


When analysing the 1 pm plot we notice that it follows a similar behavioural pattern to that of the 12 pm plot. This is indeed supported by the fact that the mean remained constant from 12 pm to 1 pm (-0.01 to 3dp), this means that from 12 pm till 1 pm prices remained fairly constant. By looking at the y-axis of the 1 pm plot we would expect the variance to be very small, because the y-axis ranges from roughly 0.1 till about -0.35. This is indeed the case as the variance is 0.0005, which up until now the smallest variance we have encountered. This indicates that price fluctuations are at a minimal at 1 pm. From 9 am till 1 pm the plots seem to have more negative events occurring. However the 2 pm plot has a mixture of positive and negative events occurring, but the negative events seem to be more apparent and drastic compared to the positive events, as a result, we would expect the mean to be small or even negative, which is indeed the case. The mean at 2 pm is -0.01 (3dp), which tells us that the price of electricity between 1 pm and 2 pm remained constant. When analysing the graph, we see that the fluctuations for 2 pm seem to be fairly similar to that of 1 pm, which is indeed the case, the variance for 2 pm is 0.0003, which is slightly smaller than the variance of 1 pm.

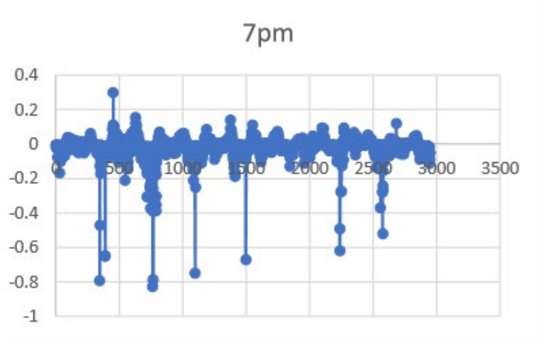


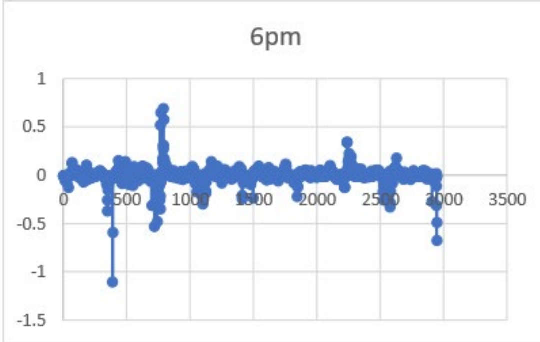


At 3 pm we see a change in extreme events, there seems to be more positive extreme events occurring at 3 pm, thus we would expect the price to increase from 2 pm till 3 pm. This is indeed the case as the mean price of electricity at 3 pm is -0.005, which is larger than the mean at 2 pm which was -0.01. At 4 pm we also expect a similar trend to follow, because the positive extreme events seem to be far greater than that of 3 pm. This is the case as the mean at 4 pm is 0.005, which also indicates that the price of electricity goes up from 3 pm to 4 pm. This makes sense as 4 pm is usually the time when children get home from school, so a lot more electricity will be utilised during that period.

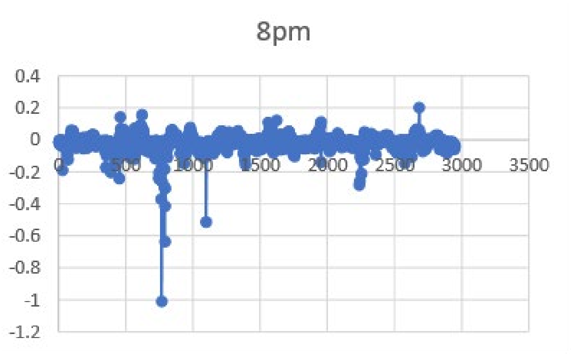


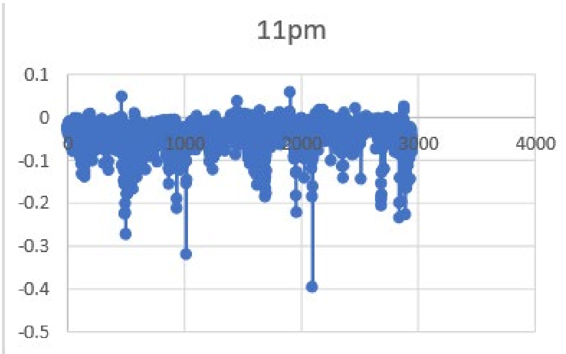
At 5 pm we notice a similar pattern, but we would expect the mean to be higher at 5 pm than that of 4 pm, because at 5 pm there seems to be far more positive values, but hardly any visible negative extremities occurring. This is supported by the fact that the mean at 5 pm is higher than the mean at 4 pm. The mean at 5 pm is 0.01. Therefore, electricity prices increased from 4 pm to 5 pm. When analysing fluctuations, we look at the y-axis for the graphs, upon inspection, it becomes fairly apparent that we would expect the variance to be highest at 5 pm followed by 4 pm then finally 3 pm. This is because when analysing the y-axis, we notice that the values for 3 pm ranges from [-0.2-0.5], for 4 pm the values range from [-0.2 - 0.8], and finally for 5 pm the values ranged from [-0.2 – 1.3]. From the ranges, we can see that fluctuations were highest at 5 pm then 4 pm and then 3 pm. This is supported by the variance. The variance at 5 pm was 0.003, followed by the variance of 4 pm which was 0.002, and then finally 0.001 for 3 pm. Generally speaking, the variance for all three graphs were fairly small and very close to one another (difference of 0.001), this means that from 3 pm till 5 pm there was not much change in the price during that period.



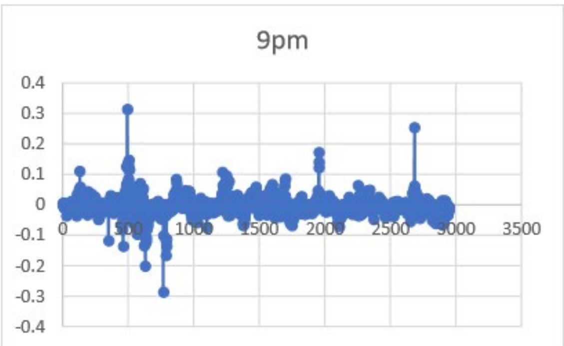


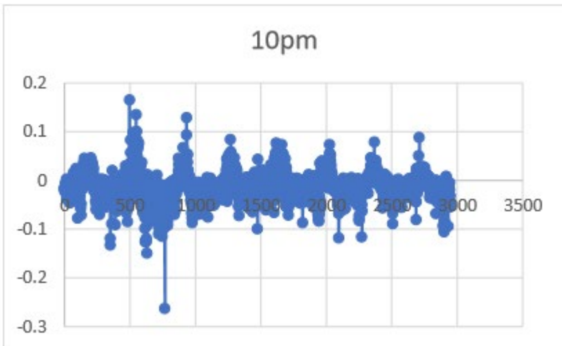
From 6 pm to 7 pm we have negative and positive extreme events, however, the negative events at 7 pm are a lot more noticeable and apparent. The mean value at 6 pm is -0.0001, which is very close to 0, this means that from 5 pm to 6 pm price decreased ever so slightly, the change is so small that it could be considered that the price pretty much remained constant from 5 pm to 6 pm. The mean at 7 pm was -0.01, which means that from 6 pm to 7 pm the price of electricity ended up reducing. From 6 pm till 7 pm the variance remained the same (to 3dp), the variance at 6 pm and 7 pm was 0.003, which means that price fluctuation remained constant from 6 pm to 7 pm.





At 8 pm and 11 pm we have negative and positive extreme events, however, the negative events at 6 pm and 11 pm are a lot more noticeable and apparent. The mean value at 8 pm is -0.01, this means that from 7 pm to 8 pm price decreased ever so slightly. The mean at 11 pm was -0.03, which means that from 10 pm to 11 pm price of electricity ended up reducing. From 6 pm till 7 pm the variance remained the same (to 3dp), the variance at 8 pm and 11 pm was 0.02 and 0.001, which means that price fluctuation was very similar at 8 pm and 11 pm.

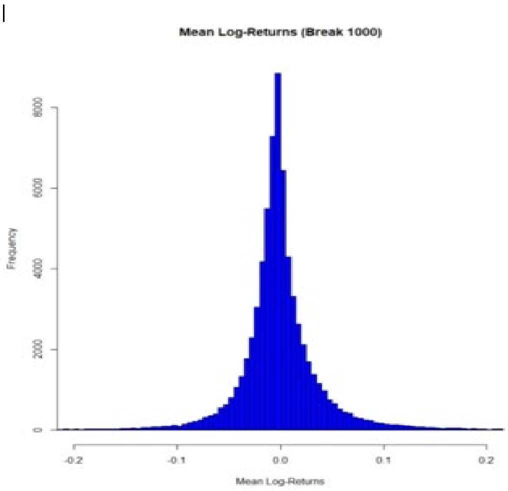


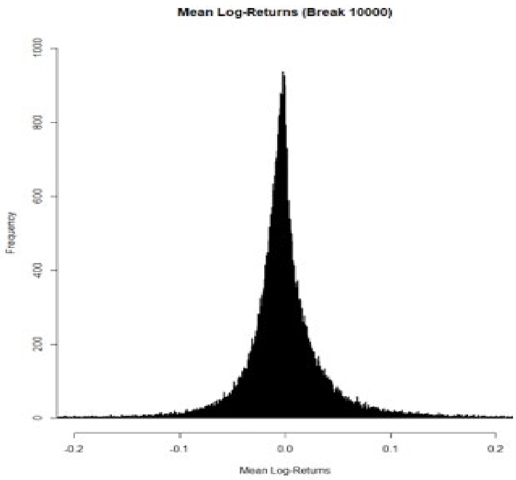


From 9 pm to 10 pm we have negative extreme events. The mean value at 6 pm is -0.004, which is very close to 0, this means that from 8 pm to 9 pm price decreased ever so slightly, the change is so small that it could be considered that the price pretty much remained constant from 8 pm to 9 pm. The mean at 10 pm was -0.01, which means that from 9 pm to 10 pm price of electricity ended up reducing. At 9 pm the variance was 0.0005, which means that price fluctuations were very small, it was so small that it could be considered that the prices remained fairly constant. The variance at 10 pm was slightly larger, it was 0.0006, which is still very small, hence it could be considered that the price remained fairly constant from 9 pm to 10 pm.

4.0 Probability Density Functions

We will now be plotting the probability density function (PDF) of the whole data set and the data set conditioned per hour. The PDF can be plotted in terms of a histogram in RStudio, however, to plot the histogram in RStudio we need to ensure that the whole data set and the data set conditioned per hour are centred around the mean. This can be done by simply subtracting the log-returns of the data set by the average of the log-returns. For example, if our data for log-returns is in column D, to calculate the mean log-returns we subtract column D with the average of column D. We repeat this step for the conditioned data set. To plot the histogram in RStudio we use the function hist(….), in the histogram function we specify the breakpoints, x boundaries and y boundaries. Ideally, we would like the PDF to be smoothed out, this requires trial and error.





The image on the left is the mean Log-Returns with 1000 breaks and with the y-axis ranging from [0:9000]. As we can see the plot on the left could be smoothed out. Which brings us to the plot on the right. The plot on the right has 10000 breaks and the y-axis ranges from [0:1000]. As we can see the plot on the right is far smoother. However, we should note that different data sets will require different breakpoints to turn out smooth.

When we plot the probability, density functions we notice that some distributions have long right tails and some distributions have long left tails. The distributions that contain long right or long left tail, we will plot the log-log plot. Once again, the log-log plot can be calculated via

RStudio, by taking the logarithm (base 10) of the x and y-axis. The log-log plot tells us whether a power law exists or not. If the log-log plot is more or less a straight line, then we say that a power law in the form of $y = a * x^n$ exists (where a is a constant). The next plot that we will be analysing is the semi-log plot. If our log-log plot is not close to a straight line, then we will plot the semi-log plot. The semi-log plot is calculated by taking the logarithm (base 10) of the y-axis, whereas the x-axis remains unchanged. Just as in the case of the log-log plot we will also be analysing whether the semi-log plot is more or less a straight line. If the semi-log plot appears in a straight line, then we say that an exponential fit in the form of $y = a * 10^{mx}$ exists (where m is the gradient of the line) for the right/left tail of the distribution. However, if a straight line does not exist then we claim that an exponential fit does not exist.

