Impact of Daily Rhythms on Extreme Events in Electricity Price Market Data

Ghazala Nasir

Student Number: 140067631

Mathematical Finance Dissertation MTH775P, 2017/18

Supervisor: Dr Wolfram Just

School of Mathematical Sciences

Abstract

My dissertation investigates how extreme events, such as weather changes or political decisions impact electricity prices. For empirical results, we will observe the Nord Pool electricity prices from 1999 to 2007, that is given in an hourly manner, and see whether those results are what we would expect. In this thesis, we will look at the whole data set from 1999 to 2007 as well as the data sets that are conditioned on the hours of the day. We will see interesting patterns of extreme events. For example, at 1 am we observe a periodical pattern of the extreme events and analyse what could cause that. Moreover, we will investigate on what hours of the day the extreme events have a strong impact, and surprisingly, we will see that at 2 am the extreme events have such a strong impact on the electricity price, that we cannot use statistical tools such as the mean and the variance to describe the behaviour of the electricity prices for that data set. We will then conclude on our findings, which includes an insight on whether any of the data sets follow a Geometric Brownian Motion process.

Declaration of original work

This declaration is made on 04/09/2018.

I, Ghazala Nasir, hereby declare that the work in this thesis is my original work. I have not copied from any other students' work, work of mine submitted elsewhere, or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

Referenced text has been flagged by:

1. Using italic fonts or by referencing the author together with the year of publication in round brackets, and
2. Using quotation marks "…", and
3. Explicit mentioning of the sources in the last section of this thesis

Table of Contents

1.0 Introduction

The theory and practise of asset valuation over a particular time period is referred to as financial time series analysis (Tsay, 2010). Events like the financial crash in 2007, also known to as 'The Great Recession', have made it even more crucial to analyse financial data to prevent a crisis like this in the future. The recession had vast consequences. Major banks went bankrupt such as Lehman Brothers and because of economic uncertainty, the general public reduced their consumption of goods. Consequently, the profit of firms decreased which lead to higher rates of unemployment. Even today, countries such as Greece and Spain are suffering from the recession and are trying to recover and build a stronger economy. To do that, data analysis from past values is essential as one may find patterns which can help with forecasting and in return enable them to make more efficient economic decisions. A contemporary example is Brexit, whereby in June 2016, the UK voted to leave the EU. Many speculate what changes Brexit will bring, especially with regards to prices of goods as well as political changes. The uncertainty of Brexit can lead to drastic changes in the housing prices as well as electricity prices, which in return will affect the general public. The online newspaper *Independent* (2017) predicts that 10,500 jobs in finance will be lost on the first day of Brexit. Moreover, since the Brexit vote, the economy growth has slowed down. This is evident in the sterling exchange rate that dropped immensely after the Brexit vote and which is growing slowly since.

Figure 1: Musaddique S., 2018, *Sterling Exchange Rate,* The Independent. [Online] Available at: https://www.independent.co.uk/news/business/news/brexit-economy-sterling-currency-investment-cost-impact-business-financial-banks-insurance-retail-a7695486.html) [Accessed 14/08/2018]

To understand what impact such a political decision can have, it is essential to study data sets from the past and observe how extreme events like Brexit impact prices. In this thesis, we will analyse the behaviour of the electricity prices in Nord Pool from 01/01/1999 to 26/01/2007. We will investigate how the prices behave overall in this period as well as when they are conditioned on the hour of the day. First, an overview will be given of the financial market, the electricity market and Nord Pool. After this, we will discuss the concept of stationarity and returns. We will then analyse our data set in detail whereby different statistical tools will be used to understand the electricity price behaviour. Finally, we will conclude on our findings.

2.0 Background

2.1 Financial Market

A financial market is a place in the market "where buyers and sellers participate in the trade of assets" [Investopedia, 2018] . Those assets can be bonds, equities, derivatives or currencies. In every country, financial markets exist; however, they can be of very different sizes. Some markets are very small such as in Malta or Kosovo and some are huge such as the New York Stock Exchange (NYSE), where trillions of dollars are traded on a daily basis. There are various types of Markets. The following are 3 examples:

1. Stock Markets
   "Stock Markets allow investors to buy and sell shares in publicly traded companies" [Investopedia, 2018]. They contribute to one of the most important areas of market economy because it enables companies to gain capital and in return investors get some ownership of the company. If the company is successful, the gains of the investors will increase as he/she owns a fraction of the company .

2. Bond Markets
   "A bond is a debt investment in which an investor loans money to an entity (corporate or governmental" [Investopedia, 2018]. Those entities then borrow these funds for a particular period of time, whereby the interest rate is fixed. Investors buy and sell bonds on credit markets, which can be used by companies, states and governments in order to fund projects.

3. Derivatives Market
   A derivative is an economic contract whose value depends on the value of an underlying asset. Possible underlyings could be stocks or commodities such as gold, oil or electricity. People that transact derivatives can be hedgers (who take an offsetting position to reduce the risk of the investment), speculators (people that try to anticipate the price changes and buy and sell contracts to make profits) or middlemen (such as investment banks). The major types of derivative contracts are futures, which is an agreement to buy or sell an asset for a specified price at some future time, and options where the buyer of the contract has the right, but not the obligation, to buy or sell an underlying asset at a specific price, referred to as the strike price, at a specified date. Another type of derivative is swaps, whereby 2 counterparties exchange future cashflows (Boyle, 2018).

2.2 OVERVIEW OF THE EUROPEAN ELECTRICITY MARKET

In the past couple of years, there have been huge changes in the European Electricity system. The reason for that is that countries are aiming to create a low-carbon economy such that they minimize the emissions of greenhouse gases (GHG). To achieve that, changes have been taken place in the electricity market. For example, electricity is used from renewable sources, such as using solar or wind energy. As a result, electricity is used in a more efficient way as the GHG emissions are reduced from electricity generation. Moreover, changes have been made in the transport sector such as using electric vehicles and electric heating.

The Electricity System

Figure 2 gives an overview of how the electricity system works.

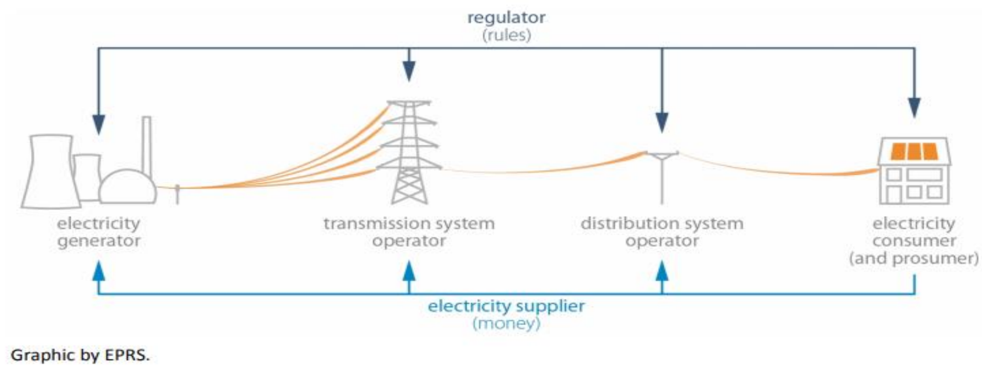


Figure 2: Erbach, G., 2016, *schematic overview of the electricity system*, European Parliamentary Research Service. [Online] Available at: http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/593519/EPRS_BRI(2016)593519_EN.pdf [Accessed 21/07/2018]

The electricity generator is a power plant whereby electricity is generated. The transmission system operator then increases the voltage so that the electricity can be transmitted. To carry electricity for long distances, transmission lines are used (U.S. Energy Information Administration, 2018). The distribution lines then transfer the electricity to consumers. This electricity system is sometimes referred to as the electricity grid and for it to be stable, one needs to ensure that the supply of electricity meets electricity demand on a constant basis otherwise failures are likely to happen (U.S. Energy Information Administration, 2018).

According to Erbach (2016), the flow of money involves the following:

- The electricity suppliers, who purchase electricity from the generators and then sell it to the consumers.
- The consumers then use the electricity and pay the suppliers through bills.
- TSO, transmission system operators, are paid for the transport of electricity that they carry out over a long distance. They also make sure that the system is stable.
- DSO, distribution network operators, deliver electricity to the consumers and are thus paid for doing so.

- Lastly, the regulators are paid for setting the rules of the system and making sure that it functions well.

The Balance of Supply and Demand

As mentioned before, for the electricity system to work efficiently, the supply of electricity has to meet the demand continuously. To achieve that, generators are used that are non-flexible (Erbach,2016). These are used to meet the normal electricity demand. Flexible generators, on the other hand, are used for the purpose of ensuring that the peaks in demand are met. However, there is an increased demand now for using energy from renewable sources. Therefore, this increased demand requires and increased capacity of flexible generators. For short term periods, the supply and demand is balanced by:

1. primary reserves, which can be activated in seconds
2. secondary reserves, that can be used in a few minutes
3. and tertiary reserves that can be set off within a 5 minutes period

Another way to balance supply and demand is through energy storage (Erbach,2016). When there is a low demand and/or a high supply, then energy is stored. It is then used when there is a high demand and/or low supply. However, storing energy is not only expensive but some energy is also lost during that process.

2.3 Nord Pool

Nord Pool, also known as the Nordic electricity market, was established in 1993, where the electricity markets in Norway, Finland, Denmark and Sweden were integrated into one single Nordic electricity market (Bergman, 2003). The owners of this market are the two national grid companies, Statnett SF, located in Norway, which owns 50% of the market, and Affaerverket Svensa Kraftnaet in Sweden, which owns 50% as well (Erzgraeber et al., 2008). In 1991, the parliament in Norway decided to deregulate the market for power trading, so between 1992 and 1995 Norway was the only country that contributed to the market. However, in 1996 Sweden joined Norway in a power exchange and consequently the power exchange was renamed to Nord Pool ASA. In 1996, EL-EX started which was a power exchange market of Finland and then in 1997 Finland joined Nord Pool. Interestingly, the western and eastern part of Denmark joined at different times. On 1st July 1999, Funen and Jutland, which is the western part, entered the Nordic electricity market and after 1st October 2000, the eastern part joined the market as well.

Electricity Consumption in Nord Pool

According to Bergman (2003), Nord Pool can be compared to the largest national electricity markets in Europe as the electricity consumption in Sweden, Norway, Finland and Denmark is greater than the European average even though the population of the Nordic countries is small. Reasons for the massive electricity use are dark and long winters with a cold climate, an intensive industry structure and a high share of electricity in the consumption of total energy. The last point reflects that Sweden, Finland & Norway have huge amounts of hydro power resources for low cost. Particularly, there is a large capacity of nuclear power plants in Finland & Sweden that are cost efficient (Bergman, 2001). This is the reason why the electricity prices have been quite low in Finland, Norway and Sweden. So, in Nord Pool there is a lot of electricity use, but the prices are relatively low. The electricity consumption is represented in Table I whereas Table II shows the electricity prices for different groups of customers.

Table I. Electricity consumption (TWh) in the Nordic countries 1999

	Denmark	Finland	Norway	Sweden
Industry	29 %	55 %	40%	43 %
Residential, services	62 %	41 %	51%	49 %
Transport	1 %	1 %	1 %	2 %
Distribution losses	7 %	4 %	8 %	7 %
Total consumption	35 TWh	78 TWh	121 TWh	143 TWh
Annual change 1990-99	1,3 % p.a.	2,5 % p.a.	1,6 % p.a.	0,4 % p.a.
Per capita consumption	6 500 kWh	15 100 kWh	26 100 kWh	15 900 kWh

Source: Elmarknaden 2000 (The Electricity Market 2000), Swedish National Energy Administration.

Table I: Elmarknaden, 2000, (as cited in Bergman 2001), *Electricity consumption (TWh) in the Nordic countries 1999, (The Electricity Market 2000),* Swedish National Energy Administration. [Online] Available at: https://pdfs.semanticscholar.org/1864/9cffe78f204586e786d29d5204fabc9502dc.pdf [Accessed 08/072018]

Table II. Electricity prices[1] (US c/kWh) in selected countries 1999

	Denmark	Finland	Norway	Sweden	Germany	France	Japan
Large industry[2]	4,8	2,9	2,0	2,3	5,1	4,0	4,7
Middle industry[3]	5,2	4,0	2,7	2,9	6,7	5,5	9,5
Small industry[4]	5,4	4,5	3,7	4,0	8,2	6,5	11,7
Household, 20 MWh/yr	14,8	4,9	5,5	7,5	8,2	9,2	7,5
Household, 3,5 MWh/yr	17,5	8,1	9,4	8,7	17,0	13,2	17,1

Source: Elmarknaden 2000 (The Electricity Market 2000), Swedish National Energy Administration.

Table II: Elmarknaden, 2000, (as cited in Bergman 2001), *Electricity prices (US c/kWh) in selected countries 1999, (The Electricity Market 2000),* Swedish National Energy Administration. [Online] Available at: https://pdfs.semanticscholar.org/1864/9cffe78f204586e786d29d5204fabc9502dc.pdf [Accessed 08/072018]

It is noticeable in Table I that, in the 1990's, there was a slow growth in the electricity consumption. The reason for this is that in the first part of that decade, there was a major slowdown of economic growth. However, in the last couple of years the economies have grown and so has the consumption of electricity, so the growth rate of electricity has increased (Bergman, 2001). Moreover, the demand for electricity has grown in Sweden due to the use of oil burners and electric boilers in the households. Thus, due to the increasing price of oil, the demand for electricity has grown. When looking at Table II for the electricity prices, we observe a major difference between the prices in Denmark and the other Nordic countries. The low prices in the other Nordic countries are a reflection of low cost hydro power that I mentioned earlier.