Skewness

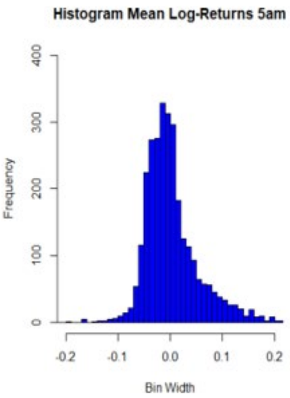
As previously mentioned, we will analyse the PDF of the plots which have a right or left tail. One way of grouping the PDF's into a right tail and left tail is through the measure of skewness. What skewness helps us see is the degree of symmetry of a plot. Most statistical programmes will have a means of calculating skewness without any computations. One of the easiest ways to calculate skewness is through the use of the function "SKEW(…)" in excel. The skew function in excel will return a value, which can be interpreted.

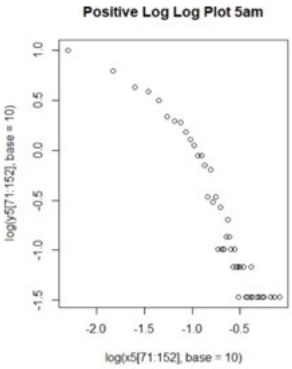
The following interpretations will allow us to classify whether the plots are symmetric, right-tailed or left tailed
(Taken from Bulmer, 1979, p. 63).

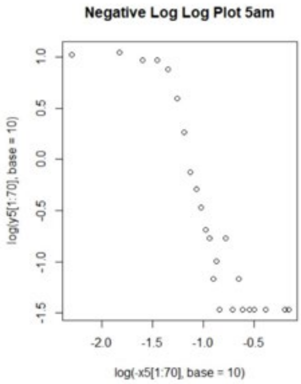
- If skewness is greater than 1 then we say the distribution is highly (positively) skewed, i.e. the right tail is longer than the left. If the skewness is less than -1 then the distribution is also highly (negatively) skewed, i.e. the left tail is longer than the right tail.
- If skewness is between -1 and -0.5, or between 0.5 and 1, then we say that the distribution is fairly symmetrical, or we can say that the distribution is moderately skewed.
- If skewness is between -0.5 and 0.5 then we say that the distribution is fairly/approximately symmetrical.

4.1 Log-Log Plot and Semi Log Plot (Right Tail)

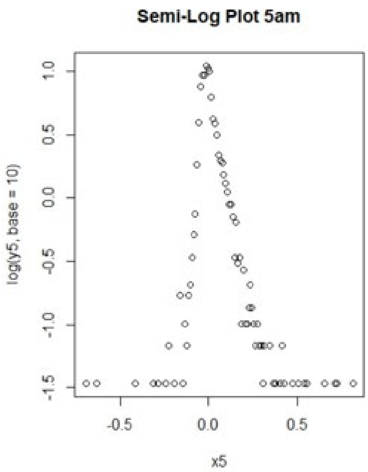
We will start by analysing plots that appear to be right-tailed, and further back this up by calculating the skewness. As mentioned earlier we will analyse the mean values of each data conditioned per hour. When it comes to the log-log plot, I have split it up into two types, we have the positive and the negative log-log plot. These variations arise since we cannot take the logarithm of negative values, some of our x values may be negative, hence, to take a logarithm of the negative values we would have to take the logarithm of minus the negative values. For example, if our x value is -3, to take the logarithm of that we would take log(-(-3)). Once again, the logarithm is to the base 10. If a power law does not exist, we will analyse the semi-log plot. We will first consider the 5 am data set.



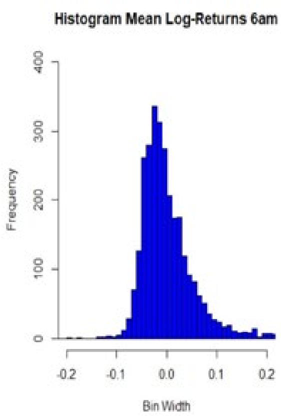


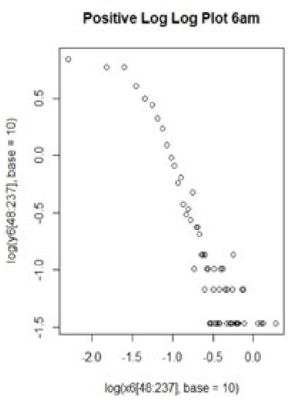


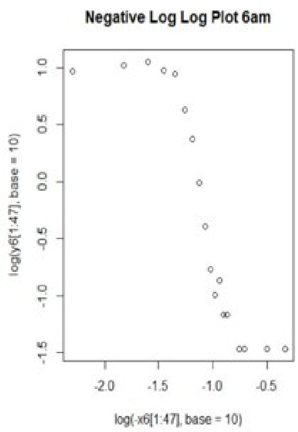
Upon inspection, it seems like the 5 am plot is right tail, which is indeed the case as the value of the skew is 2.302 (to 3dp), which means that the distribution is highly (positively) skewed. When looking at the positive log-log plot it becomes obvious that the points do not lay on a straight line, and its, even more, obvious when looking at the negative log-log plot that the points are more of a curvature rather than a straight line. Hence, we conclude that a power law does not exist. Since a power law does not exist, hence we will take a look at the semi-log plot.



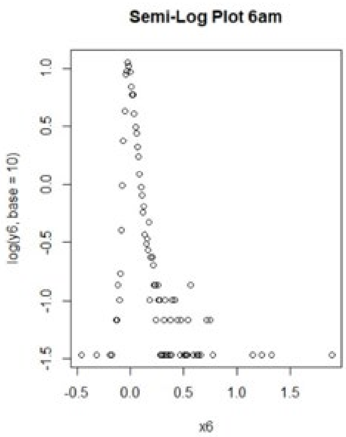
The semi-log plot is somewhat interesting, we can break down the semi-log plot into two instances. The first instance being the positive half of the graph, and the second being the negative portion of the graph. When analysing the positive half from roughly points [0.0, 0.3] there seems to be a straight which is a clear indication of an exponential fit, however after around 0.3, we see that the points flatten out. The same trend follows for that of the negative portion of the graph, from points [0.0,-0.2] we have roughly a straight line, but after -0.2 the points seem to flatten up. This indicates that overall, there doesn't seem to be a strong exponential fit.



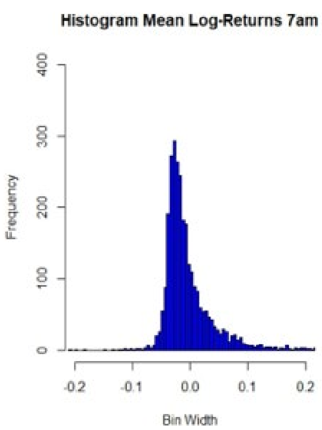


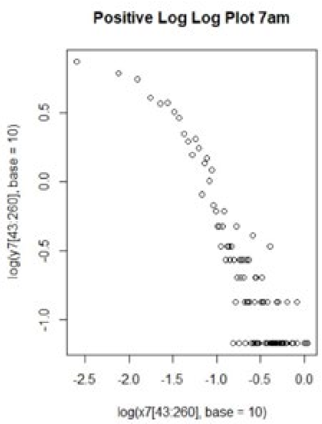


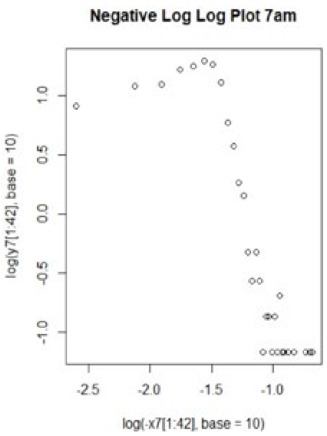
We will now consider the 6 am data set. Similarly, to the 5 am distribution there seems to be evidence of the distribution having a right tail. The skewness for the 6 am distribution is 6.984 (to 3dp), which tells us that the 6 am distribution is highly positively skewed. Hence, we will consider the log-log plot (positive and negative portion). For the positive portion of the log-log plot, we see that for the interval [-1.5, -0.5], there seems to be somewhat of a straight line and after -0.5 the points flatten out, so there is not a strong evidence of a power law. However, we will still consider a power law (since certain points in the positive log-log plot do form a straight line). The equation of the line of best fit that best follows the positive log-log plot is y=-1.52311x-1.98153. The gradient of the equation is -1.5 (to 1dp). Therefore, we can consider a fit with exponent -1.5. The equation of the power law is in the form of: $y = a * x^{-1.5}$ $where\ a \in \mathcal{R}$. Now when looking at the negative log-log plot we can see that the data points are a curvature rather than a straight line. Therefore, we conclude that a power law does not exist in the negative log-log plot. Hence, we can analyse the semi log-log plot and focus on the negative portion of the data.



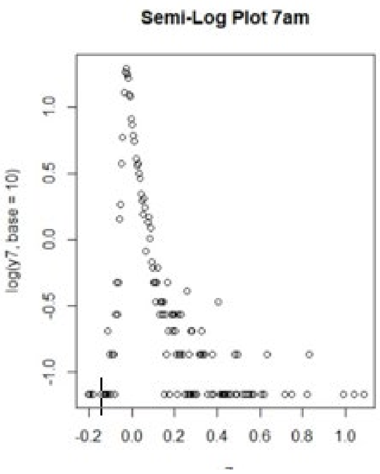
The semi-log plot is somewhat similar to the semi-log plot of the 5 am distribution. What we can see is that from [0,0.2] and [0,-0.2], we have somewhat of a straight line, thus indicating an exponential fit, however after 0.2 we see that the points flatten out, and the same can be said for points before -0.2. Thus, we conclude that a strong exponential fit does not exist.



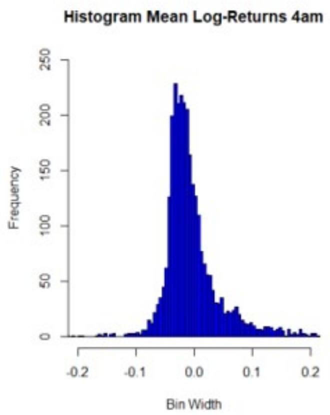


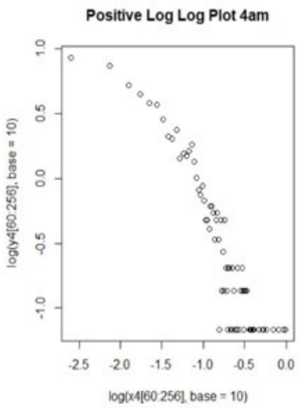


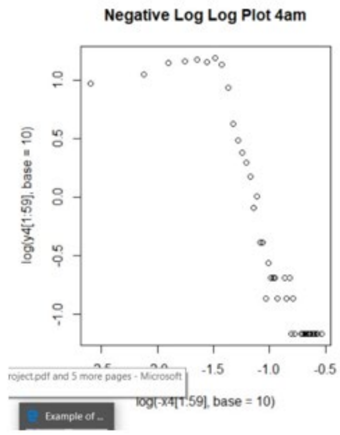
Next, we shall look at the 7 am plot, the skewness for the 7 am plot was 5.346, which indicates that the distribution is highly (positive) skewed. We shall now consider the log-log plot. Just as the case for 5 am we see that the positive log-log plot of 7 am does not lay on a straight line. In the positive log-log plot, the points between [-1.5: 0.3] are roughly in a straight line, but overall, they are not straight. The negative plot however does not lay in a straight line. Hence, we shall consider the semi log-log plot.



We shall analyse both portions of the semi-log plot. The semi-log is somewhat straight between [0.0,0.2], but since this is such a small interval it would not make sense to claim that the graph is exponential for certain points. The same argument can be used for the negative portion of the semi-log plot. The graph is fairly straight for the interval [0,0.1], which is a small interval, hence we claim that an exponential fit does not exist for the right tail.







Next, we analyse the final graph with the right tail, which is the 4 am dataset. The skewness for the 4 am dataset is 4.606, which indicates that the distribution is highly (positively) skewed. We shall now consider the log-log plot (positive and negative). Generally speaking, the positive log-log plot is more of a curvature rather than a straight line, however there are certain points in which the graph appears roughly straight. The interval [-1.2, -0.5] is somewhat of a straight line, hence we consider a line of best fit in the interval [-1.2,-0.5]. A similar trend follows for negative log-log plots, we see that for the interval [-1.5, -0.8] the distribution is straight. The equation of the line of best fit is y=-1.19476x-3.43047. The gradient is -1.2 (to 1dp), hence we consider a fit with exponent -1.4. Which means the equation of the power law is in the form of: $y = a * x^{-1.2}$ $where\ a \in \mathcal{R}$.

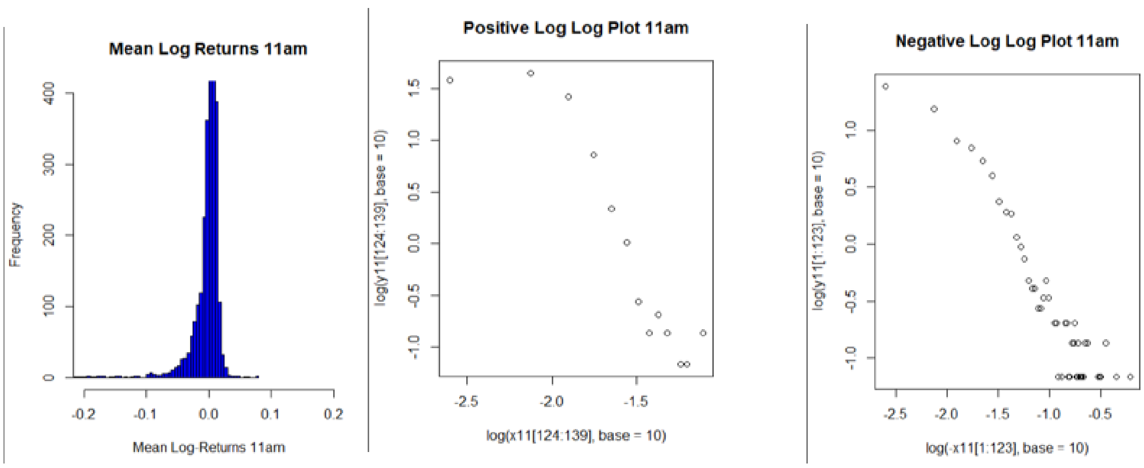
We can summarise the findings of the right tail distributions in a table. The table will consist of the exponent of the power-law fit (if any) and the skewness of the distribution. This can help see if any potential relationships between skewness and the exponent exists.

Time	Exponent	Skewness
4 am	-1.2	4.61
5 am	No power law	2.30
6 am	-1.5	6.98
7 am	No power law	5.35

The first thing that we notice is that all the distributions are highly positively skewed, which is fairly self-explanatory when we consider the time series that we observed earlier on, more specifically the time series plot for 5 am, 6 am and 7 am, consisted of positive extremities whereas the 4 am plot contained a mixture of positive and negative plots, with the negative extremities being the more eye-catching extremity, however, positive extremities occurred more frequently, which explains why the skewness is highly positive. This means that positive extreme events generally lead to a highly positive skewness which as a result leads to long tails on the right side of the histogram distributions. Now we come to the exponent, regarding the exponent, the smaller the exponent is, the faster the tail of the distribution decays. What we notice from the table above is that the distribution that decays the fastest is for the 6 am database (smallest exponent), and the slowest for the 4 am dataset (largest exponent), moreover, the skewness for the 6 am dataset is greater than that of the 4 am dataset. Therefore, there seems to be a relationship between skewness and exponent. More specifically, it seems like the larger the skewness is the faster the right tail decays. We shall clarify if this is indeed the case when we analyse the distributions that contain left tails. Finally, what we see is that the 5 am and 7 am plots did not follow a power law or an exponential fit. This concludes the analysis of distributions with right tails.

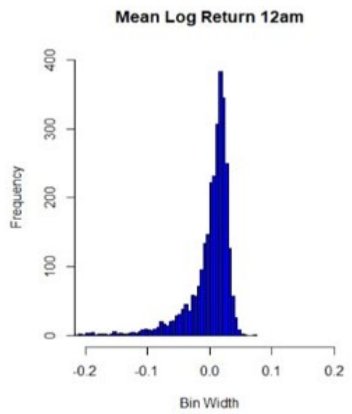
4.2 Log-Log Plot and Semi Log Plot (Left Tail)

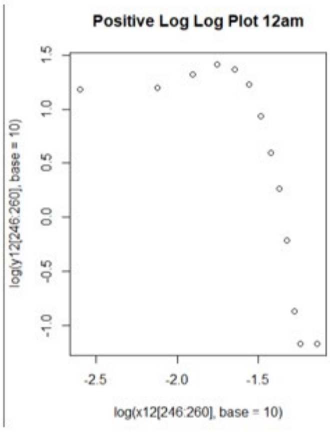
We now move to distributions that have long left tails, and as we did for the right tail, we shall calculate the skewness and investigate whether a power law exists for these distributions. Keeping in mind that we are dealing with the left tail, our values will be negative, therefore we cannot take the logarithm of those values. To overcome this, we shall take the absolute value of these numbers, this will allow us to take the logarithm of these number, hence we can calculate the log-log and the semi-log plot. This can be done in RStudio, when calculating the log of the x vector we put a minus in front of it. For example, in RStudio the command could be log(-x[....], base=10), this will give us the absolute values of the negative numbers. In the case of left tailed distributions, if a power law does indeed exist then it will be in the form of $y = a * |x|^{-n}\ where\ a \in \mathcal{R}$ and 'n' is the value of the exponent.

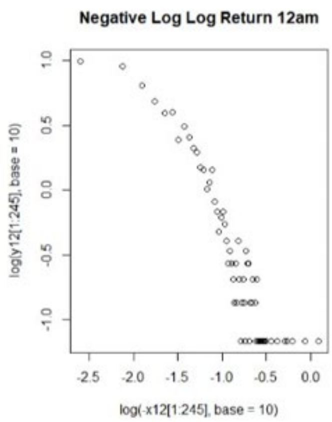


We shall first analyse the 11 am dataset.

The skewness of the plot is -7.339, which indicates that the data is highly (negatively) skewed. Generally speaking, when looking at the log-log plot (positive and negative), it is fairly obvious that the line is not straight, however upon close inspection we notice that in certain intervals the line does seem to be almost straight. From the positive log-log plot, the interval [-2,-1.3] seems to be fairly straight and for the negative log-log plot the interval [-1.8, -1] seems to be roughly straight. Hence, we consider a line of best fit for those intervals. The equation of the line of best fit is: $y = -1.42325x - 2.02252$. Hence, we consider a fit with exponent -1.4 (1 dp), which means the equation of the power-law will be in the form of $y = a * |x|^{-1.4}$ $where\ a\ \varepsilon\ \mathcal{R}$.

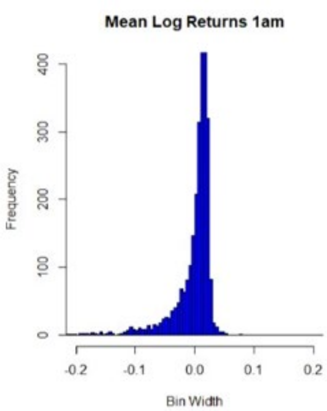


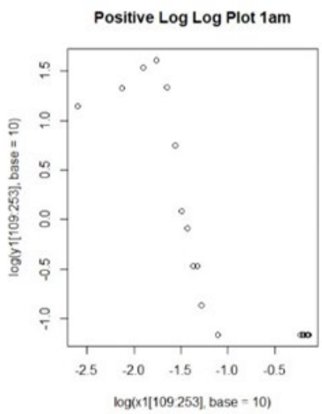


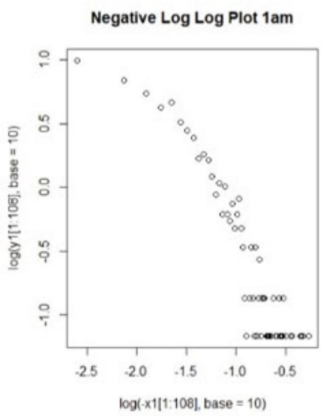


Next, we move onto the 12 am plot. The skewness of the 12 am distribution is -4.643, which indicates that the distribution is highly (negatively) skewed. As with the 11 am plot, we notice that the positive log-log plot is a clear curvature, but certain intervals are fairly straight. For example the positive log-log plot has a slight positive trend in the interval [-1.6, -1.3] but the interval is way too small to generalise for the whole data set, however the negative log-log plot is straight for the interval [-1.5, -0.8], hence we can consider a line of best fit for that interval. The equation for the line of best fit is: $y = -0.19212x - 0.51959$. Hence, we consider a fit with exponent -0.2 (1 dp), which means the equation of the power-law will be in the form of $y = a * |x|^{-0.2}$ $where\ a\ \varepsilon\ \mathcal{R}$.

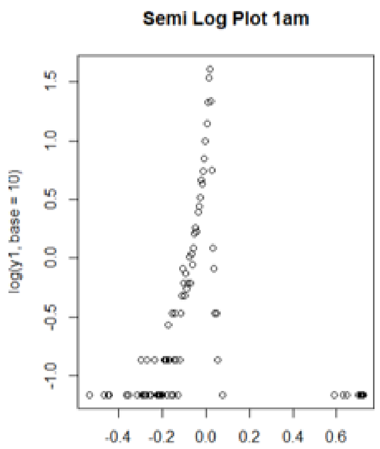
Next up is the 1 am plot.



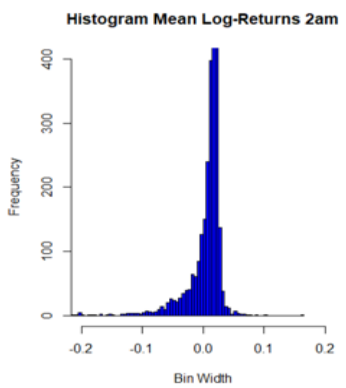


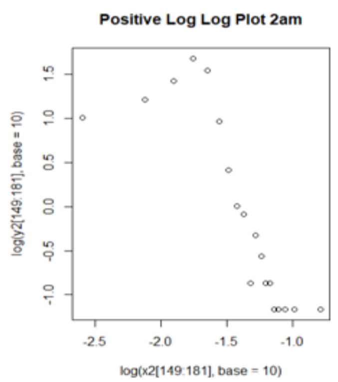


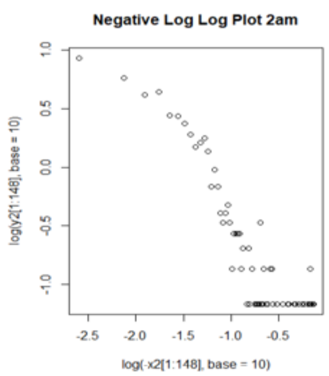
The skewness for the 1 am distribution is -2.006, which once again indicates that the distribution is highly (negatively) skewed, but not as skewed as the 11 am and the 12 am distributions. When looking at the positive log-log plot, we can see that the distribution is that of a curve rather than a straight line, and the same can be said for the negative log-log plot. The positive log-log plot is fairly straight for the interval between [-1.5, -1.1], but the interval is too small to generalise for the whole dataset of the positive log-log plot. With the negative log-log plot, there does not seem to be an interval where the distribution seems to be straight at all. This indicates that there does not seem to be any power-law involved. Hence, we will consider the semi-log log plot to investigate whether there is evidence of an exponential fit involved.



From the semi-log plot, we can see that the plot is not a straight line, hence we can conclude that an exponential fit for the 1 am does not exist.

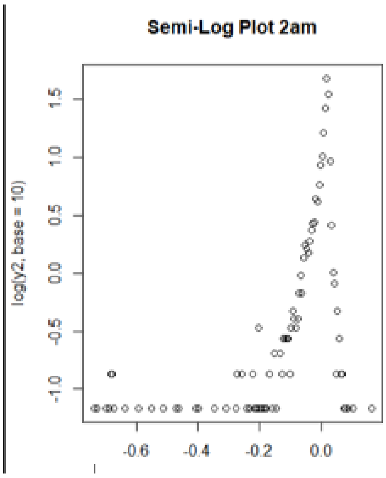




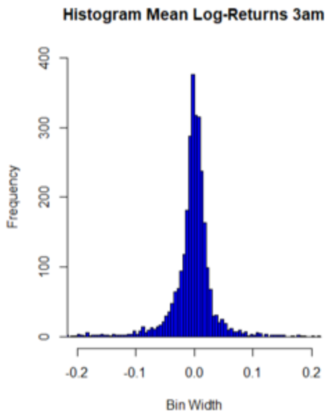


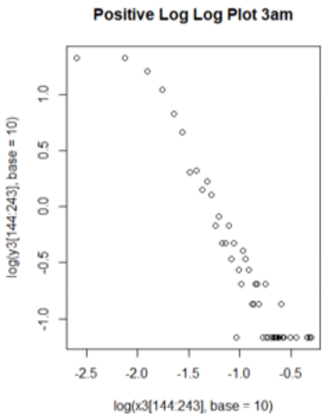
Next up we have the 2 am plot.