Bergman (2002) states that there are 2 reasons for why there is a significant difference between Denmark and the other Nordic countries, when it comes to electricity consumption and prices. The first reason is that in Sweden, Finland and Norway there is a big focus on electricity intensive industries and the second reason is that electric heating is used far more in these countries when compared to Denmark.

Table III. Electricity production (TWh) in the Nordic countries in 1999

	Denmark	Finland	Norway	Sweden
Hydro power	0	13	122	70
Wind power	3	0	0	0
Nuclear power	0	22	0	70
Condensing power (fossil fuels)	24	7	1	0
Combined heat and power	10	25	0	10
Total production	37	67	123	150

Source: Elmarknaden 2000 (The Electricity Market 2000), Swedish National Energy Administration.

Table III: Elmarknaden, 2000, (as cited in Bergman 2001), *Electricity production (TWh) in the Nordic countries 1999, (The Electricity Market 2000),* Swedish National Energy Administration. [Online] Available at: https://pdfs.semanticscholar.org/1864/9cffe78f204586e786d29d5204fabc9502dc.pdf [Accessed 08/072018]

Table III shows the production of electricity in Nord Pool. It is interesting to observe that the 4 countries have very different electricity productions. For example, there is no nuclear power in Norway & Denmark whereas there is a substantial amount of electricity production on nuclear power in Sweden and Finland. One of the controversial results is that the use of environmentally friendly electricity technologies is encouraged by all Nordic countries, but a significant use of wind power only takes place in Denmark (Bergman, 2001).

Overall, it can be argued that the Nordic electricity market has been a success to a great extent. According to Bergman (2002), the market reforms in electricity in Nord Pool have been more successful when compared to the EU electricity Market directive. Some of the successful features involve the maintenance of the supply reliability, and not only have the electricity prices decreased but also productivity has increased in the electricity supply industry. This demonstrates that, in fact, competition can lead to lower prices and more efficiency, which is ideal to an electricity consumer. One might think that market power could be a problem since there is only one common exchange of 4 nations, however, the integration of the national markets does not appear to be a big problem in regards to market power (Bergman, 2002).

2.4 Stationarity

An important aspect of time series analysis is stationarity of a data set. Before introducing the formal definition of stationarity, we define the autocovariance function:

Definition 1 (Giraitis, 2017)

Let X_t be the observations at time $-\infty < t < \infty$. The Autocovariance function of (X_t) is $Cov(X_t, X_s) = E[(X_t - E[X_t])(X_s - E[X_s])]$.

There are 2 types of stationarity, that are referred to as strict stationarity and second order/covariance stationarity.

Definition 2 (Giraitis, 2017)

(X_t) is a strictly stationary sequence, if for any integers $t_1, t_2, \dots, t_k$ and for any h, the joint distribution of $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$ is the same as the joint distribution of $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})$: $\left(X_{t_1}, X_{t_2}, \dots, X_{t_k}\right) =^{d} (X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})$.

The definition above means that we require the joint distribution of $\left(X_{t_1}, X_{t_2}, \dots, X_{t_k}\right)$ to be invariant, i.e. not change when we shift the time. However, according to Tsay (2010), strict stationarity is difficult to prove by using empirical data. Thus, one can consider second order/covariance stationarity, which is a weaker version of strict stationarity.

Definition 3 (Giraitis, 2017)

We say that the time series (X_t) is second order or covariance stationary, if:

I. $E[X_t] = \mu$, where μ is independent of t
II. $Var(X_t) = \sigma^2$, where σ^2 is independent of t
III. $Cov(X_t, X_{t+k}) \equiv \gamma_k$, (for all t and k) only dependent on k

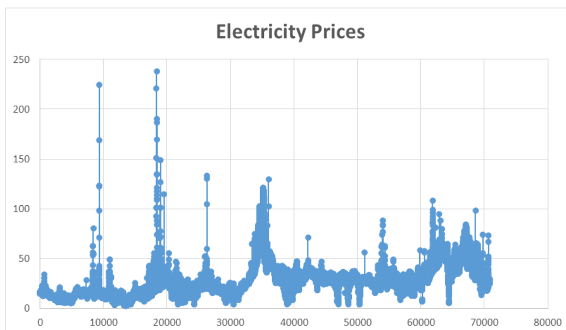
Thus, for a covariance stationary time series, we need the mean and variance to be finite and independent of time. The autocorrelation function (γ_k) also needs to be independent of time. Moreover, strict stationarity implies covariance stationarity if $E[X_t] < \infty$ and $Var(X_t) < \infty$. However, in general, the converse is not true (Giraitis, 2017). If one of the conditions above fails, then we have a non-stationary time series.

2.5 Returns

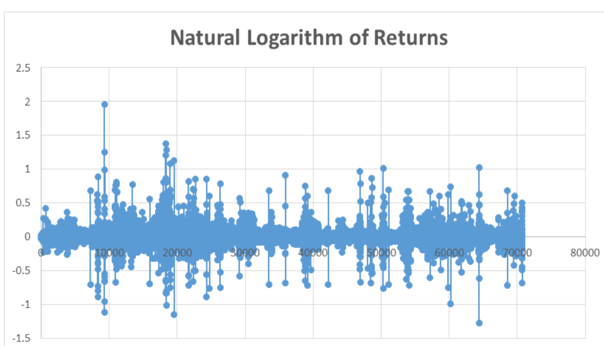
When studying a financial data set, it often desired to use returns, in particular log-returns. The reason for this is that we can normalize, meaning that returns enable us to measure our variables with a metric that one can compare with. This helps us to evaluate analytical relationships between 2 or more variables. There are 2 definitions of returns.

1. Simple return (Giraitis, 2017):
   Let us define P_t to be the price of an asset at time t. Holding the asset for one period from time $t-1$ to t gives us the simple gross return: $1 + R_t = \frac{P_t}{P_{t-1}}$
   The corresponding simple return (or simple net return) is defined as: $R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}}$
2. Log return (Giraitis, 2017):
   The natural logarithm of the simple gross return is called log-return, which is defined as:
   $r_t = \ln(1 + R_t) = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(P_t) - \ln(P_{t-1})$
   We note that for small R_t we have that $\ln(1 + R_t) \sim R_t$.Thus, $r_t \sim R_t$, and so the log-returns are close to the raw returns in terms of their value.

The Nord Pool data set is given from 01/01/1999 to 26/01/2007. It contains the electricity prices (in EUR/MWh) at a sampling rate of one hour. For our analysis of this data set (as well as the data sets that are conditioned on the hour of the day), we will be using log-returns, with a logarithm of base e, so a natural logarithm. The following is a plot of the electricity prices of the whole data set, that contains about 71000 data points. The x-axis represents the consecutive hours and the y-axis shows the electricity prices.



We observe from the plot above that the prices have huge fluctuations and the data seems to be seasonal with a slight upward trend. The plot of the natural logarithm of returns is the following:



From the plot above, we see high and low peaks, which represent extreme events and thus show extreme price fluctuations. In this thesis, we will analyse on what hours of the day these extreme events occur and how strong their impact is on the electricity prices.

3.0 Analysis

3.1 Mean/Variance & Conditioned Time Series Plots

We will now consider the mean and variance of the log-returns of the whole data set as well as for the data sets that are conditioned on the hours of the day. To get the conditioned data sets, we use the function 'OFFSET' in Excel. The following will be an overview of the mean and variance values. We will then analyse the time series plots of the conditioned data sets, where a more detailed discussion of the mean and variance values will be given. To find the mean value of a data set in Excel, we use the function 'AVERAGE' and to get the variance we use 'VAR'.

The mean of the whole data set is: 0.000006692666, which is very small and can be considered to be zero. The following table illustrates the mean values of the data sets that are conditioned on the hour. The values are given to 5 decimal places.

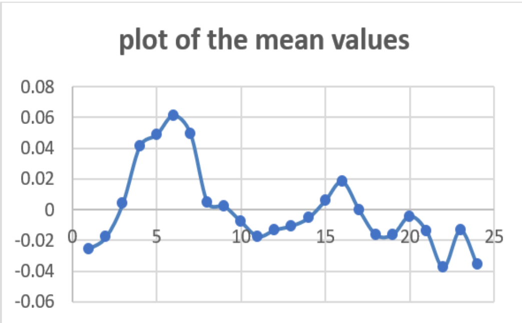
Time	12 am	1 am	2 am	3 am	4 am	5 am	6 am	7 am	8 am
Mean	-0.03579	-0.02568	-0.01788	0.00415	0.04123	0.04870	0.06142	0.04965	0.00467

9 am	10 am	11 am	12 pm	1 pm	2 pm	3 pm	4 pm	5 pm
0.00218	-0.00782	-0.01743	-0.01329	-0.01061	-0.00547	0.00586	0.01827	-0.00015

6 pm	7 pm	8 pm	9 pm	10 pm	11 pm
-0.01634	-0.01630	-0.00428	-0.01391	-0.03762	-0.01342

What we notice is that the mean value of the whole data set is very small when compared to the mean values of the conditioned data sets. The mean values for the conditioned data set ranges from -0.03762 to 0.06142, so the average return is the smallest at 10 pm and the highest at 6 am. Moreover, the mean values are negative from 5 pm to 2 am. Negative mean values mean that prices go down and positive mean values mean that prices go up. Overall, just from looking at the mean values, it seems sensible to have negative mean values from 9 pm to 2 am as people usually sleep during this time, so not much electricity is used and hence the price goes down. However, it is surprising that the mean values of 5 pm and 6 pm are negative as one would expect them to be positive because these are rush hour times. The reason for why that could be will be discussed later when we will look at the variance of the data sets.

After looking at the mean values, I thought it would be interesting to plot the mean values in a graph to observe if there is a specific pattern. The following is the plot of the mean values which goes from 1 am to midnight.



From the plot one observes interesting oscillatory patterns. The values of the oscillations become smaller throughout the day. It is noticeable that the mean keeps on changing and is thus not constant. This goes against the first property of stationarity, which we defined earlier.

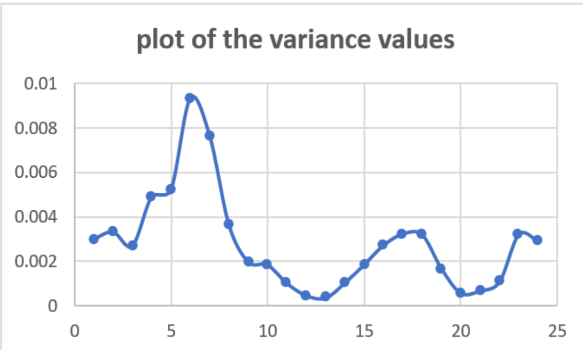
We will now look at the variance of the whole data set and the conditioned data sets. The variance of the whole data set is: 0.003491878006. The following table consists of the variance values for the time series that are conditioned on the hour.

Time	12 am	1 am	2 am	3 am	4 am	5 am	6 am	7 am	8 am
Variance	0.00293	0.00301	0.00333	0.00272	0.00492	0.00522	0.00935	0.00766	0.00364

9 am	10 am	11 am	12 pm	1 pm	2 pm	3 pm	4 pm	5 pm	6 pm
0.00198	0.00186	0.00102	0.00047	0.00037	0.00105	0.00186	0.00273	0.00325	0.00321

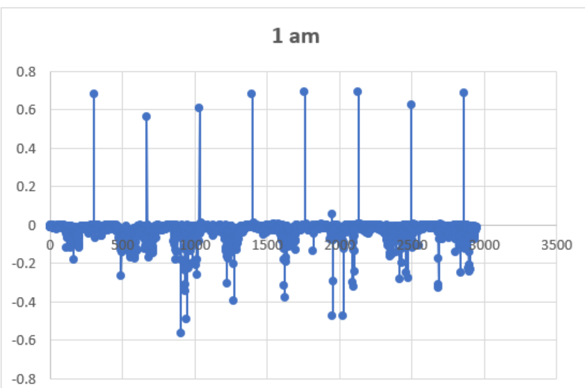
7 pm	8 pm	9 pm	10 pm	11 pm
0.00167	0.00058	0.00068	0.00109	0.00321

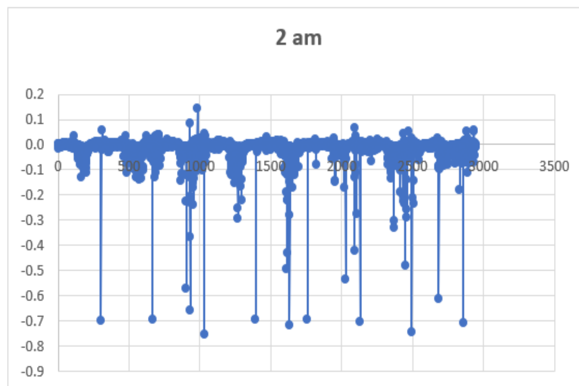
The variance is a measure for the volatility of the market. If the variance is high, then that means that the market is highly volatile, so the prices go up and down drastically. The smaller the variance the more ‘constant’ the spread of the data is. If the variance is zero, then the data points are evenly distributed, so we do not have outliers that lead to fluctuations in the data set. From the table above, we can see that the variance ranges from 0.00037 to 0.00935, so the variance is smallest at 1 pm and highest at 6 am, thus we have the least fluctuations of the price at 1 pm and drastic fluctuations in the price at 6 am. The following hours of the day have the same variance as the whole data set to 3 decimal places: 12 am, 1 am, 2 am, 3 am, 4 pm, 5 pm, 6 pm & 11 pm. The hours of the day that have a small variance and thus very small fluctuations of the electricity price are at 12 pm, 1 pm, 8 pm and 9 pm. From 11 pm to 3 am the variance is constant, i.e. the same (to 3 decimal places). As mentioned for the mean values, during this period most people are sleeping, so we would not expect the price of the electricity to fluctuate much. I mentioned in the analysis of the mean that it was surprising to see that the mean at 5 pm and 6 pm is negative. One reason for this may be the variance. The variance at 5 pm and 6 pm is 0.003 (to 3 decimal places), which is higher than most hours. A higher variance means that the data is more spread out, so we have outliers. Thus, it is likely that we have negative outliers in the 5 pm and 6 pm data set and as a result of that, the mean value turned out to be negative. To see if the variance shows a specific pattern I plotted the graph of the variance of the data sets that goes from 1 am to midnight.