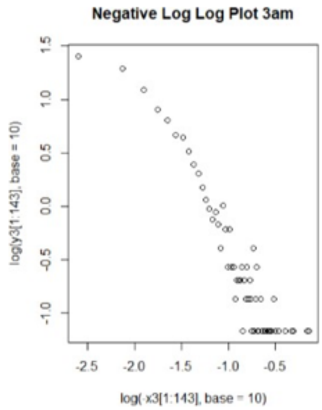
The skewness for the 2 am distribution is -8.051 (3 dp), which means that the 2 am distribution is highly (negatively) skewed. When analysing both the positive and the negative log-log plot we notice that there does not seem to a straight line visible. It could be argued that in the positive log-log plot there seems to be somewhat of a straight line between the interval [-1.5,-1.2], however once again the interval is far too small to generalise a power law for the whole 2 am distribution. For the negative log-log plot, we can see the plot is of a curvature. Hence, we shall analyse the semi-log plot.



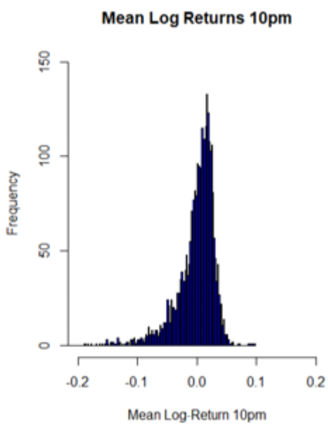
Just as in the case for 1 am, we notice that the plot is not of a straight line, hence we conclude that an exponential fit does not exist for the 2 am plot.

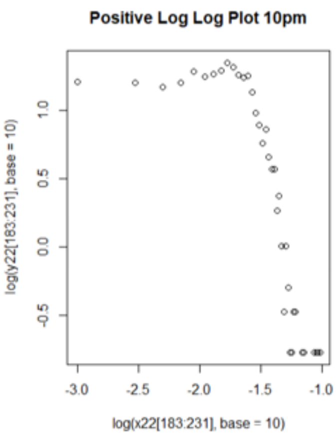


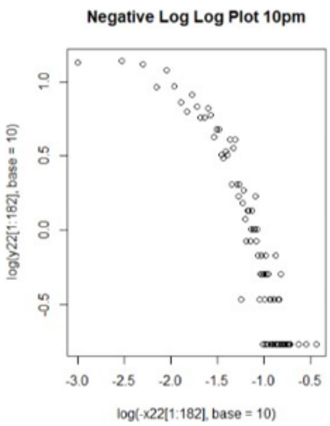




Upon inspection, the 3 am distribution seems to be centred, but there is a hint of skewness to the left, hence I have considered the 3 am plot to be left tailed. Surprisingly the skewness for the 3 am distribution is -1.897, which indicates that the 3 am distribution is highly (negatively) skewed. When compared to the other left tailed distributions the skewness is fairly small, which is to be expected. Now we analyse both the positive log-log and the negative log-log plot. With the positive log-log plot, there seems to be an interval in which the plot seems to contain a straight line. Thus, we shall consider the interval [-1.7,-0.8] for the positive log-log plot, and for the negative log-log plot, we shall consider the same interval when calculating the equation for the line of best fit. Since most of the points seem to lay in a straight line, this indicates that a strong power law may exist. The



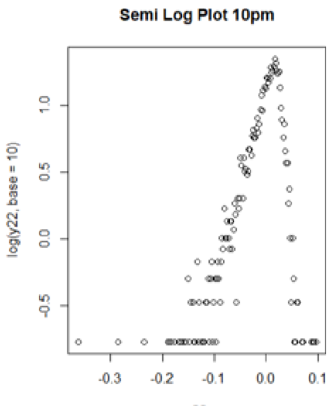




equation for the line of best fit is: $y = -1.5841x - 1.98744$. Hence, we consider a fit with exponent -1.6 (1 dp), which means the equation of the power-law will be in the form of $y = a * |x|^{-1.6}$ $where\ a\ \varepsilon\ \mathcal{R}$.

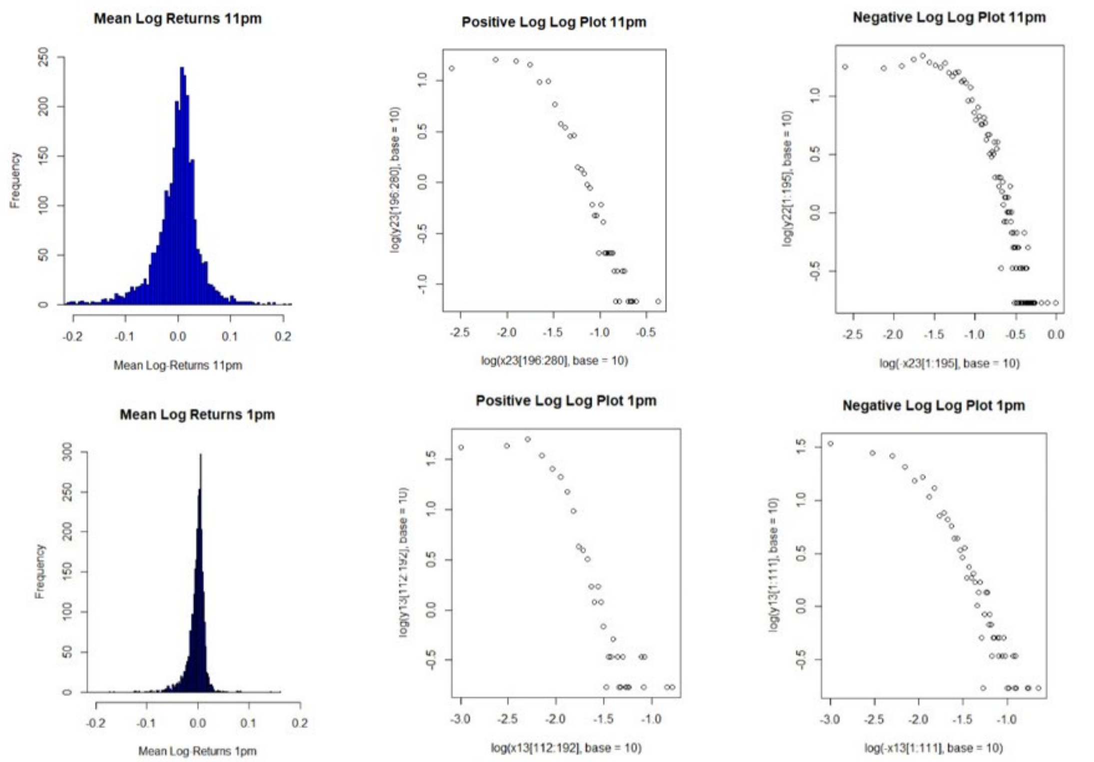
Next, we look at the 10 pm dataset.

The skewness for the 10 pm distribution is -2.256, which indicates that the 10 pm distribution is highly (negatively) skewed. We now analyse the log-log plots. The negative log-log plot is a clear curvature, whereas with the positive log-log plot the interval [-1.6,-1.3] seems fairly straight. However, the interval is far too small to consider a power law. Hence, we can conclude that a power-law may not exist for the 10 pm dataset. Hence, we consider the semi-log plot to conclude whether there is an exponential fit for the 10 pm distribution.



Just like every other semi-log plot we have come across, there does not seem to be a straight line present, hence an exponential fit is not present in the 10 pm plot.

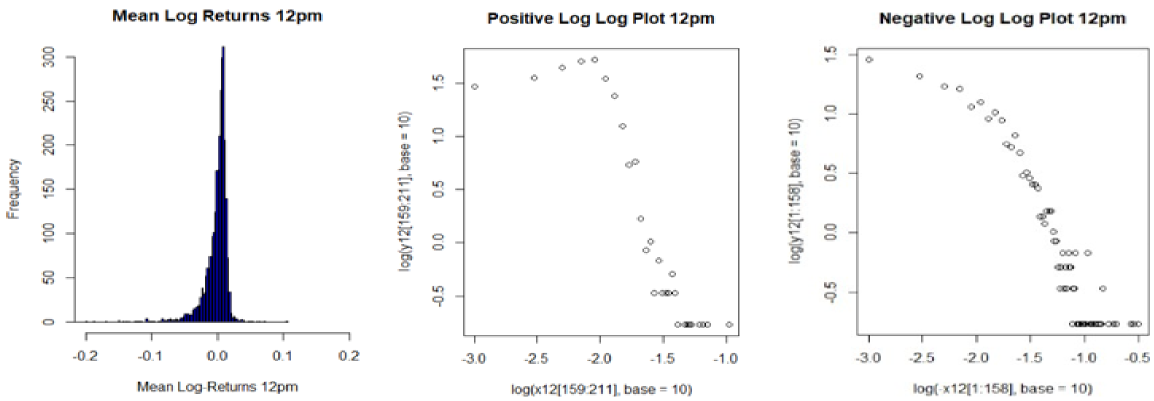
Now we consider the 11 pm plot and the 1 pm plot.



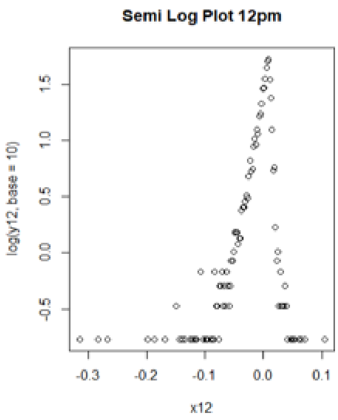
We shall first start with the 11 pm dataset. Now when looking at the histogram for the 11 pm dataset, we have a similar situation as we did with the 3 am dataset. That is the fact that the 11 pm distribution seems to be fairly centred, with just a hint of skewness. The skewness of the 11 pm plot is -3.931, which indicates that the distribution is highly (negatively) skewed. This came as somewhat as a surprise since I would have expected the skewness to be closer to -1 rather than -3. Next, we move to the log-log plots. We start with the positive log-log plot. The plot is somewhat straight, more specifically the interval [-1.7, -0.9] seems to be fairly straight. For the negative log-log plot, the interval [-1,-0.4] seems to be somewhat straight. Hence, we consider a power law. The equation of the line of best fit is: $y = -1.49847x - 1.84427$. Hence, we consider a fit with exponent -1.5 (1 dp), which means the equation of the power-law will be in the form of $y = a * |x|^{-1.5}$ $where\ a\ \varepsilon\ \mathcal{R}$.

Now we move onto the 1 pm dataset. From the histogram, we can see that the distribution is skewed towards the left. The skewness of the 1 pm dataset is -2.006, which indicates that the distribution is highly (negatively) skewed. Just as in the case of the 11 pm dataset there are some intervals in which a power law is possibly present. We first consider the positive log-log plot. The interval [-2, -1.3] seems to indicate a power law, and for the negative log-log plot, the interval [-2, -1.1] seems to indicate some sort of a power law. Hence, we shall consider an equation for the line of best fit in the interval [-2, -1.1]. The equation of the line of best fit is: $y = -1.45739x - 2.00288$. Hence, we consider a fit with exponent -1.5 (1 dp), which means the equation of the power-law will be in the form of $y = a * |x|^{-1.5}$ $where\ a\ \varepsilon\ \mathcal{R}$.

Now we consider the final plot, which is the 12 pm plot.



The skewness of the 12 pm distribution is -4.643 which indicates a highly (negatively) skewed distribution. Both log-log plots do not seem to indicate a power law; hence we consider a semi-log plot.



Once again, we see that the semi-log plot is not that of a straight line, hence we can conclude that the distribution does not contain an exponential fit.

We can summarise the findings of the left tail distributions in a table. The table will consist of the exponent of the power-law fit (if any) and the skewness of the distribution. This can help see if any potential relationships between skewness and the exponent exists.