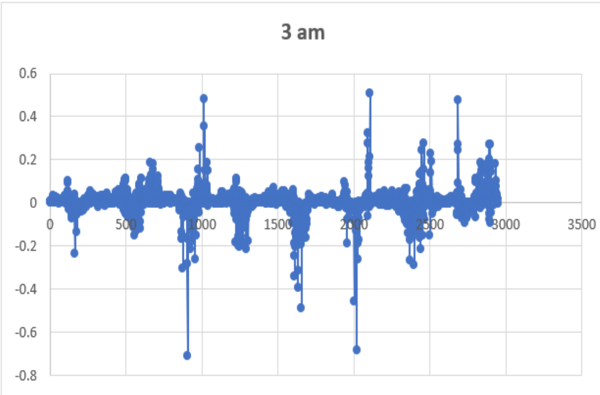
From the variance plot we observe that there are huge fluctuations between 1 am and 9 am and we see that the fluctuations are almost periodic after 10 am. As with the mean values, we notice that the variance is not constant but changes over time, which goes against the second property of stationarity. Thus, we can say that our data sets are not covariance stationary.

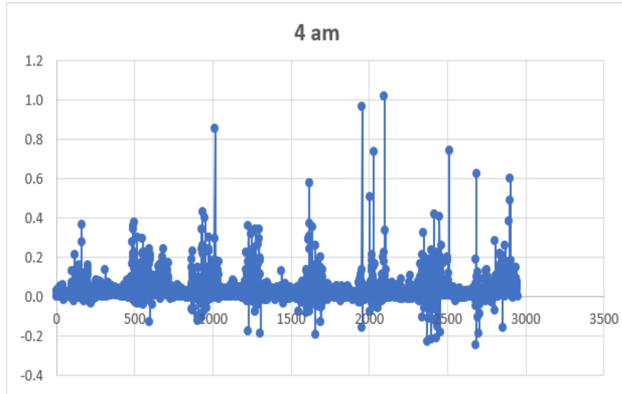
We will now consider the plots of each hourly data set and investigate in more detail how the mean and variance changes from one time to another. We will also observe if any extreme events are present and whether those are positive or negative.



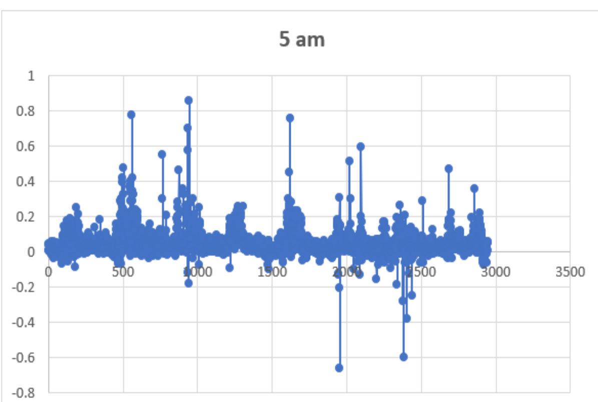


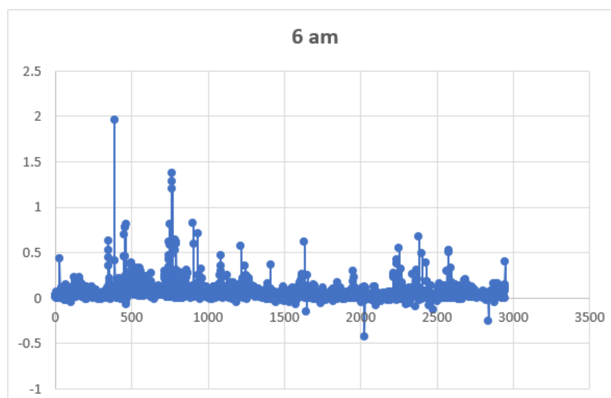
In the 1 am plot we notice positive extreme events that seem to be periodic. The extreme events happen once every year and as they are positive, it means that there is a significant price increase from 12 am to 1 am at one specific day every year. As there seems to be a periodic pattern, it is likely that the price increase is not caused by economic demand but rather by some computer update; for example, power networks might have an upgrade during those days. It would make sense to do a power upgrade at 1 am as most people would be sleeping at this time and would thus not be affected by it. During this time the mean is negative (-0.02568), so on average the electricity price goes down from 12 am to 1 am and the variance is the same as for the whole data set (to 3 decimal places). In the 2 am plot there are negative extreme events, so there is a price decrease from 1 am to 2 am. However, there seems to be no specific periodic pattern of the extreme events. The mean of 1 am and 2 am are very similar to 2 decimal places (-0.03 & -0.02 respectively) and the variance is the same to 3 decimal places (0.003), so the dispersion of the data is constant from 1 am to 2 am. It is noticeable that at 1 am we have extreme positive spikes and at 2 am we have extreme negative spikes but both plots have peaks of around 0.7 (in terms of the modulus). So, when considering the modulus, the value of the spikes of 1 am and 2 am are very similar and thus the similar variance of these 2 time series plots makes sense.



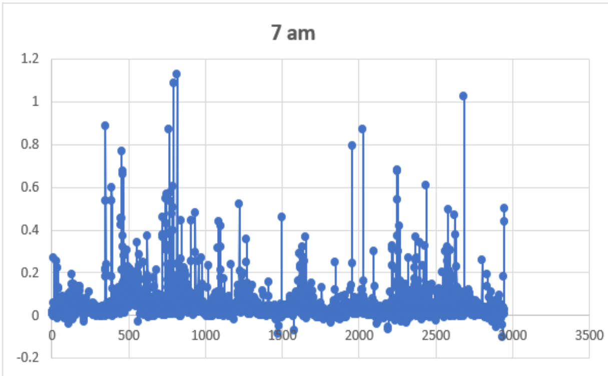


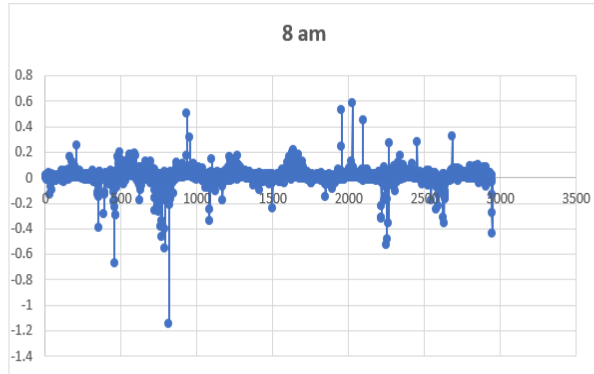
At 3 am there are positive and negative extreme events. The mean at 3 am (0.00415) is larger than at 2 am (-0.01788). Thus, on average, there is a price increase from 2 am to 3 am. However, the variance from 2 am to 3 am is the same to 3 decimal places (0.003), showing that the price fluctuations are fairly constant between 2 am and 3 am. This is what we would expect as normally there is not a huge change in the use of electricity between 2 am and 3 am as it is night time. At 4 am we observe positive extreme events, so there is a price increase from 3 am to 4 am. This can also be seen from the mean as the mean at 4 am (0.04123) is significantly larger than at 3 am (0.00415). The variance at 4 am (0.00492) is larger than at 3 am (0.00272). Thus, the price fluctuations are not constant anymore.



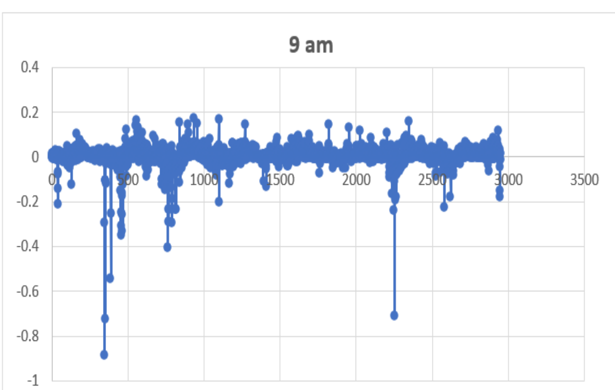


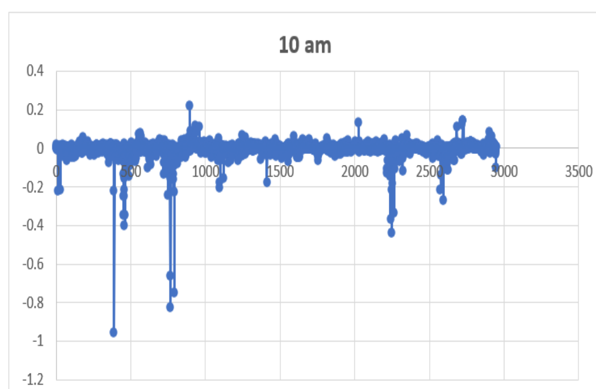
At 5 am there are positive extreme events, so the electricity price goes up between 4 am and 5 am. This is what we would expect as some households wake up during this time and thus more electricity is used. The mean at 4 am (0.04) and 5 am (0.05) are nearly the same to 2 decimal places and their variances are the same to 3 decimal places (0.005), so the price fluctuation between 4 am and 5 am is constant. At 6 am we also have positive extreme events, and the values of some extreme events are higher than the values at 5 am; for example, the highest value that an extreme event has at 6 am is 2 whereas the highest peak of the extreme events at 5 am is 0.85. This may be the reason why the variance of 6 am (0.009) is much higher than the variance at 5 am (0.005) and thus we have major price fluctuations between 5 am and 6 am. Moreover, the mean value of 6 am (0.06142) is larger than at 5 am (0.04870), indicating a price increase between 5 am and 6 am.



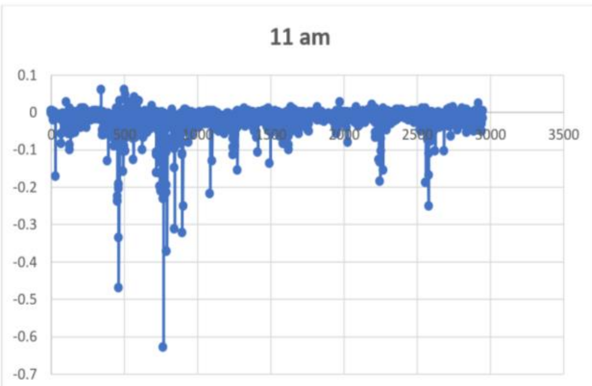


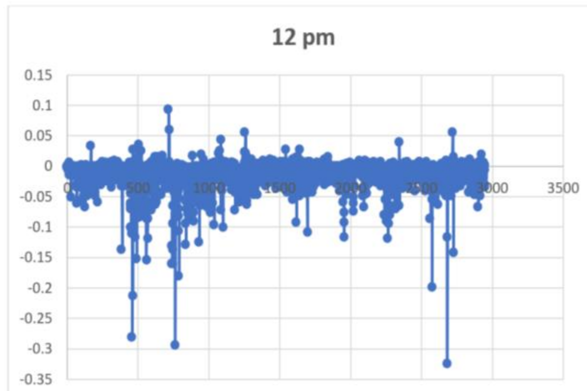
At 7 am we have positive extreme events, so the electricity price goes up between 6 am and 7 am. During this time there are the most extreme events when compared to any other hour of the day. This is expected as most people start waking up during this time for work/school and this is a rush hour time. Thus, more electricity is used in households and company buildings, so the price of electricity increases. When looking at the plot, one would expect the variance of 7 am to be larger than at 6 am, but in fact the variance at 7 am is smaller. The reason for this may be that at 7 am the extreme events are smaller than 1.2 whereas at 6 am some extreme events are higher than 1.2. So, the data is more spread out at 6 am and this may be the reason why the variance at 6 am (0.00935) is higher than at 7 am (0.00766). As we have a smaller variance at 7 am, it would indicate that there are less price fluctuations when compared to 6 am. However, the variance at 7 am is still high when compared to other hours of the day, so in average the price at 7 am does in fact fluctuate more than in other hours. At 8 am we observe a few positive and negative extreme events, but the prices seem to be fairly constant. This is also evident in the variance as the variance at 8 am (0.00364) is much smaller than at 7 am (0.00766), thus we do not have many outliers at 8 am that would lead to high price fluctuations. Moreover, the mean at 8 am (0.00467) is significantly smaller than the mean at 7 am (0.04965), thus indicating that the electricity price decreases between 7 am and 8 am.



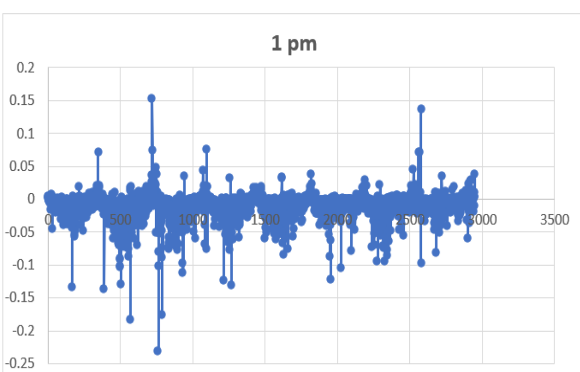


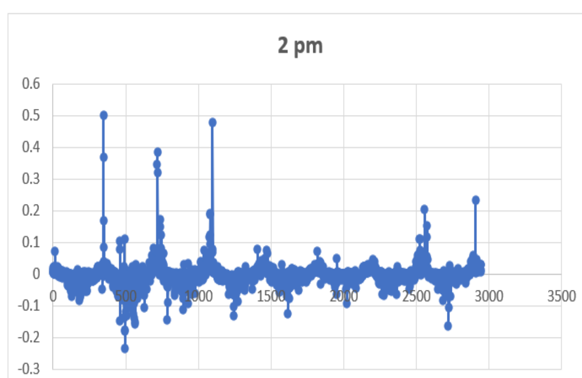
We observe some negative extreme events at 9 am and 10 am, which seem to occur randomly that could be due to ecological reasons such as unexpected weather changes, but generally the electricity prices seem to be constant during these times. Moreover, the variance between 9 am and 10 am is the same to 3 decimal places (0.002) and therefore the price fluctuations between 9 am and 10 am are constant. The mean at 9 am (0.00218) is smaller than the mean at 8 am (0.00467) and the mean at 10 am (-0.00782) is smaller than the mean at 9 am, indicating that the prices go down. The reason for the price decrease may be that most people are at school/work during that time and not much electricity is used in the households.



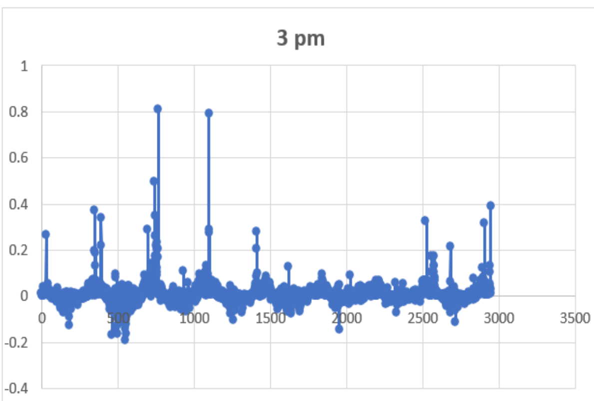


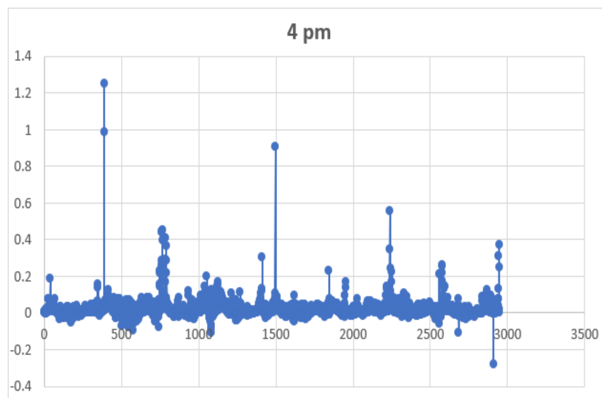
The 11 am data set has negative extreme events and looks similar to the 10 am plot. This is evident in the variance as the variance at 11 am (0.001) and at 10 am (0.002) are similar to 3 decimal places, thus the price fluctuations are fairly constant in that time period. However, the mean at 11 am (-0.01743) is smaller than the mean at 10 am (-0.00782), meaning that there is a price decrease between 10 am and 11 am. At 12 pm we have negative extreme events as well. It is interesting to observe that the variance at 12 pm is very small (0.00047) since when looking at the plot one may expect that we have a lot of extreme events and thus the variance would be higher. The reason for the small variance may be that the values of the extreme events are quite small, with the highest being 0.325 (when considering the modulus). The mean values between 11 am (-0.02) and 12 pm (-0.01) are very similar to 2 decimal places, so there is not a huge change in the price during that time period.



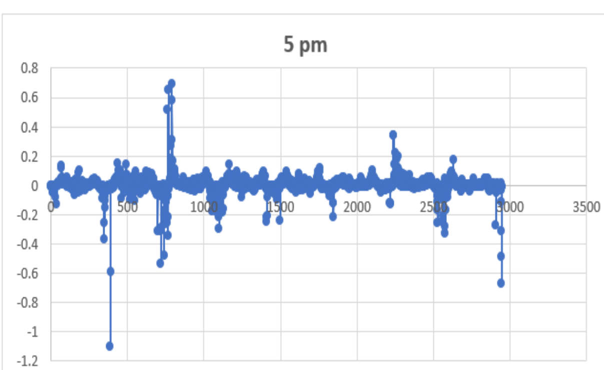


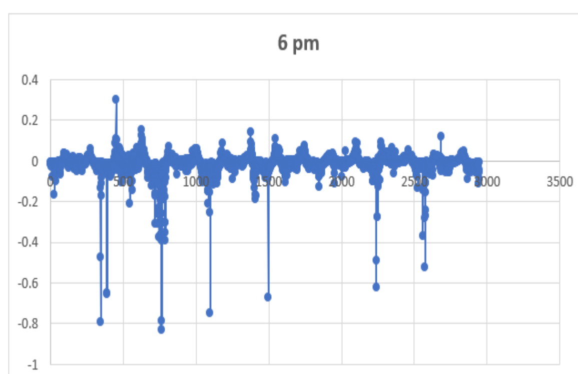
In the 1 pm plot we observe a similar behaviour as in the 12 pm plot. This is evident in the variance values as the variance is very similar to 4 decimal places between 12 pm (0.0005) and 1 pm (0.0004) and the mean values are the same to 2 decimal places (-0.01), thus we do not see any drastic changes in the electricity price. However, we notice a change in the price at 2 pm as the mean at 2 pm (-0.00547) is larger than at 1 pm (-0.01061) so the prices increase in that period. Moreover, we see positive extreme events at 2 pm, which seem to occur in a random manner. When considering the modulus, we observe that the extreme events at 2 pm have higher values when compared to the extreme events at 1 pm; for example, the highest value of an extreme event at 2 pm is 0.5 whereas at 1 pm it is 0.225. Since the values at 2 pm are higher, we would expect that the data is more dispersed. This is supported by the variance at 2 pm (0.00105) as this is larger than at 1 pm (0.00037) and thus we experience more price fluctuations during this time.

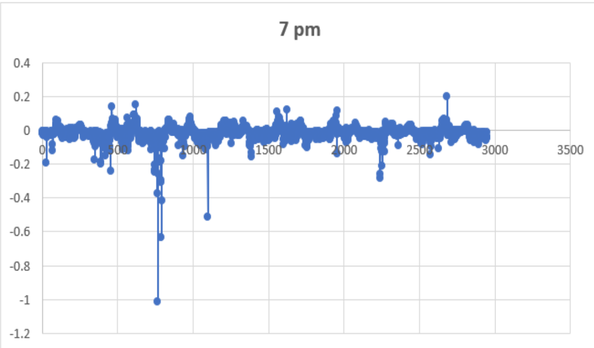


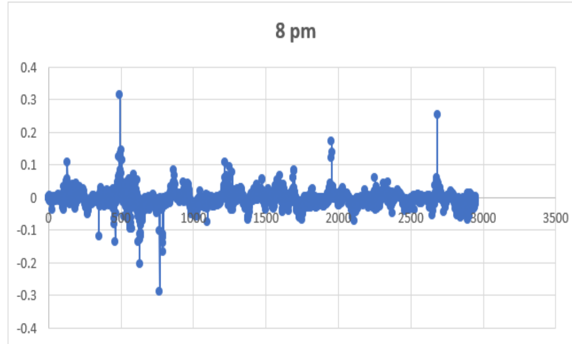


There is a price increase between 2 pm and 3 pm as the mean at 3 pm (0.00586) is much higher than at 2 pm (-0.00547). Also, the mean at 4 pm (0.01827) is larger than at 3 pm, so the prices also increase between 3 pm and 4 pm. The reason for this may be that most children come home from school during that time, so more electricity is used in the households. Even though there are some positive extreme events in both plots, the price fluctuations are fairly small as the variance between 2 pm (0.00105), 3 pm (0.00186) and 4 pm (0.00273) are close in value.

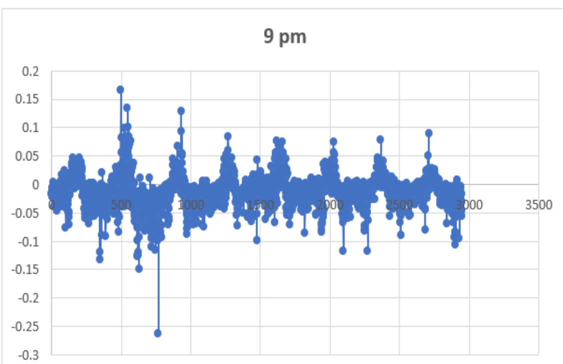


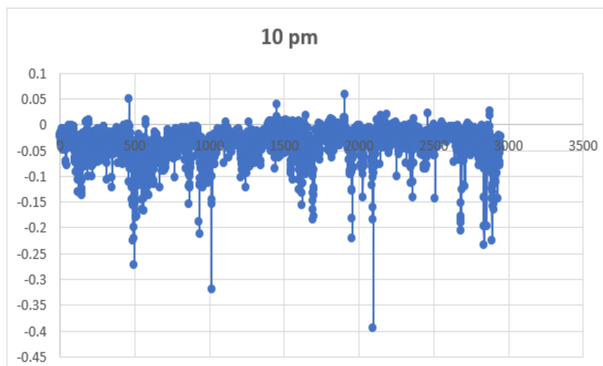


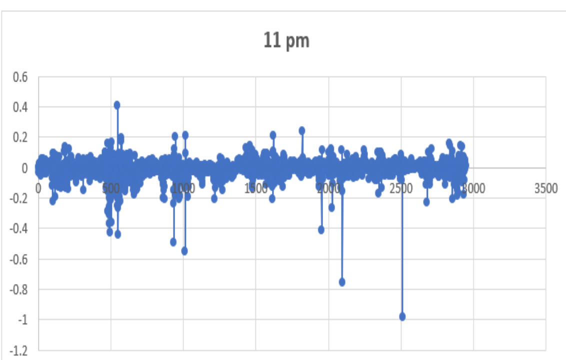


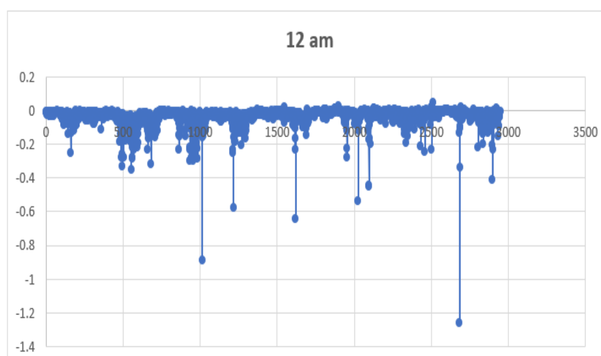


The plots from 5 pm to 8 pm look quite similar. The plots of 5 pm and 8 pm have positive and negative extreme events whereas at 6 pm and 7 pm we have negative extreme events. The mean value becomes smaller from 5 pm (-0.00015) to 7 pm (-0.01630), so the prices decrease in that period. The variance between 5 pm and 6 pm are the same to 3 decimal places (0.003) but then it decreases at 7 pm (0.00167) and at 8 pm (0.00058), thus getting closer to zero, so the price fluctuations become smaller from 5 pm to 8 pm.









From 9 pm to 12 am we observe negative extreme events, which one would expect as most people sleep during this time and thus in most households electricity is not used, so the prices decrease. There is an interesting pattern of the mean values between 9 pm to 12 am: -0.01, -0.04, -0.01, -0.04, so the prices tend to go up and down almost in a periodic manner. Moreover, the variance increases between 9 pm (0.00068) to 11 pm (0.00321) so the prices tend to fluctuate much during this period.

However, the variance at 12 am (0.00293) is the same as the variance at 11 pm to 3 decimal places and thus the fluctuations of the electricity price are fairly constant between 11 pm to 12 am.

When looking at the plots of the different hours of the day, we observed that there were sometimes positive and/or negative extreme events that do not have any specific pattern. The following are a few reasons why electricity prices can fluctuate and lead to those extreme events:

1. **Weather**: severe weather such as storms can lead to power cuts, which in return will affect the electricity price as the prices will increase (Exchange Utility, 2018).
2. **Supply & Demand**: Since the discussion on green energy started, more businesses want to become environmentally friendly in their electricity use. In those cases, sometimes supply is greater than demand, so the prices go down and hence affect the electricity price.
3. **Infrastructure**: electricity prices can be affected by the cost of maintenance and the cost of construction of power stations. If the cost of these is too high, then the electricity prices will be increased (Exchange Utility, 2018).
4. **Randomness**: However, some extreme events may just occur by ‘randomness’, where an explanation may simply be that some people used more or less electricity for no specific reason.

In the analysis above, we used the mean and variance to observe the behaviour of the price changes. However, using the mean and variance has advantages and disadvantages. The advantages are that both can be computed easily and both values take into account the whole data set rather than just specific points of a data set. On the other hand, the mean is highly affected by extreme values, so it may not be a true representation of the data points that have outliers. Moreover, the mean may not be appropriate to use for highly skewed data (Laerd Statistics,2018) and the same holds for the variance (Manikandan, 2011). This is important to consider as later we will analyse the distributions of the data sets and we will see that some distributions are in fact highly skewed and it may be that the mean and variance cannot be used to describe those data sets.

3.2 Probability Density Functions

We will now consider the probability density functions of the whole data set and the data sets that are conditioned on the hour. To get the probability density function (PDF) we need to do the following:

Let N be the number of data points. We first sort out the data points from smallest to largest, such that we get a series of $X = \{x_1, \dots, x_N\}$. To get the PDF, we need to consider how to get the cumulative distribution function (CDF), as the PDF is the derivative of the CDF. The CDF gives us the probability of a variable that is less than or equal to x. In our case, we have to get a numerical CDF so we define the CDF to be: $Y = \{c(x_1), c(x_2), \dots, c(x_N)\}$.