Time	**Exponent**	**Skewness**
11am	-1.4	-7.34
12am	-0.2	-4.64
1 am	No power law	-2.01
2 am	No power law	-8.05
3am	-1.6	-1.90
10 pm	No power law	-2.26
11pm	-1.5	-3.93
12 pm	No power law	-4.64
1pm	-1.5	-2.01

The first thing that we notice is that all the distributions are highly negatively skewed, which is fairly self-explanatory when we consider the time series (plots) that we observed earlier on, more specifically the time series plot for 11 am, 1 am, 3 am, 11 pm, 12 pm and 1 pm consisted of negative extremities whereas the 12 am and 10 pm plots contained a mixture of positive and negative plots, with the latter being the more eye-catching extremity, furthermore, the negative extremities were larger than that of the positive extremities, which is a possible explanation as to why the skewness is highly negative. There seems to be one dataset that does not seem to fit in with the trend, and that is the 2 am dataset. The time series plot for the 2 am dataset consisted of periodic positive extremities, but when looking at the skewness it was highly negatively skewed, and the skewness was far more severe compared to the other datasets. One possible explanation could be that the 2 am dataset had positive extremities which were greater than the negative extremities in terms of value, but the frequency in which the negative extremities occurred was greater than that of the positive extremities. Therefore, the frequency of the negative extremities may have had a greater impact than the value of the extremities. This means that negative extreme events generally lead to a highly negative skewness which as a result leads to long tails on the left side of the histogram distributions.

What we notice from the table above is that the distribution that decays the fastest is for the 3 am database (smallest exponent), and the slowest for the 12 am dataset (largest exponent). However there does not seem to be any relationship between exponent and the skewness, because the 3 am dataset had the smallest exponent and had the largest skewness, followed by the 11 pm and 1 pm dataset, but that trend is not followed by the 11 am dataset. The 11 am dataset had a smaller exponent than the 12 am dataset, yet the skewness for the 11 am dataset was far smaller than that of the 12 am dataset. Earlier on, for the right tail, we concluded that the larger the skewness is the faster the right tail decays, however that is not backed up by the left tail. One reason could be due to the fact for the right tail there wasn't enough datasets to compare and come up with a conclusive relationship between exponent and skewness (only 2 data sets were compared for the right tail distributions), whereas for the left tail we had 5 distributions to compare and contrast. Hence, we can conclude that there does not seem to be a relationship between exponent and the skewness.

5.0 Auto-Correlation function (ACF)

We shall now analyse the ACF for the distributions with right and left tail. The aim of this is to see whether the exponent or skewness has an impact on the shape of the ACF. Before we get into analysing the ACF plot, we shall give a mathematical definition and then a more informal definition of the ACF.

Definition 5 (W Yoo): The autocovariance function of a stationary time series $\{x_t\}$ at lag k denoted by $\gamma(k)$, is defined as

$$\gamma(k) = Cov(x_{t+k}, x_t)$$

Definition 6 (W Yoo): The autocorrelation function of a stationary time series $\{x_t\}$ at lag k is defined as

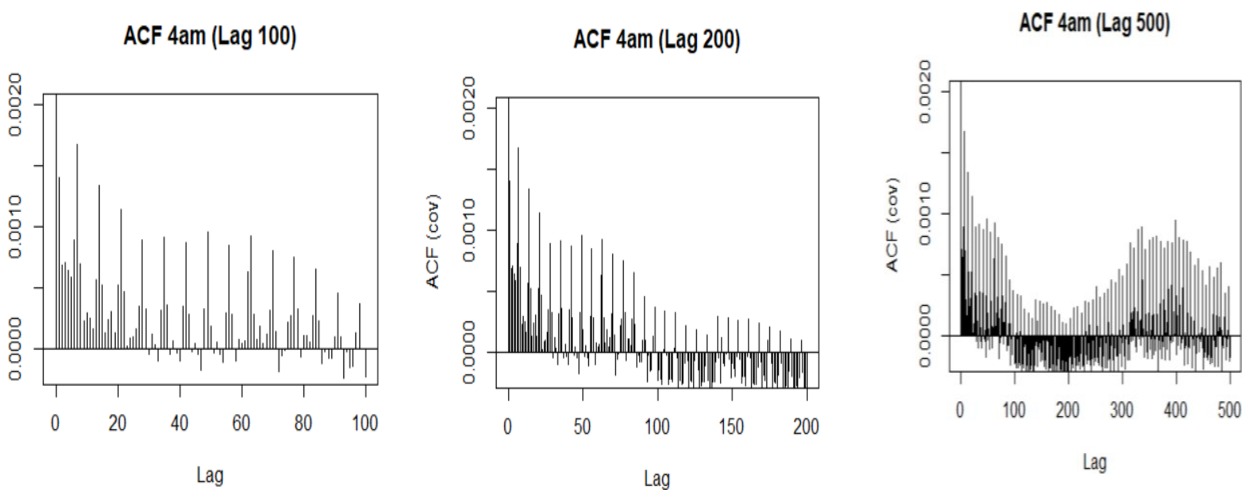
$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = Corr(x_{t+k}, x_t)$$

Now we shall give a more informal definition of autocovariance function and autocorrelation function. The autocovariance is a function that gives the covariance of the process with itself at certain points in time (known as lag- denoted by k), whereas the correlation gives us the correlation of the process with itself at certain points in time.

What the ACF plot will tell us is whether electricity prices today have an effect on electricity prices in say 10 days, 20 days or even 100days' time.

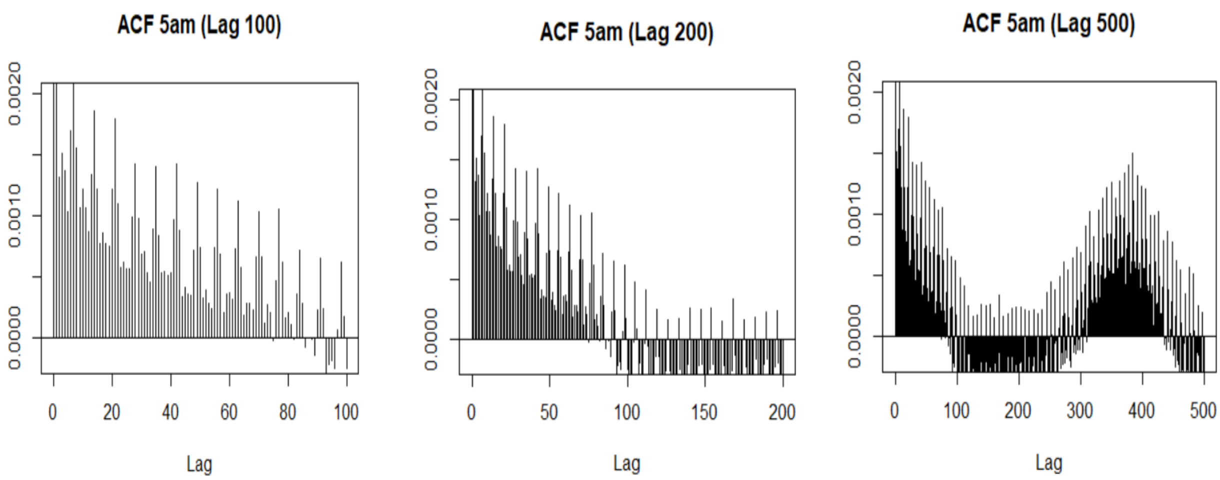
We shall start by analysing and comparing the ACF for the right tail distributions. To be able to compare the ACF for each right tail distribution I ensured that the axis range were the same for each ACF. For the dataset conditioned per hour, 3 lag values were used, namely 100,200 and 500 lag values (i.e. k=100,200 and 500).

We first start with the 4 am dataset.

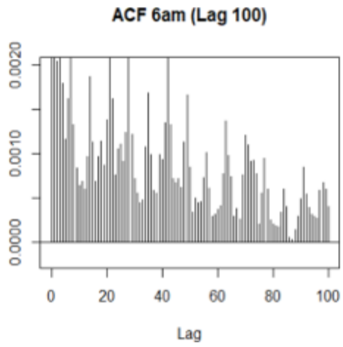


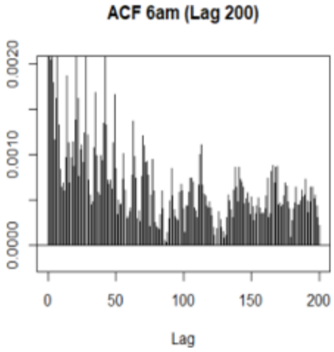
For the ACF 4 am with lag 100 we notice an interesting trend that seems to occur. We see high positive spikes every (roughly) 7 days. This indicates that the electricity prices seem to be positively correlated every 7 days. When looking at the ACF 4 am with lag 200 we notice that the trend of positive spikes every 7 days seems to carry on occurring, however this time we notice that a slight oscillating pattern seems to be emerging. The oscillating pattern becomes evident when we consider the ACF 4 am with lag 500.

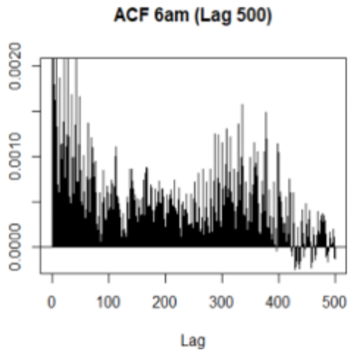
Next up is the 5 am plot, we shall first analyse the 5 am plot but also compare it with the 4 am ACF plot.



First, we analyse the 5 am plot with lag 100. What we notice is that the lag 100 plot follows the same pattern as the 4 am plot with lag 100, in the sense that the most evident positive spikes seem to occur every 7 days, the difference between the 4 am and 5 am with lag 100 is that the 5 am plot seems to have a lot more positive spikes compared to the 4 am set. This means that for the first 100 days the electricity prices seem to be positively correlated, with the correlation being the greatest every 7 days. Next, we look at the ACF 5 am with lag 200, what we notice is that the pattern is similar to the 4 am plot, in the sense that an oscillating pattern seems to be emerging. Finally, we look at the ACF 5 am with lag 500, the plot is very interesting because we see a very evident oscillating pattern, namely, we see positive correlation followed by negative, followed by positive and then finally a negative correlation. This might be due to changes in season. The positive correlations could represent winter and summer since during those seasons people would be using their central heating and using fans for the summer, which utilises electricity.

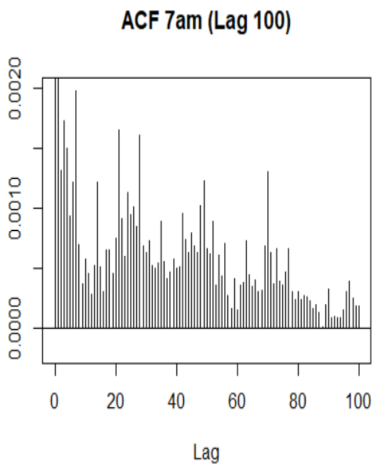


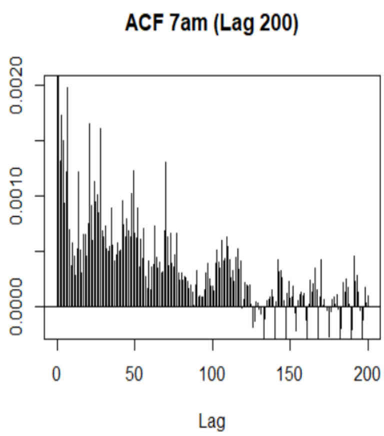


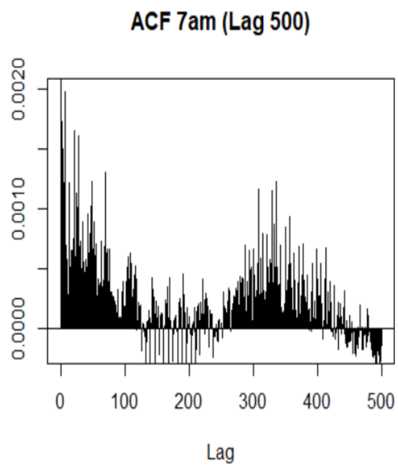


The first thing we notice with the 6 am ACF with lag 100 is that the points are all positively correlated, unlike the 4 am and 5 am plots which had some points that had some negative correlations present. We notice the same trend for the 6 am ACF with lag 200. This means that the electricity prices are positively correlated with each other for the first 200 days. Now we come across the 6 am ACF with lag 500, what we notice is that most points are positively correlated to each other, but some points are negatively correlated. These occurrences seem to be random.

Finally, we take a look at the 7 am dataset.

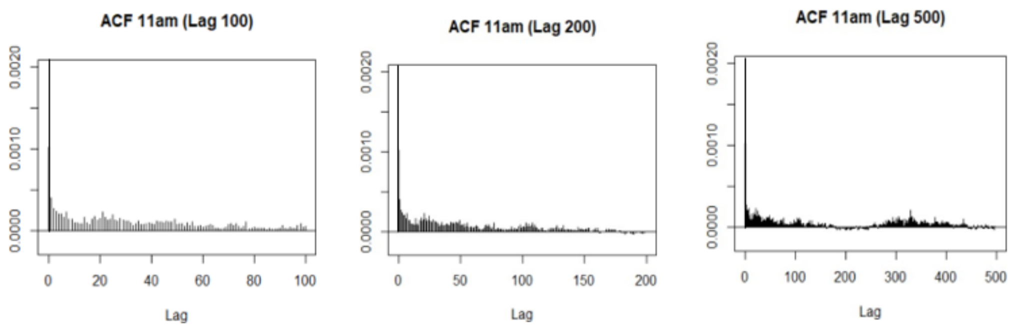




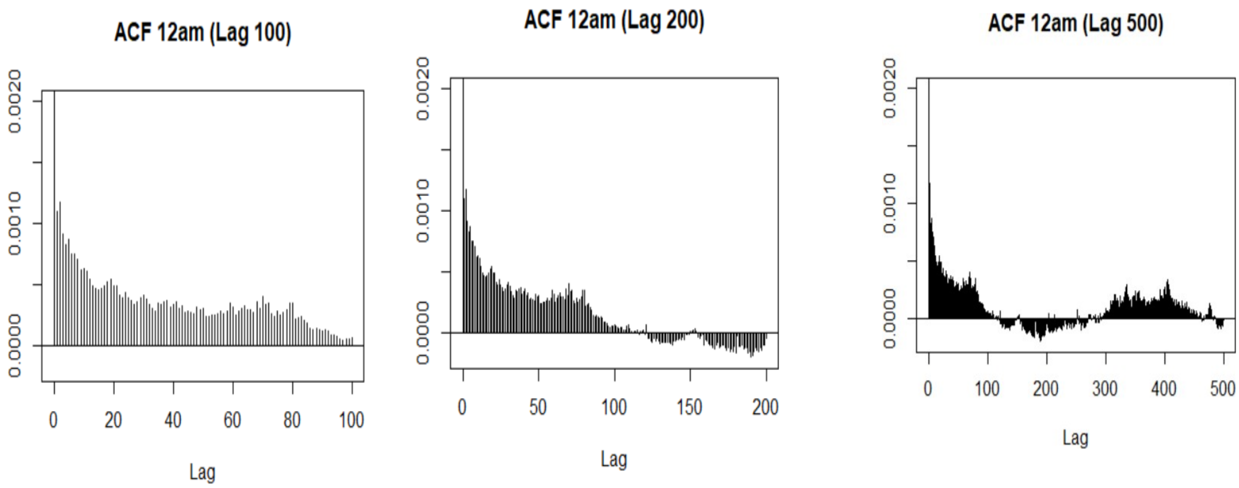


The 7 am plot with lag 100 follows a similar pattern to the 6 am plot with lag 100, all the points are positively correlated. When looking at the ACF 7 am with lag 200 we notice some negative correlation after (roughly) 125 days. Now we look at the ACF 7 am with lag 500, just as the case 5 am ACF with lag 500 we notice some sort of repetition of positive and negative correlation. Again, this might be an indication of a seasonal effect. From the lag 500 graphs for 7 am and 5 am we notice that the seasonal effects had a greater effect on the 5 am data set than the 7 am dataset. Now we shall briefly state whether the exponent has a possible effect on the ACF plot, as well as whether skewness impacts the ACF plot. The exponent does not seem to have any obvious impact on the trend of the ACF plot. Next, we look at the skewness, the 6 am dataset had the largest skewness, now the 6 am plot also had more frequent as well as larger positive correlations when compared to the other datasets. The prospect of a larger skewness leading to a larger and more frequent positive correlations is not supported by the 4 am dataset. The 4 am dataset had a larger skewness compared to the 5 am dataset, however, the 5 am ACF had larger and more frequent positive correlations than the 4 am ACF. One possible reason could be since the time series plot of the 4 am consists of positive and negative extremities, hence why the ACF of the 4 am plot had smaller and less frequent positive correlations. So, it seems like extremities might impact the ACF plots. We shall take a look and see if the left tail distributions support the prospect of a larger skewness leading to a larger and more frequent positive correlations, also we can analyse whether the extreme events impact the ACF plots.

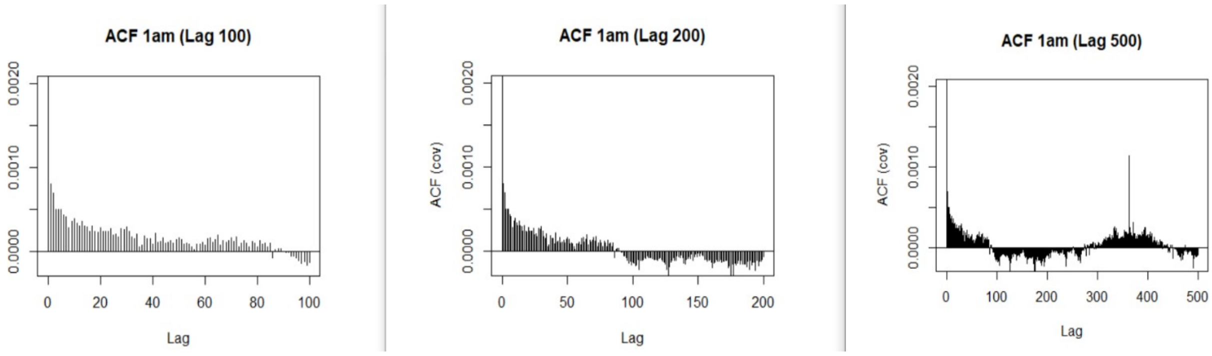
Next, we consider the left tailed distributions. First up we consider the 11 am ACF.



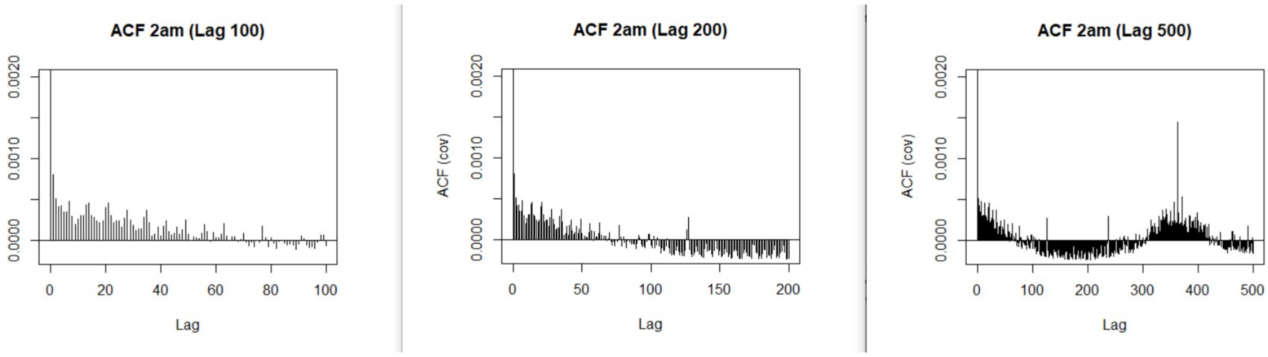
For the 11 am data we see that lag 100,200 and 500 all follow a similar trend, in the sense that we see that on the first day we see the highest positive correlation, and generally the positive correlation seems to get smaller and smaller. In lag 500 we have a few negative correlations but to a small degree. Next up is the 12 am dataset.



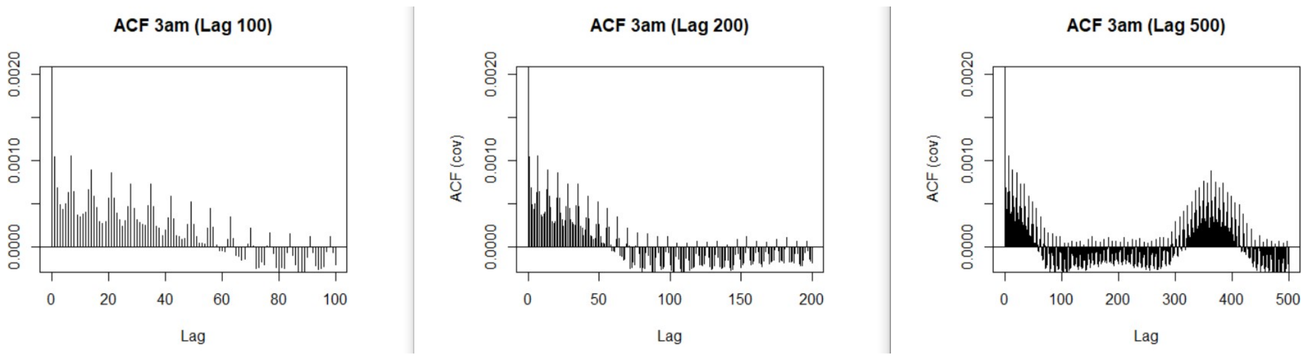
The ACF 12 am with lag 100 follows a similar pattern to the 11 am dataset, in the sense that all the points are positively correlated, with the first point being the greatest correlated. However, the 12 am with lag 100 has higher positive correlations compared to the 11 am dataset. Now we move onto the 12 am with lag 200, here we some differences arise compared to 11 am ACF with lag 200. ACF 12 am with lag 200 we notice that the plot starts positively correlated, but after about 120 days (roughly) we notice the points become negatively correlated. This may be due to seasonal changes. This is further backed up by the lag 500 plot, we can see a repetition of positive and negative correlations, which further indicates seasonal changes.



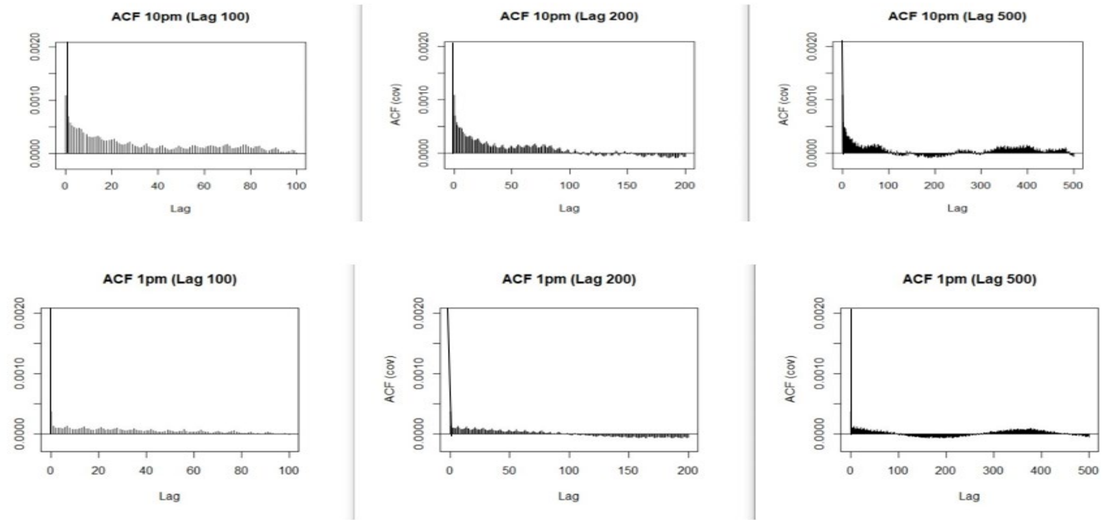
The ACF 1 am plot follows a similar trend to the 12 am plot, we notice seasonal changes in lag 200 and lag 500 plots. Around roughly day 365 we notice a spike in the value of the correlation. This might indicate a periodic pattern that occurs after every year.