To get the numerical CDF, we calculate the values by doing the following:

$$c(x_1) = \frac{1}{N}, c(x_2) = \frac{2}{N}, \dots, c(x_N) = \frac{N}{N} = 1$$

Thus, we plot X against Y to get the CDF. To find the PDF, we need to take the derivative of the (numerically defined) CDF by using a suitable difference quotient. Difference quotients are used to approximate a numerical differentiation. Since the x- values are not equally spaced, we can consider the inverse function first as the y-values are equally spaced and then take a simple first order difference quotient which is: $\frac{\mathrm{x_{k+1}-x_k}}{\mathrm{y_{k+1}-y_k}}$, for $k = 1, \dots, N-1$. That is the derivative of the inverse function, so the inverse of the fraction, which is the derivative of the CDF, i.e. the PDF is: $\frac{y_{k+1}-y_k}{x_{k+1}-x_k}$, for $k = 1, \dots, N-1$. However, one can get more accurate results by taking a higher order difference quotient. For my data set I used the second order difference quotient which is: $\frac{y_{k+1}-y_{k-1}}{x_{k+1}-x_{k-1}}$, for $k = 2, \dots, N-1$.

The above is an overall explanation on how to get the PDF. I will now explain step-by-step how to get the PDF in Excel explicitly and what problems may occur.

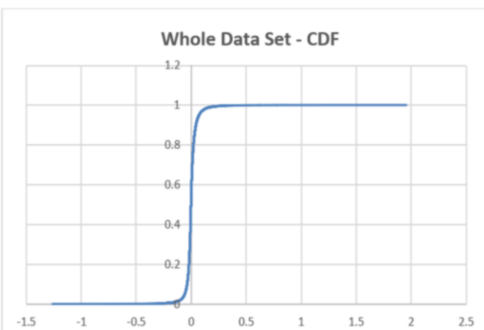
PDF in Excel

One of the important things to remember, when doing the calculation in Excel, is that when we create a column that contains the data of $\mathrm{x_{k+1} - x_{k-1}}$, we might get values that are zero. In Excel, it either shows as an error that we are dividing by zero, or sometimes if the data set is large (here about 3000 data points for each hour of the day), what Excel does is convert the number that we are dividing by zero into a huge number, such as 13 000 000 (as mathematically 1/0 diverges to infinity). To prevent that, we just delete the data points that contain those zeros. Deleting data points like this is sometimes referred to as 'cleaning' the data set. It is important to remember to not only delete the zeros in that one column, but to delete the sorted x- values that lead to these zeros. This is important as the number of data points, which is N, will change since we have deleted some data points. Thus, to get accurate results for the PDF we do the following steps:

1. We first sort the data points from smallest to largest in column A (which is the 1st column) of Excel. This gives us $x_1, \ldots, x_N$. However, $'N'$ is not the last data point that we use for the calculation of the CDF (as we might have some zeros in the quotient, so the number of data points will change and so will the CDF). Thus, before calculating the CDF, we need to find the data points that need to be deleted first, which is our second step.

2. In the second column, i.e. column B, we calculate $\mathrm{x}_{\mathrm{k}+1} - \mathrm{x}_{\mathrm{k}-1}$.To find the zeros, we use CTRL+F. After we find those zero points, we delete them from column B and the data points in column A that correspond to it.

3. We now have the accurate number of data points, that we define as $'n'$. In column C we calculate the CDF, so: $C(X) = \{\frac{1}{n}, \frac{2}{n}, \ldots, 1\}$. We then plot column A versus column C. The following is the CDF of the whole data set:



We observe that for $x < 0$, the function decays to zero and for $x > 0$, the function gets closer to 1. However, we are more interested in the probability density function of the whole data set and the conditioned data sets.

4. We now calculate $\mathrm{y}_{\mathrm{k}+1} - \mathrm{y}_{\mathrm{k}-1}$ in column D. In column E we then divide column D by column B, so we calculate $\frac{y_{k+1} - y_{k-1}}{x_{k+1} - x_{k-1}}$. This gives us the values of the PDF. However, since the data set is very noisy, the PDF values need to be smoothed, and then we can plot the smoothed PDF.

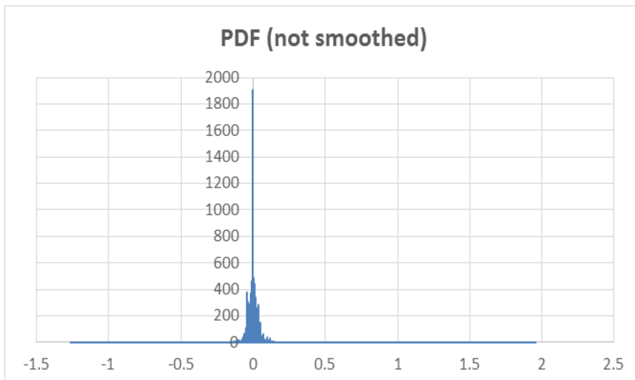
Smoothing a data set

As mentioned above, the data set is noisy, so it is desired to smooth the data set. The aim of smoothing a time series is to eliminate the noise. A way to smooth the data set is by using a simple moving average, which is implemented in the following way:

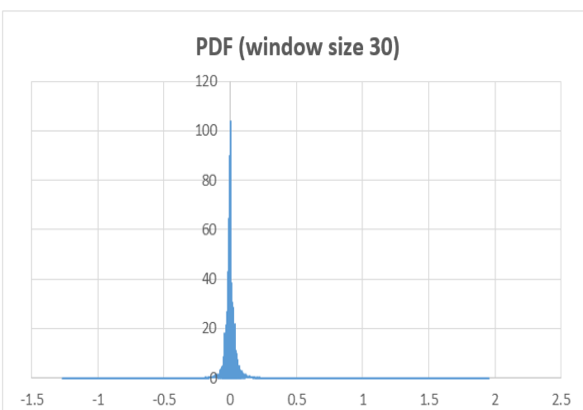
Let us assume that we have a time series containing data points $x_1, \ldots, x_n$. Then, a moving average with a window size of N means to average over the next N points. For example, if we have $N = 5$ then we need to average over the next 5 points. Thus, we define a new time series, which we call z_t, that will have the following values:

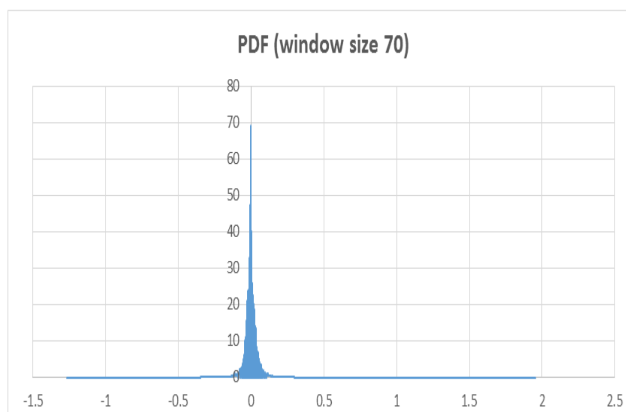
$$z_1 = \frac{1}{5}(x_1 + x_2 + x_3 + x_4 + x_5)$$
$$z_2 = \frac{1}{5}(x_2 + x_3 + x_4 + x_5 + x_6)$$
$$\dots$$
$$z_n = \frac{1}{5}(x_{n-4} + x_{n-3} + x_{n-2} + x_{n-1} + x_n)$$

Once the moving average has been calculated, we plot the sorted data set (that we defined as X earlier) versus the moving average values to get the smoothed PDF. For my data set, I tried different window sizes. The following is the result of the PDF of the whole data set when it is not smoothed:



We can clearly see that the PDF needs to be smoothed. One needs to try different window sizes in order to get a smooth PDF. For a window size of 30 and 70 the PDF looks like this:





When using the 30-order simple moving average, the PDF becomes smoother but a 70-order moving average gives the smoothest results. I will thus be using a 70-order moving average for the conditioned data sets as well.

When plotting the probability density functions in Excel, we notice that some distributions have long right and /or long left tails. For the distributions that have long right or long left tails we will investigate if a power law exists. To do that we use log-log plots. This is when we take the logarithm of the x-values and y-values of the distribution. For our data set, we will be using the logarithm of base 10. If we can see a straight line in the log-log plot, then that means that we have a power law that takes the following form: $y = b * x^n$. The power, i.e. the exponent $'n'$, is the slope of the line from the log-log plot. The constant $'b'$ is found by trying out different numbers until a plot fits the curve. However, if we do not see a straight line in the log-log plot, then we consider a semi-log plot. In this case the x-values stay the same and we take the logarithm of the y-values of the distribution. If we observe that there is a straight line in the semi-log plot, then we have an exponential fit for the right/left tail of the distribution that has the form: $y = k * 10^{mx}$. $'m'$ is the gradient of the line in the semi-log plot. Again, the constant $'k'$ is found by trying different numbers until a function fits. However, if we do not see a straight line in the semi-log plot, then that means that we do not have an exponential fit either.

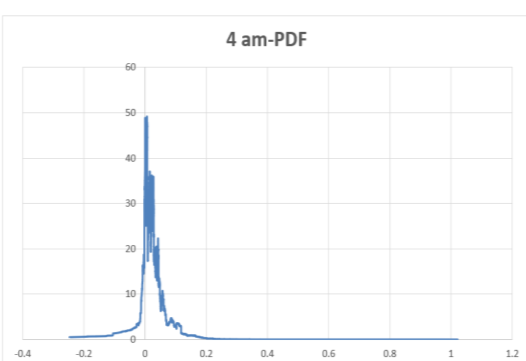
Skewness

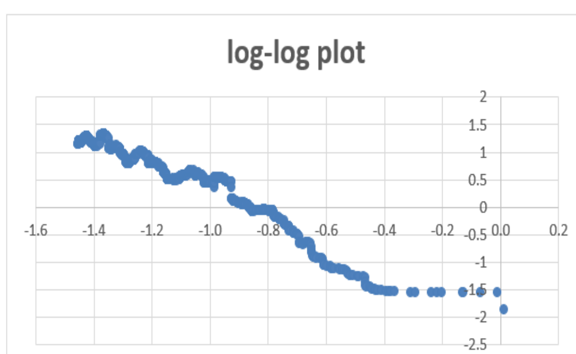
We will also analyse the skewness of the distributions. The skewness is a measure for the asymmetry of the distribution. If the skewness is positive, then the right tail of a distribution is longer than the left tail and if it is negative then the left tail of a distribution is longer than the right tail. The following is how we characterise the extent of the skewness (taken from GoodData Corporation, 2018):

1. If the skewness is less than -1 or greater than +1, then we have a highly skewed distribution
2. If the skewness is between -1 and -0.5 or between 0.5 and 1, then we have a moderately skewed distribution
3. And lastly, if the skewness is between -0.5 and +0.5, then the distribution is approximately symmetric.

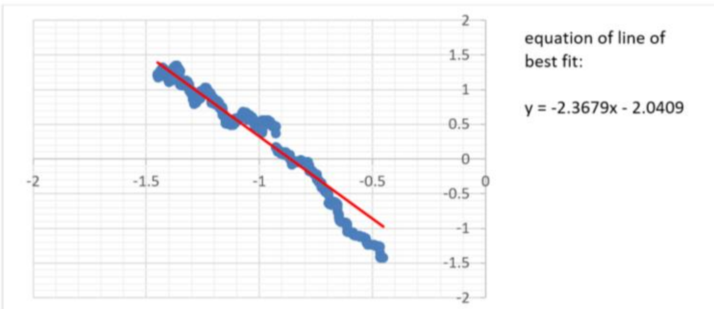
In Excel, the skewness can be calculated by using the function 'SKEW'.

We start by looking at the distributions that have long right tails. The right tail is taken from the mean value as the mean is the x-value of the peak of the distribution. Next to the PDF, we will have the log-log plot. As we are dealing with real life data, it is rare to see a perfect straight line, so to get a good fit for the tail, we consider an interval in the log-log plot where we plot a line of best fit, which can be done in Excel. In case a power law does not exist, the semi-log plot will be plotted after it. The first distribution that we consider is for the 4 am data set.



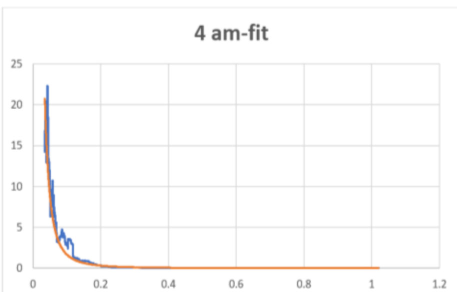


The values between -1.4 and -0.5 seem to show a straight line. We will thus consider a line of best fit in Excel in that interval. However, we bear in mind that the points flatten out after -0.5, so there is not a strong evidence of a power law. In this case, we consider a power law and identify later on (when we analyse the exponent values in detail) if the fit makes sense for the given data set.



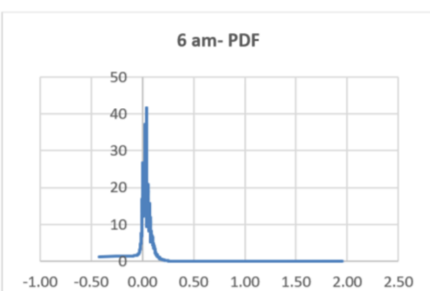
We see from the equation of the line of best fit that the gradient is -2.4 (to 1 decimal place). We thus consider a fit with the exponent -2.4. However, the exponent does not have to be -2.4 but that number gives us a good idea what the approximate value of the power could be.