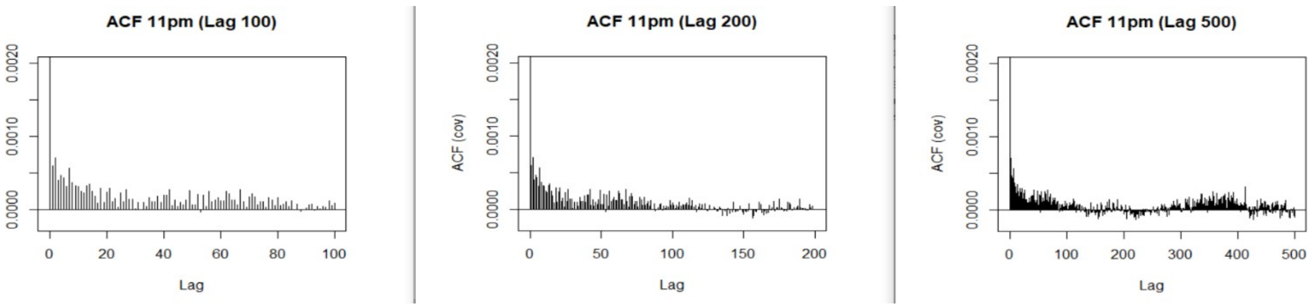
The 2 am ACF for all lag values follows pretty much the same trend as the 1 am dataset, we notice a slight periodic trend occurring. Next up we look at the 3 am plot.



The 3 am plot seems to be affected by some sort of seasonal change, but unlike the 2 am and 1 am plot, the fluctuations do not seem to be periodic. Another difference is that with the 3 am dataset the data is more positively and negatively correlated compared to the 2 am and 1 am dataset. So, the electricity prices at 3 am are more positively correlated at certain days compared to the 2 am and 1 am dataset, and electricity prices are more negatively correlated at certain days compared to the 2 am and 1 am dataset.

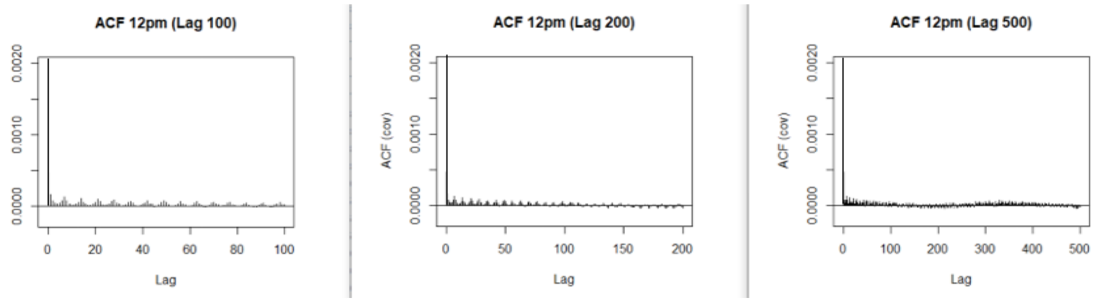


The 10 pm and 1 pm follow somewhat of a similar trend, what we see is that at lag 100 the ACF for 10 pm and 1 pm are all positively correlated, the only difference coming from the fact that the 1 pm dataset are more positively correlated than the 10 pm dataset, this means that for both 10 pm and 1 pm the electricity prices are correlated to each other in the first 100 days, more correlated for the 1 pm dataset. Next up is the lag 200 graph, what we notice is that for both 10 pm and 1 pm plot after roughly 100 days the electricity prices become negatively correlated, once again much more negatively correlated for the 1 pm dataset. This means after about 100 days the electricity is negatively correlated to one another. This means that electricity prices today have a negative impact on the electricity after 100 days. Now we come across lag 500, immediately we can see some sort of seasonal effect, but for the 500 lag for 1 am the seasonal impact is far less severe compared to the 10 pm dataset.



For the 11 pm dataset what we notice is that for lag 100 the electricity prices are all positively correlated, but lag 200 and 500 we notice some sort of oscillation taking place, so there are certain days where the prices are negatively correlated and others where they are positively correlated. Most days the electricity prices are positively correlated.

Finally, we come to the final dataset which is the 12 pm dataset. Now what we notice is that correlations are very tiny, more specifically smaller than 0.0005, so electricity prices today have a very small impact on electricity prices in the future.



We now summarise our findings and look for any potential relationship between skewness and the ACF plots also any potential relationship between extremities present in the time series plot and the ACF plots.

When looking at skewness there does not seem to be any apparent relationship between skewness and the ACF plot. Now we look at extremities. When analysing the extremities for the left tailed distributions we realise that there does not seem to be any relationship between extremities and the ACF plots, for example, the 12 pm dataset had negative extremities yet the values of the correlations in the ACF was the smallest, however, the 3 am plot also contained negative extremities yet the correlation values for the ACF plot was fairly large when compared to the other left tailed distributions. Hence, we can conclude that skewness, exponent and extremities do not have an impact on the ACF.

One potential area worth researching further could be whether the dataset conditioned per hour and per season has an impact on how correlated electricity prices are to each other.

6.0 Geometric Brownian Motion

We have observed datasets with either a left tail or a right tail, now we shall look at the datasets that do not have a left tail nor right. We shall look at the skewness of these datasets and try to determine which dataset is the closest to following a Geometric Brownian Motion. Why Geometric Brownian Motion? This is because the Geometric Brownian Motion (GBM for short) Is a common model used to simulate stock prices, this means that the dataset that best follows a GBM we can simulate a rough idea of the electricity prices in the future. According to the GBM, the future price of a financial stock has a lognormal probability distribution. Therefore, we can estimate the future prices to a certain level of confidence. In a Geometric Brownian Motion, we will expect a normal distribution, hence the datasets with a skewness closest to 0 will be the dataset that is the closest to following a Geometric Brownian Motion (GBM).

Before we start analysing, we will state a few definitions, starting with Brownian motion, also known as the Weiner process.

Definition 7 (Phillips, 2019)

The Wiener process $W(t)\ t \geq 0$, (also known as Brownian motion) is a stochastic process which satisfies the following criteria:

- $W(0) = 0$
- $The\ sample\ paths\ of\ W(t) are\ continuous.$
- $The\ increments\ of\ W(t) are\ independent,\ i.e.\ for\ any\ set\ of\ times\ 0 \leq t_1 < t_2 < \cdots < t_n\ the\ random\ variables$

  $$W(t_2) - W(t_1), W(t_3) - W(t_2), \ldots \ldots., W(t_n) - W(t_{n-1})$$

  $are\ independet.$
- $For\ any\ 0 \leq s < t, the\ increments\ are\ normally\ distributed$

  $$W(t) - W(s) \sim \mathcal{N}(0, t - s)$$

However, the Brownian motion has a huge flaw to it, and that is the fact that the Brownian motion assumes equal probability of stock price going up and stock price going down. In reality stock prices on average tend to increase over time. Hence, we come across an improved version of the Brownian motion which is known as the Brownian motion with drift. We will define Brownian motion with drift.

Definition 8 (Phillips, 2019)

The process: $X(t) = X(0) + mt + \sigma W(t)$ is defined as a Brownian motion with drift. Where,

- $X(0) = The\ initial\ value\ of\ the\ process$
- $m = Drift\ in\ the\ process\ per\ unit\ time$
- $\sigma = Volatility\ of\ the\ process$
- $W(t) = Wiener\ process$

The Brownian motion with drift also has a major flaw to it, that is the fact that the Brownian motion with drift can take negative values. In reality, stock prices cannot be negative, therefore the Brownian motion is not the best process to describe fluctuations of stock prices. We hence come across an even better process known as the Geometric Brownian motion.

Definition 9 (Phillips, 2019)

The process: $Y(t) = Y(0)\exp\,[mt + \sigma W(t)]$ is defined as a Geometric Brownian motion. Where

- $Y(0) = The\ initial\ value\ of\ the\ process$
- $m = Drift\ in\ the\ process\ per\ unit\ time$
- $\sigma = Volatility\ of\ the\ process$
- $W(t) = Wiener\ process$

The Geometric Brownian is a better process at pricing fluctuations of stock prices than the Brownian motion with drift, because the GBM does not take negative values.

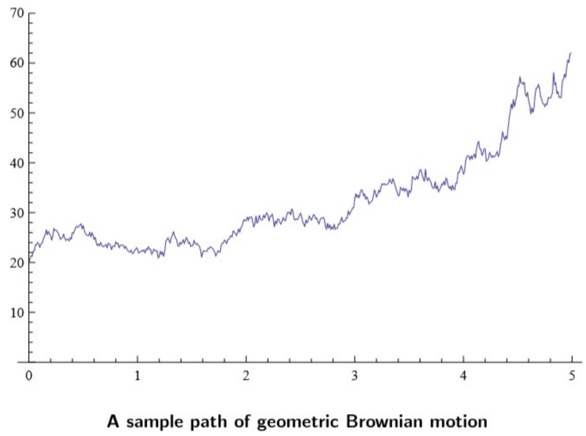
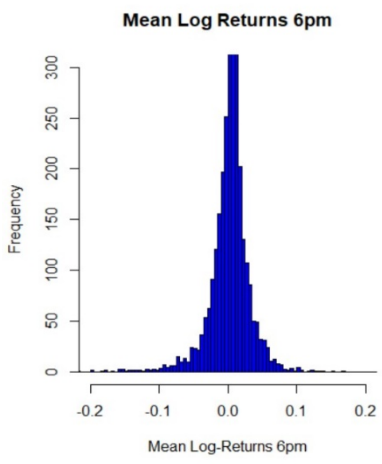
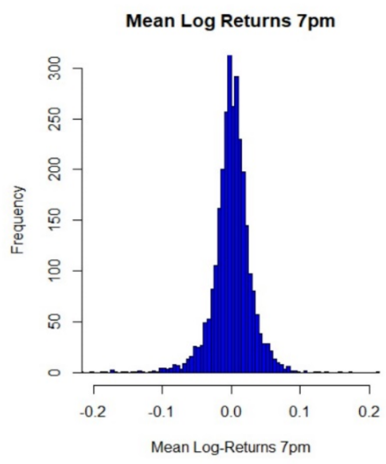
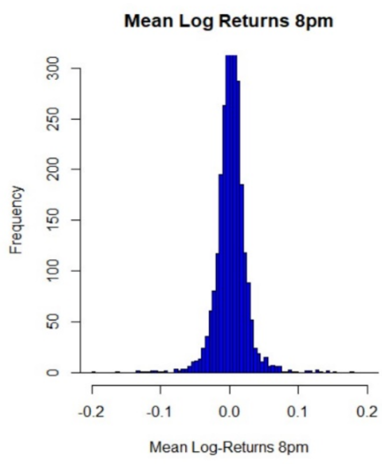


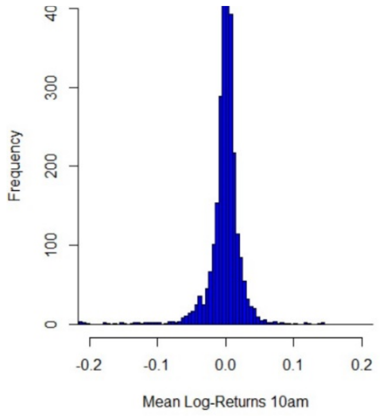
FIGURE 6: SAMPLE PATH OF GEOMETRIC BROWNIAN MOTION. TAKEN FROM PHILLIPS, M., 2019, LECTURE NOTES, FOUNDATIONS OF MATHEMATICAL MODELLING IN FINANCE, MTH771P, QUEEN MARY UNIVERSITY OF LONDON.

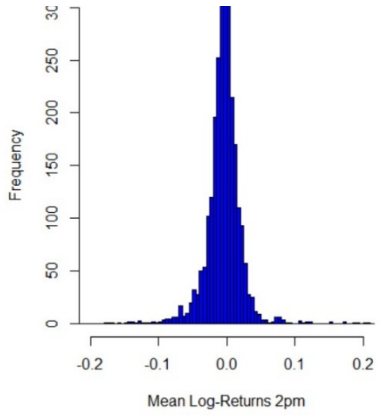
Figure 6 illustrates a sample path of Geometric Brownian Motion. We saw a similar upward trend when we plotted the whole data set of the prices. As earlier mentioned, the distributions with a skewness close to 0 will best follow a GBM, hence the distributions with left or right tails (long) do not represent a GBM. Furthermore, from figure 6 we notice that there does not seem to be any extreme events present, hence this further backs up the fact that the right and left tailed distributions certainly do not follow a GBM.

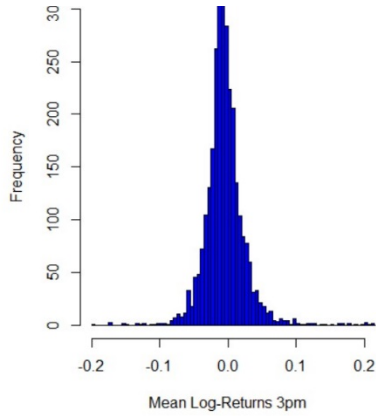


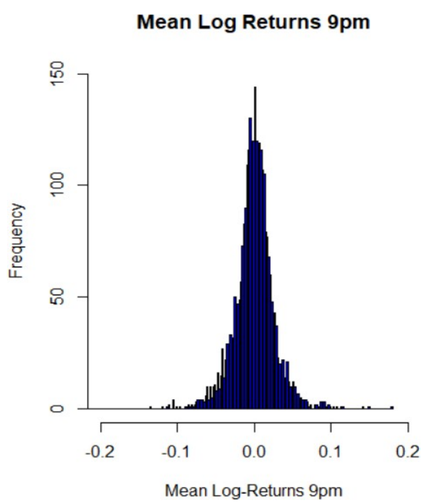










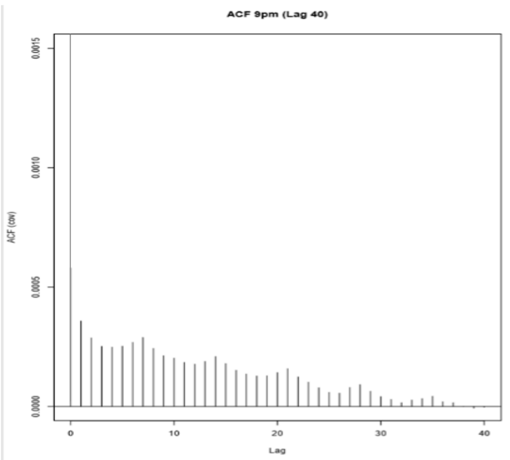


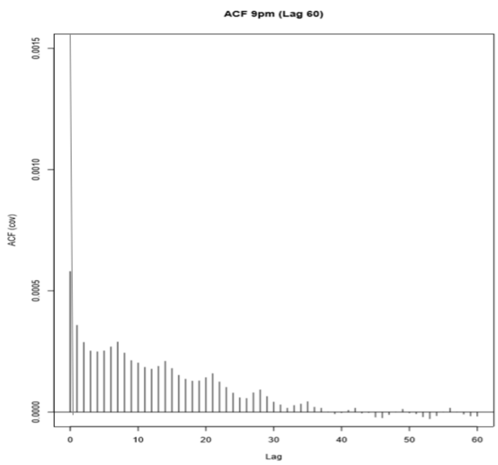
At first glance all the distributions above seem fairly symmetrical, however, to get a better picture of whether these distributions show a strong evidence of a GBM, we would need to take a look at the skewness. The table below shows the skewness of these distributions.

Time	Skewness
Whole dataset	1.94
8 am	-3.73
9 am	-7.71
10 am	-1.83
2 pm	4.40
3 pm	6.58
6 pm	-7.09
7 pm	-8.10
8 pm	0.56
9 pm	-0.23

From the table above, some of the skewness comes as somewhat of a surprise as some distributions are highly skewed. For example, the whole dataset, 2 pm and 3 pm distributions are all highly positively skewed. On the other hand, the 8 am, 9 am, 10 am and 6 pm distributions are all highly negatively skewed. As previously mentioned, any highly skewed distributions will most certainly not follow a GBM, hence we can conclude that these highly skewed distributions do not follow a GBM. Now we come across the 8 pm distribution, from the skewness we can see that the 8 pm distribution is moderately skewed, furthermore, the skewness is far less when compared to the others. However, it is not the 8 pm distribution that shows the strongest evidence of a GBM, rather it is the 9 pm distribution that shows the evidence of a GBM. With a skewness of -0.23, it is the distribution that is the closest to being symmetrical. Hence the 9 pm distribution best supports a GBM.

We have taken a look at the skewness to determine which dataset best demonstrates a GBM. We concluded that the 9 pm dataset best demonstrates a GBM. However, skewness is a necessary condition for GBM, but not a sufficient condition. This is because a large skewness rules out the possibility of a GBM, but a small skewness does not necessarily mean the dataset follows a GBM. It is possible to have a dataset with a small skewness that is not a GBM. Hence, it will be wise to look at the ACF (correlation function) for the 9 pm dataset. For GBM the autocorrelation function vanishes, or is very small i.e. close to zero, this is because the increments are independent. To get a good idea of whether the 9 pm dataset follows a GBM, we shall take a look at the correlation when the data is short-ranged (in our case we shall take a look at instances when the lag is 40 and lag is 60).





Now what we notice for both lag values is that the largest correlation value is slightly greater than 0.0005, which is very close to zero. Furthermore, what we can see is that the correlation seems to reduce to almost zero as the lag value increases. Other datasets have correlation values that are also very small, such as the 1 pm dataset, but as previously mentioned the 1 pm had a large negative skewness, namely -1.5. As stated, a large skewness does reject the notion of a GBM, which is why even though both 9 pm and say 1 pm had very small correlations, it is the 9 pm that does indeed best follow a GBM.

7.0 Conclusion

We analysed a large set of data, more specifically around 72000 datasets, from the period of 1st January 1999 till 31st December 2007. Analysing datasets over a long period is known as a longitudinal study. There are several benefits to longitudinal studies, some of these advantages are: they are effective in determining variable patterns over time, they are effective in researching developmental trends and they can provide high accuracy when observing changes. These are just a few advantages of longitudinal studies. We first analysed the mean and variance of the datasets conditioned hourly, the results for some were expected while for others the results were somewhat surprising. For instance, the mean at 2 am was negative, which makes sense as most people would be sleeping, however, the surprising results were, for example, the 6 pm dataset had a negative mean, which is surprising because at 6 pm it would be rush hour, hence a lot of electricity would be utilised. We next analysed the log-returns plot and noticed extreme events present in some of the datasets, these extreme events were either positive, negative or a bit of both. We then analysed the probability density functions for each hourly dataset, followed by the log-log plot and semi-log plot. By looking at these plots, we were able to approximate an exponent for some of these datasets.

For the penultimate step, we analysed the ACF and tried to derive any possible relationship between the exponent and the ACF plot, as well as any possible relationships between the extreme events and the ACF. We concluded that there did not seem to be any possible relationship between extreme events and the ACF. Hence, it may be difficult to come up with a relationship between extreme events and the ACF when solely looking at a given dataset. Finally, we looked at the distribution without any tails present, and tried to see which dataset best demonstrated a Geometric Brownian Motion. The conclusion we reached was that the 9 pm distribution best illustrated a Geometric Brownian Motion, as it had a skewness close to zero, and had a correlation value that was very close to zero.

8.0 Reference

William, W., 2019. Lecture 1: *Introduction to time series analysis*, lecture notes, *Time series*, MTH6139, Queen Mary University of London. [Accessed 24/01/2020]

William, W., 2019. Complete lecture notes: *Introduction to time series analysis*, lecture notes, *Time series*, MTH6139, Queen Mary University of London. [Accessed 20/03/2020]

Liverani, S., 2020. Lecture 3: *Time series analysis for Business*, lecture notes, *Time series*, MTH783P, Queen Mary University of London. [Accessed 01/04/2020]

Phillips, M., 2020. Lecture 3: *Foundations in mathematical modelling in finance*, lecture notes, MTH771P, Queen Mary University of London. [Accessed 01/04/2020]

Real-statistics.com. 2020. Autocorrelation Function | Real Statistics Using Excel. [online] Available at: http://www.real-statistics.com/time-series-analysis/stochastic-processes/autocorrelation-function/ [Accessed 2 January 2020]

Bulmer, M., 2012. *Principles Of Statistics*. Dover Publications. [Online] Available at: https://books.google.co.uk/books?hl=en&lr=&id=dh24EaSrmBkC&oi=fnd&pg=PA1&dq=bulmer+1979&ots=n9sapjUZ Jv&sig=t_TFO_M2iepC5gU-4bzNlnm1zUA#v=onepage&q&f=false
[Accessed 02/04/2020]

Skewness, 2., 2020. *2.6.1. Skewness*. [online]
Statisticsbypeter.blogspot.com. Available at:
http://statisticsbypeter.blogspot.com/2015/04/261-skewness.html
[Accessed 09 February 2020].

Harari, D., 2020. *Recession And Recovery: The German Experience*. [online] House of Commons Library. Available at: https://commonslibrary.parliament.uk/economy-business/economy-economy/recession-and-recovery-the-german- experience/
[Accessed 22 March 2020].

Hunt, B., Meadows, S., Lawford, M., Hunt, B., Lawford, B., Lawford, B., Lawford, B., Williams, B. and Challis, B., 2020.
How Will Coronavirus Impact UK House Prices – And Should I Hold Off Buying A House?.
[online] The Telegraph. Available at: https://www.telegraph.co.uk/property/house-prices/coronavirus-uk-house-prices-impact/ [Accessed 1 April 2020].

Davies, R., 2020. *Coronavirus Forecast To Cut UK Economic Output By 15%*. [online] the Guardian. Available at: https://www.theguardian.com/business/2020/mar/30/coronavirus-forecast-to-cut-uk-economic-output-by-15 [Accessed 1 April 2020].

BBC News. 2020. *Coronavirus: A Visual Guide To The Economic Impact*.
[online] Available at: https://www.bbc.co.uk/news/business-51706225
[Accessed 1 April 2020].

Encyclopaedia Britannica. 2020. *Market | Economics*. [online] Available at https://www.britannica.com/topic/market [Accessed 30 March 2020].

Blackwell Global. 2020. *Top 10 Biggest Trading Exchanges In The World - Blackwell Global*. [online] Available at: https://www.blackwellglobal.com/top-10-biggest-trading-exchanges-in-the-world/
[Accessed 30 March 2020].

Corporate Finance Institute. 2020. *Financial Markets - Overview, Types, And Functions*. [online] Available at:
https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/financial-markets/ [Accessed 22 March 2020].

Energyrating.gov.au. 2020. *Legislation | Energy Rating*. [online]
Available at:
https://www.energyrating.gov.au/suppliers/legislation
[Accessed 24 March 2020].

Gausden, G., 2020. *How Has The Energy Market Changed Over The Past Ten Years?*. [online] This is Money. Available at: https://www.thisismoney.co.uk/money/bills/article-7805641/How-energy-market-changed-past-ten-years- decade-bring.html [Accessed 24 March 2020].

Eia.gov. 2020. *Delivery To Consumers - U.S. Energy Information Administration (EIA)*. [online] Available at: https://www.eia.gov/energyexplained/electricity/delivery-to-consumers.php [Accessed 1 April 2020].

The Switch. 2020. *The UK Wholesale Electricity Market & Energy Prices*. [online] Available at: https://theswitch.co.uk/energy/guides/electricity-market [Accessed 1 April 2020].

Sciencedirect.com. 2020. *Nord Pool - An Overview | Sciencedirect Topics*. [online] Available at:
https://www.sciencedirect.com/topics/engineering/nord-pool [Accessed 2 April 2020].

Nordicenergyregulators.org. 2020. *An Overview Of The Nordic Electricity Market | Nordreg*. [online] Available at: https://www.nordicenergyregulators.org/about-nordreg/an-overview-of-the-nordic-electricity-market/ [Accessed 29 March 2020].

KDnuggets. 2020. *Detecting Stationarity In Time Series Data - Kdnuggets*. [online] Available at: https://www.kdnuggets.com/2019/08/stationarity-time-series-data.html [Accessed 2 April 2020].

2020. *Time Series*. [e book] London, .Page 1. Available at: http://www.maths.qmul.ac.uk/~bb/TimeSeries/TS_Chapter4_2.pdf [Accessed 2 April 2020].

Quantivity. 2020. *Why Log Returns*. [online] Available at: https://quantivity.wordpress.com/2011/02/21/why-log- returns/ [Accessed 2 April 2020].