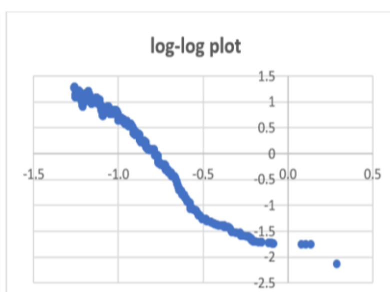
The following is the fit (orange curve) for the right tail of the 4 am distribution:



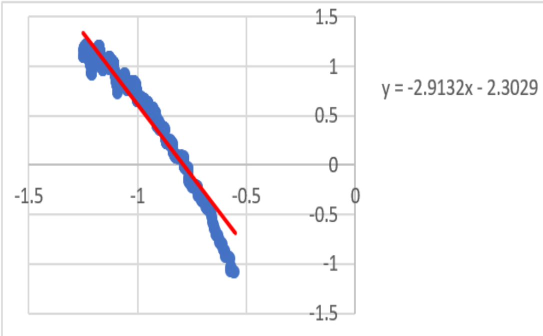
The equation of the power law is:
$y= (1/150)*x^{-2.4}$

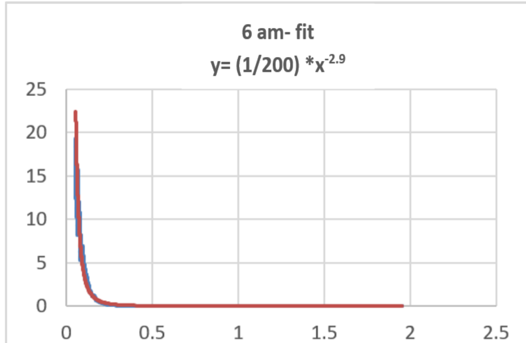
The next distribution is for the 6 am data set.



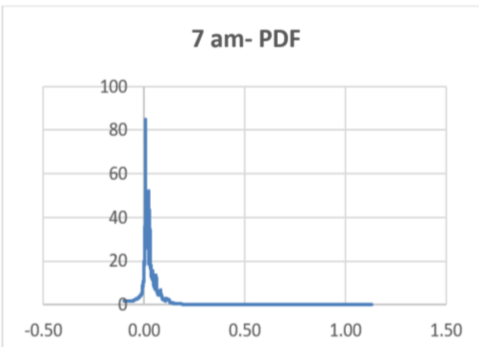


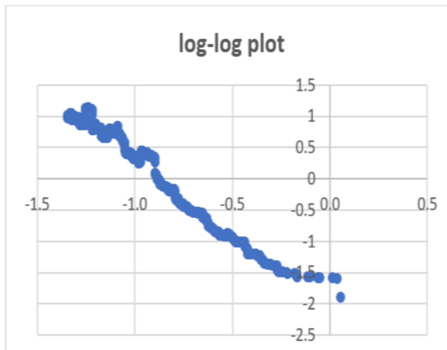
We fit a line of best fit in the interval [-1.2, -0.6]. However, as can be seen from the next plot, the points in that interval do not form a perfect straight line, so the evidence for the existence of a power law is not very strong. As with the 4 am fit, we will consider a power law for now and then analyse later if a power law for this data set makes sense.



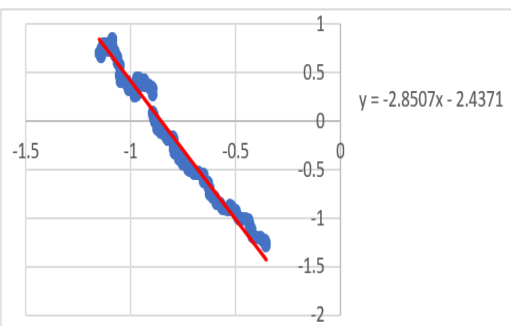


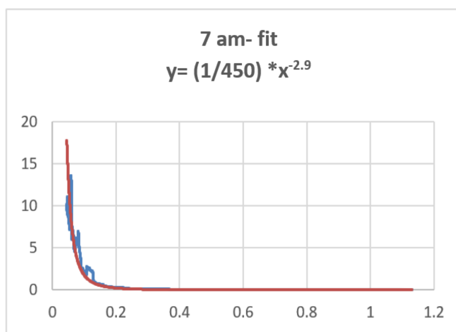
The following is the distribution of the 7 am data set.



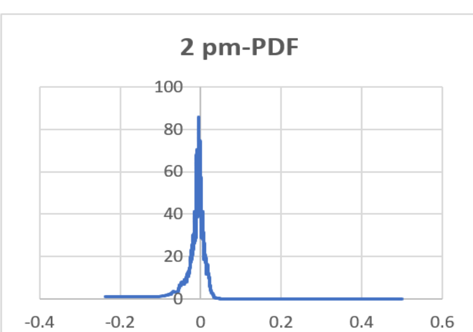


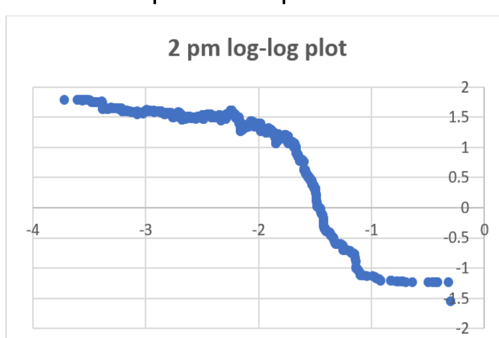
We consider the interval [-1.1, -0.4] for the line of best fit. Here, most of the points lie on the line, so there is strong evidence for a power law.

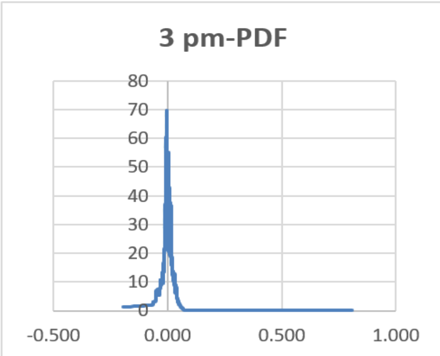


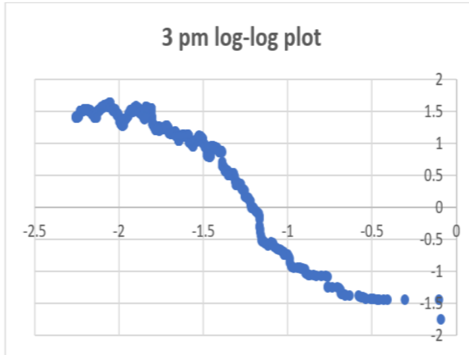


We will now consider the distributions for the 2 pm and 3 pm data set.

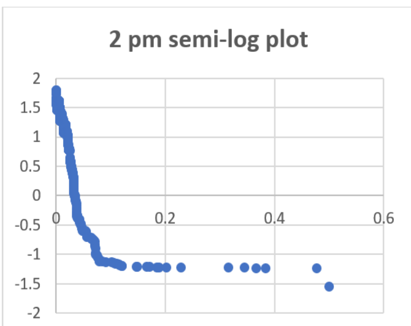


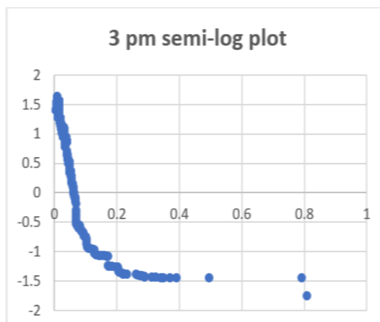






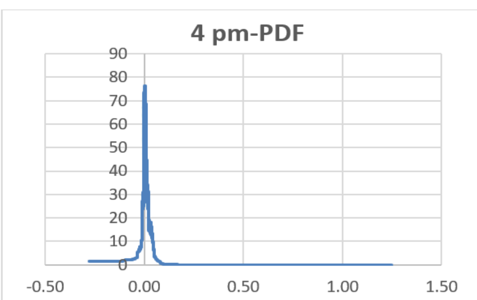
In the log-log plots of 2 pm & 3 pm we do not see a straight line. Therefore, a power law does not exist for these distributions, so we consider a semi-log plot.

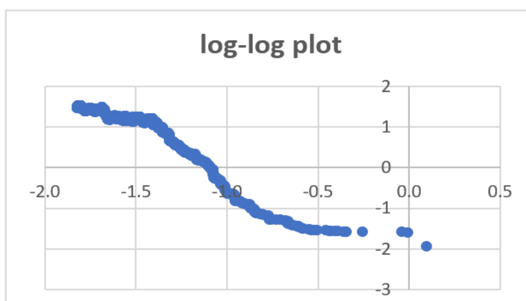




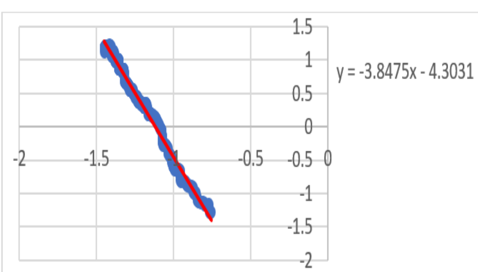
In both semi-log plots, we observe that the points show a decreasing behaviour rather than a straight line. Therefore, an exponential fit does not exist either for the right tails of both distributions.

The following is the distribution of the 4 pm data set.

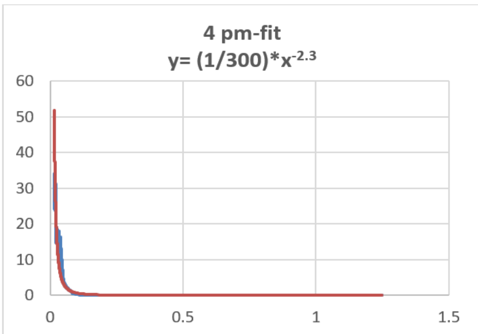




We consider a line of best fit in the interval: [-1.4, -0.8].



The gradient of the line is -3.8 (to 1 decimal place). However, when fitting a curve with an exponent of -3.8, the constant that has to be fit is about 1/200000.This example above shows that, when dealing with real life data, we might find an interval in the log-log plot that seems to be a straight line but we cannot always just take that number to be our exponent as that gradient is the gradient for only the numbers in the interval (and not the whole data set). This is why sometimes one needs to experiment with different numbers and exponents to find the right fit for the data. We can thus consider the line of best fit for the whole data set, which has a gradient of -2.3. This gives a more sensible result. However, later we will see that a power law may not exist for this data set.



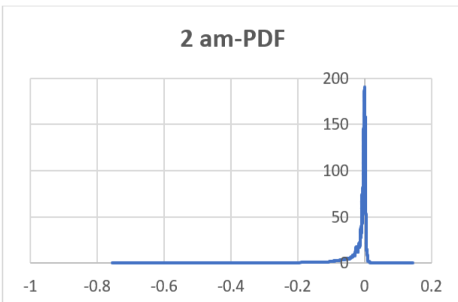
The following table consists of the skewness values of the distributions together with the exponent of the power law fit (if a power law exists).

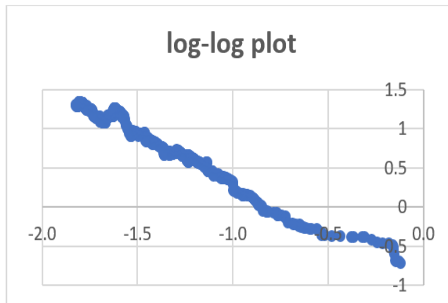
Time	Exponent	Skewness
4 am	-2.4	4.64
6 am	-2.9	6.94
7 am	-2.9	5.37
2 pm	No power law	4.19
3 pm	No power law	6.57
4 pm	-2.3	10.37

We notice that all the distributions above are highly positively skewed. In the time series plots we observed that all these hours of the day had positive extreme events. Thus, positive extreme events lead to positive skewness that in return lead to long tails on the right side of the distributions. Moreover, regarding the exponent, the smaller the exponent is, the faster the tail decays. Thus, the distribution that decays the fastest is for 6 am & 7 am and the distribution that decays the slowest is for 4 pm. It is interesting to see that, even though at 4 pm we have the highest skewness, the right tail decays the slowest when compared to the exponents of 4 am to 7 am and the tails of the 6 am and 7 am distributions have the same exponent but different skewness values, so there does not seem to be a relationship between the skewness value and the exponent. Finally, the only distributions that do not follow a power law or an exponential function are the right tails of 2 pm and 3 pm.

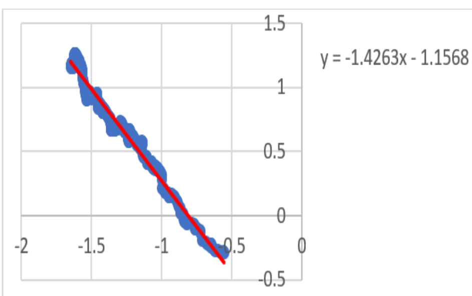
We will now consider the distributions that have long left tails and investigate whether they have power laws. Since we are considering the tails on the left side, the numbers are negative, so we take the absolute value of x, which enables us to take the logarithm of a positive number. To find the absolute value of a number in Excel we use the function ‘ABS’. If a power law exists, then it will have the form: $y = \frac{C}{|x|^n}$,where ‘C’ is a constant and 'n' is the exponent.

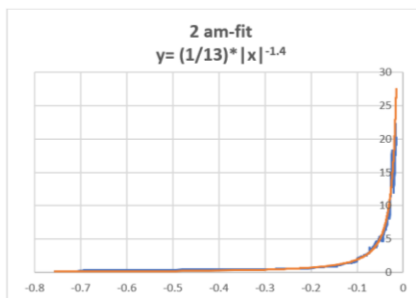
The following is the distribution for the 2 am data set.



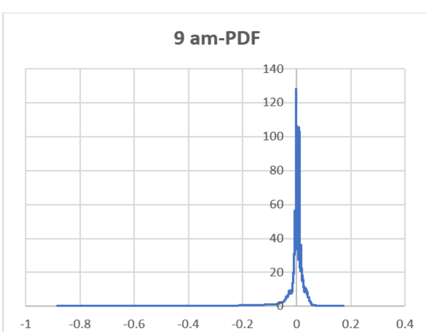


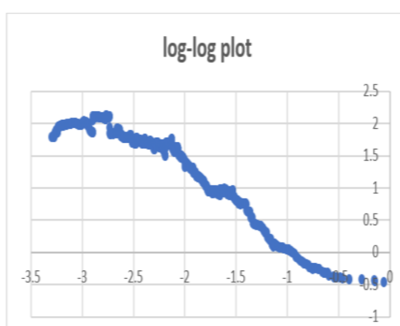
There is strong evidence for a power law. We consider the interval [-1.6, -0.6] for the line of best fit.



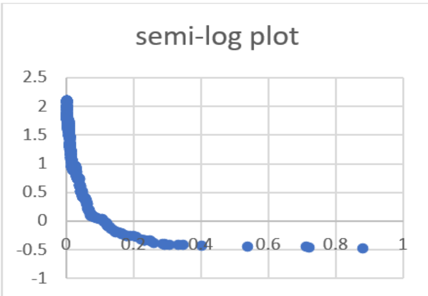


The distribution for the 9 am data set is:



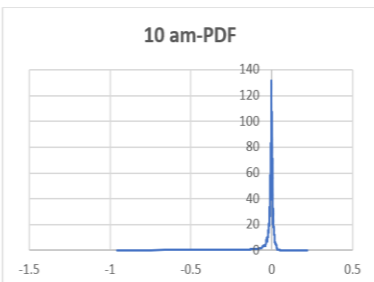


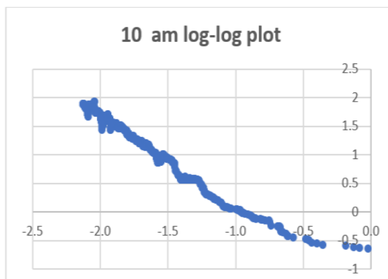
We do not see a straight line in the log-log plot. One can see straight lines for some intervals such as [-2.1 to -1.7] but that interval is too small and thus not a good reflection of the whole data set. Thus, a power law does not exist for the distribution of the 9 am data set and so we consider a semi-log plot.



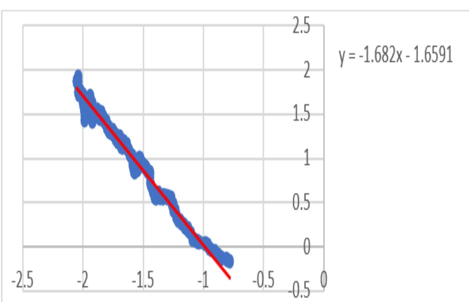
However, we do not have a straight line either in the semi-log plot and thus an exponential fit does not exist for the left tail of the 9 am distribution.

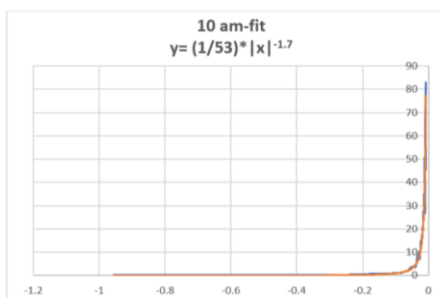
The left tails of 10 am and 11 am have a similar power law.

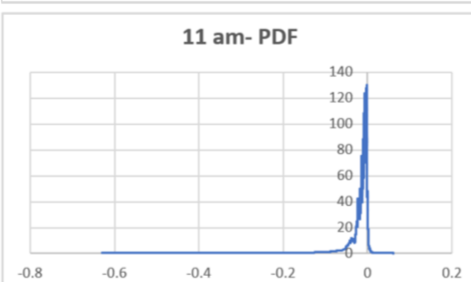


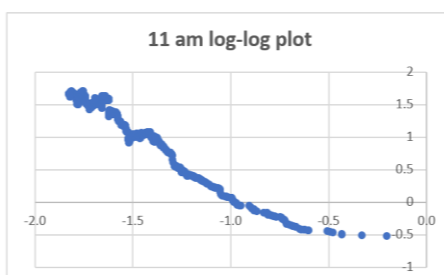


Overall, the points show a straight line. We consider the interval [-2, -0.8] for the fit.

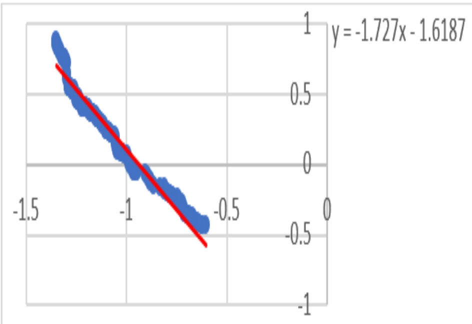


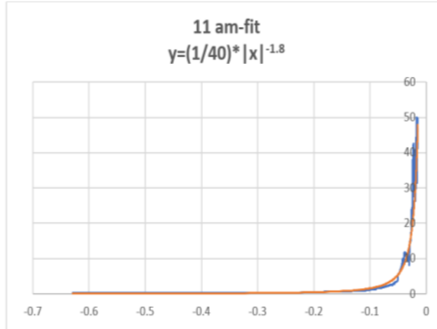




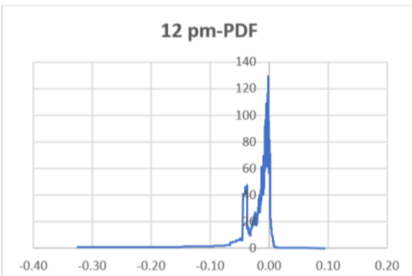


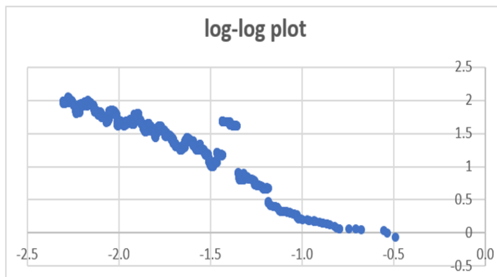
When fitting a line in the interval [-1.3, -0.6], we get a gradient of -1.7. However, we get a better power law fit when using -1.8.



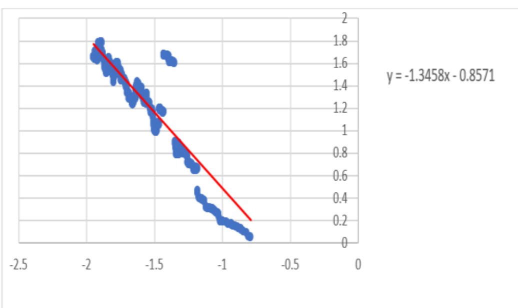


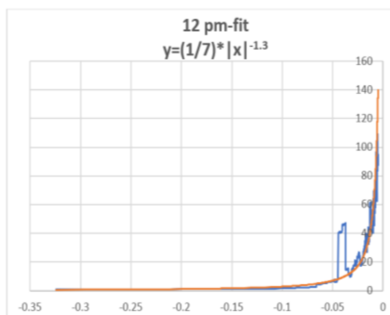
We now consider the distribution for 12 pm.



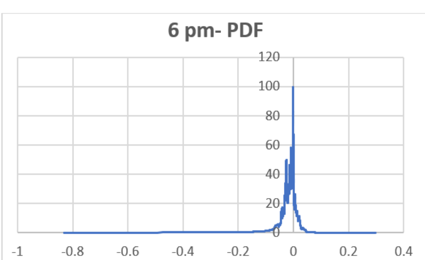


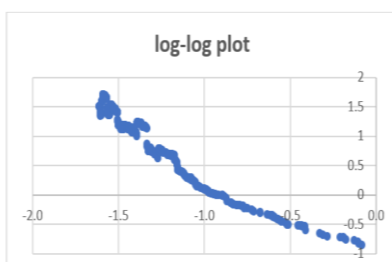
We have some outliers but since these are not many points, we do overall see a straight line. For the line of best fit, we consider the interval: [-1.9, -0.8].



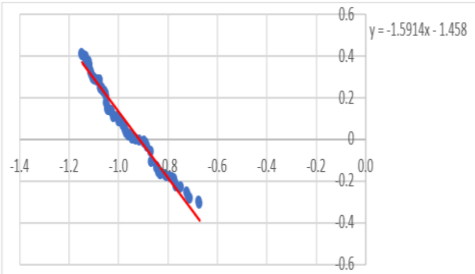


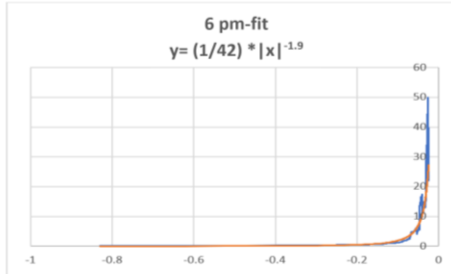
The next distribution is that of 6 pm.



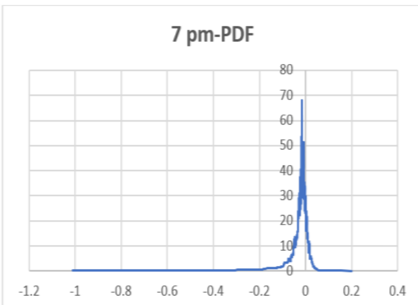


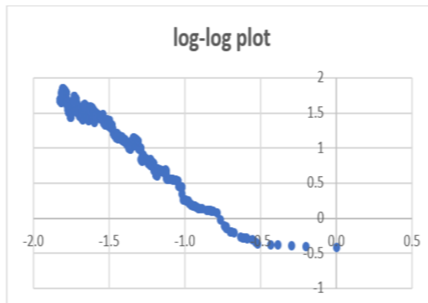
We consider a fit in the interval: [-1.1, -0.7]. We obtain a gradient of -1.6 but we get a better fit by using -1.9 as our exponent.



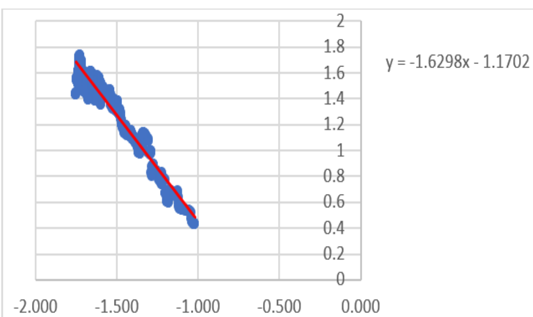


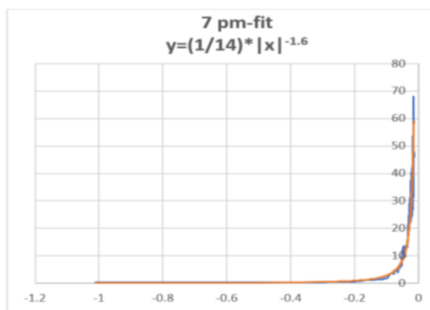
The distribution of 7 pm is the following:



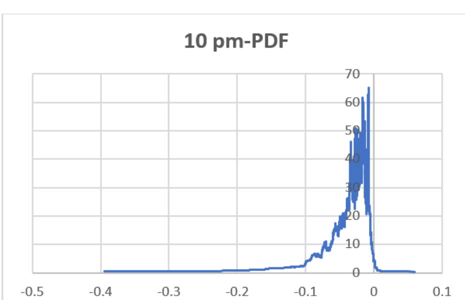


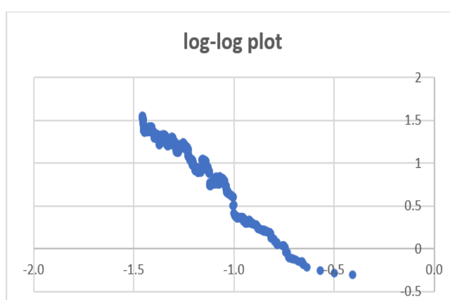
We consider the interval [-1.7, -1] for the line of best fit. However, after -0.6, the points become flatter, but these are only a few data points. Since we do overall see a straight line in the interval [-1.7, -1], there is a strong evidence for the existence of a power law.



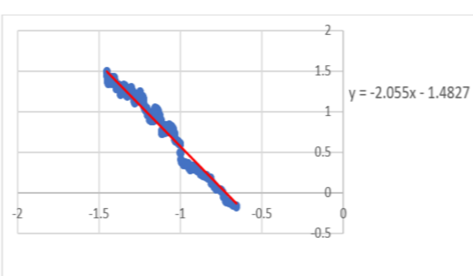


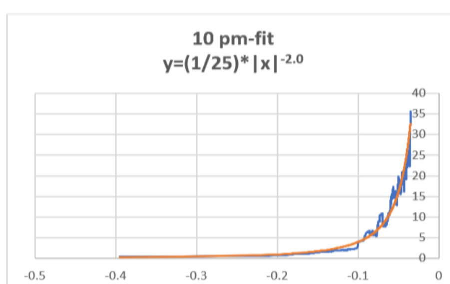
The following is the distribution of 10 pm.



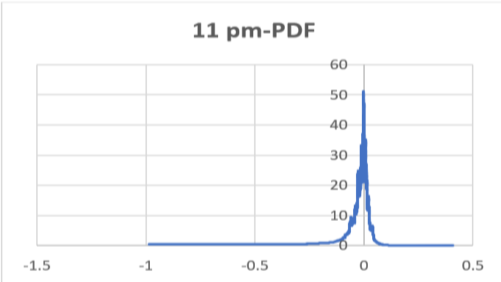


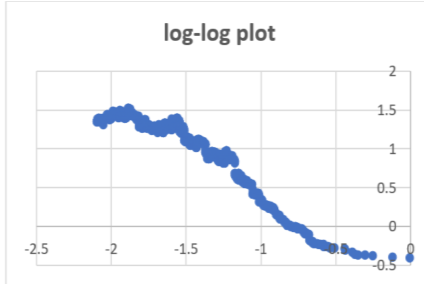
We consider a fit in the interval: [-1.45, -0.7]. We get a gradient of -2.1 (to 1 decimal place) but -2.0 gives a better fit.



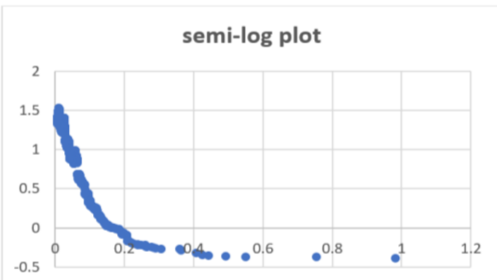


The last 2 distributions that have long left tails are for 11 pm and 12 am.

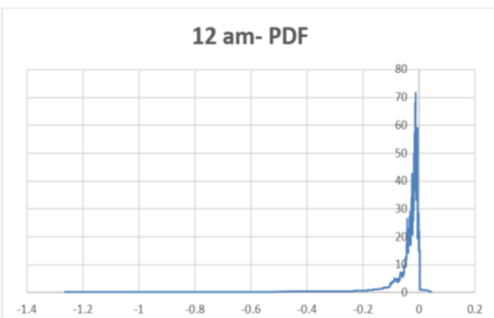


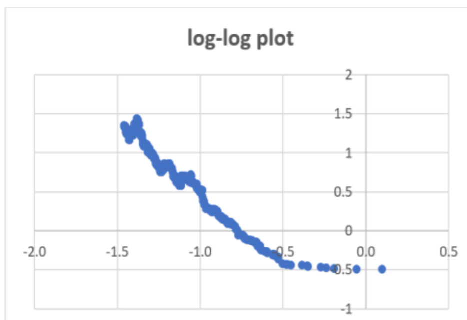


We do not see a straight line in the log-log plot, so we consider a semi-log plot.

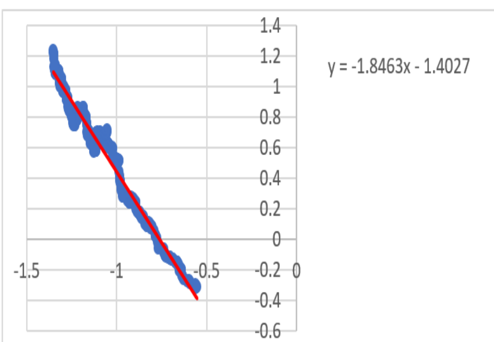


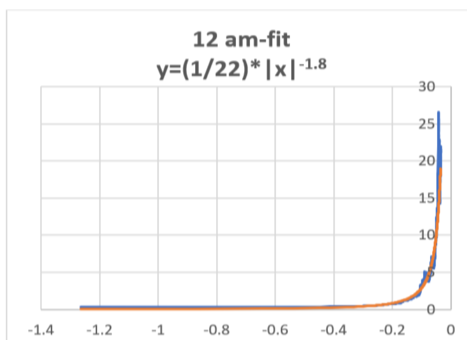
Like the other semi-log plots that we had earlier, we observe an exponential decay rather than a straight line, so we do not have an exponential fit either.





We consider a fit in the interval: [-1.3, -0.6].





The following table consists of the skewness values of the distributions together with the exponent of the power law fit (if a power law exists):

Time	**Exponent**	**Skewness**
2 am	-1.4	-7.92
9 am	No power law	-7.68
10 am	-1.7	-10.69
11 am	-1.8	-7.30
12 pm	-1.3	-4.60
6 pm	-1.9	-7.04
7 pm	-1.6	-8.02
10 pm	-2.0	-2.39
11 pm	No power law	-3.92
12 am	-1.8	-8.10

All the distributions that have left tails are highly negatively skewed and all the time series plots above have negative extreme events. Thus, negative extreme events lead to a negative skewness which in return lead to long tails on the left side of the distributions. An interesting observation is that at 10 pm we have the highest skewness and the smallest exponent and the skewness of the 10 am data set is the smallest, but the exponent is still quite small. So, there does not seem to be a relationship between the value of the skewness and the value of the exponent. Moreover, the smaller the exponent, the faster the power law increases (since we consider the modulus of x). Thus, the power law for the 10 pm distribution increases at the fastest rate whereas the power law for the 12 pm distribution grows at the slowest rate. Finally, the only distributions that do not follow a power law or an exponential function are the left tails of 9 am and 11 pm.

Existence of Moments

We will now take a closer look at the values of the exponents. The exponent value characterises the importance or strength of the extreme events that we observe in our time series plots. If the exponent is smaller than -2, then that means that the first moment (which is the expectation) still exists since $\int_{-\infty}^{\infty} xp(x)dx$ converges when x is large and so the extreme events do not have much impact on the data set. If the exponent is between -1 and -2, then the first moment does not exist anymore, meaning that the expectation diverges and cannot be used anymore to characterise the event. Lastly, if the exponent is close to -1 then this is very extreme as then even the normalisation of the distribution starts to fail because $\int_{-\infty}^{\infty} xp(x)dx$ would diverge. So, in the last case, the extreme events have the strongest impact on the data set.

The following table shows the exponents for the distributions that have long right or left tails along with their mean and variance values. We will analyse on what hours of the day the extreme events have the strongest/weakest impact on the electricity price.

Time	Exponent	Right/Left tail	Mean	Variance
2 am	-1.4	left	-0.01788	0.00333
4 am	-2.4	right	0.04123	0.00492
6 am	-2.9	right	0.06142	0.00935
7 am	-2.9	right	0.04965	0.00766
9 am	No power law		0.00218	0.00198
10 am	-1.7	left	-0.00782	0.00186
11 am	-1.8	left	-0.01743	0.00102
12 pm	-1.3	left	-0.01329	0.00047
2 pm	No power law		-0.00547	0.00105
3 pm	No power law		0.00586	0.00186
4 pm	-2.3	right	0.01827	0.00273
6 pm	-1.9	left	-0.01634	0.00321
7 pm	-1.6	left	-0.01630	0.00167
10 pm	-2.0	left	-0.03762	0.00109
11 pm	No power law		-0.01342	0.00321
12 am	-1.8	left	-0.03579	0.00293

For the following distributions the expectation exists: 4 am, 6 am 7 am and 4 pm. Thus, for these hours of the day, the extreme events do not have a strong impact on the data set, so the mean and the variance of these data sets are a good representation of their overall behaviour. This is certainly expected for the 4 am and 6 am distributions as most people sleep during that time and we would not expect the extreme events to have a strong impact on the electricity price. However, at 4 pm the price fluctuations are higher than other hours of the day (as the variance is quite high), so we would expect that the extreme events (which lead to a higher variance) would have a greater impact on the electricity price, and so the exponent value seems surprising. The reason for this may be that in the log-log plot, we did not see a perfect straight line. A straight line is only visible for the interval [-1.4, -0.8] and thus the power law that we used for the fit may not have been a good representation of the whole data set.

For the following distributions the first moment does not exist: 2 am, 10 am, 11 am, 12 pm, 6 pm, 7 pm, 10 pm and 12 am. So, the extreme events have a strong impact on these hours of the day and thus, the mean and the variance are not a good representation for these data sets. This result is expected for 10 am to 7 pm as we would expect the price to fluctuate during these hours and thus it makes sense that the extreme events have a strong impact on the price. However, the result is quite surprising for 2 am and 12 am as not much electricity is used because most households sleep during that time. We would thus not expect that the extreme events have such a strong impact on the price. One plausible reason may be that their variances are quite high (0.003) and so the mean is indeed not a good representation of the whole data set. However, in the 2 am log-log plot we do see a straight line overall, so there is a strong evidence for the existence of a power law for the left tail of the 2 am distribution, but in the 12 am log-log plot we only see a straight line for the interval [-1.3, -0.6]. However, after -0.6 the points flatten out and thus the fit for the 12 am distribution may not be a good representation of the whole data set (we had a similar situation for the 4 pm distribution). Furthermore, the extreme events have the strongest impact at 12 pm as the exponent is closest to -1. An interesting observation is that the exponents of the left-tailed distributions are all between -1 and -2 whereas the exponents for the right-tailed distributions are all below -2. Thus, negative extreme events have a stronger impact on the price than positive extreme events. We also notice that the mean and the variance are very small for the distributions that do not have a power law, i.e. for 9 am, 2 pm, and 3 pm (except for 11 pm, where the variance is quite high).

Geometric Brownian Motion

We want to observe whether the whole data set or some hours of the day follow a process called Geometric Brownian Motion (GBM). Before doing this, let us introduce its formal definition.
We start by considering the stochastic process called Brownian Motion (also known as the Wiener Process) that we define to be $B(t)$. The following is the formal definition of Brownian Motion.

<u>Definition I (Phillips, 2017)</u>
Brownian motion satisfies the following for t ≥ 0:

1. $B(0) = 0$
2. The sample paths of $B(t)$ are continuous
3. The increments of $B(t)$ are independent. This means that for any set of times $0 \leq t_1 < t_2 < \cdots < t_n$, the random variables $B(t_2) - B(t_1)$, $B(t_3) - B(t_2)$,..., $B(t_n) - B(t_{n-1})$ are independent
4. For any $0 \leq s < t$, the increments are normally distributed, such that $B(t) - B(s) \sim N(0, t - s)$
   "These conditions are both necessary and sufficient to define the Wiener Process" (Phillips, 2017).

However, when modelling stock prices, Brownian Motion is not a good model as the stochastic process assumes that the probability of a stock price going up is the same probability as the stock price going down, whereas in reality the stock price tends to move up over time (on average) (Phillips, 2017). An improved version of Brownian Motion is Brownian Motion with drift.

Definition II (Phillips, 2017)

Brownian Motion with drift is defined to be the following process: $X(t) = X(0) + mt + \sigma B(t)$, where $X(0)$ is the initial value of the process, $'m'$ is the drift per unit time, σ is the volatility and $B(t)$ is the Brownian motion process.

However, the problem with Brownian motion with drift is that the process can take negative values (Phillips, 2017) . An improved model is thus Geometric Brownian Motion (GBM).

Definition III (Phillips, 2017)

Geometric Brownian Motion is defined as the following process:

$Y(t) = Y(0)\exp[mt + \sigma B(t)]$. So, the logarithm of $Y(t)$ is Brownian Motion with drift m, $\log(Y(0))$ as the initial value and volatility σ.

This is useful as we cannot take negative values (if we assume that $Y(0) > 0$).

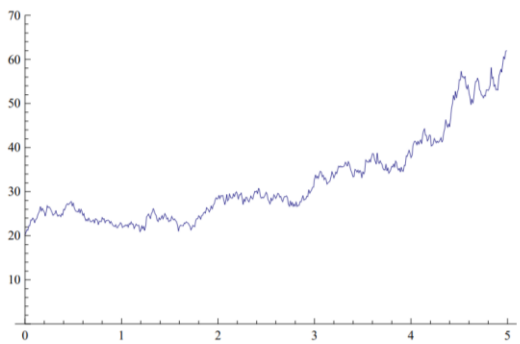
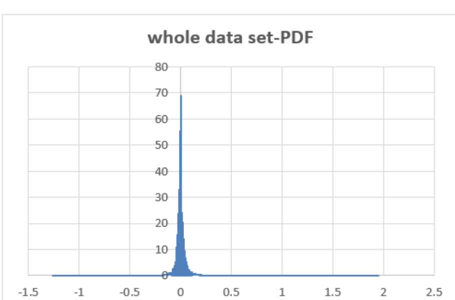
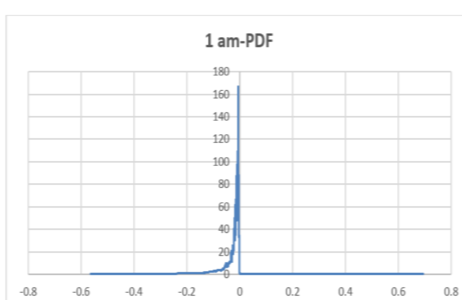
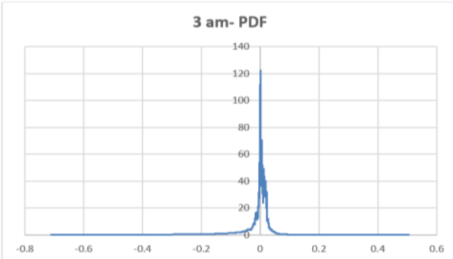


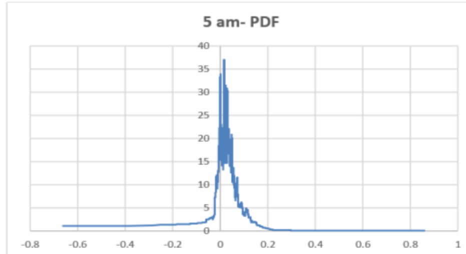
Figure above taken from: Phillips, M., 2017, *A sample path of Geometric Brownian Motion,* lecture notes, Foundations of Mathematical Modelling in Finance, MTH771P, Queen Mary University of London.

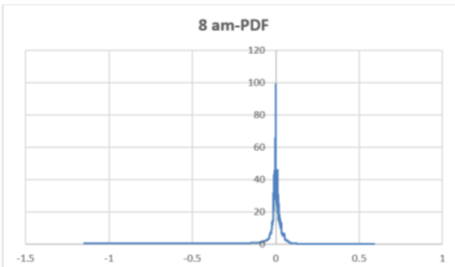
The above figure is a sample path of Geometric Brownian Motion. We saw a similar behaviour (i.e. an upward trend) when we plotted the whole data set of the prices. In a GBM process we expect a normal distribution with no extreme events. So, the distributions that we looked at before with long right or left tails were all highly skewed (and not normally distributed) and extreme events were strong in all the distributions with long left tails. Thus, those distributions certainly do not represent a GBM process. We will now consider the distributions with long tails on both sides and see whether they follow a GBM process.

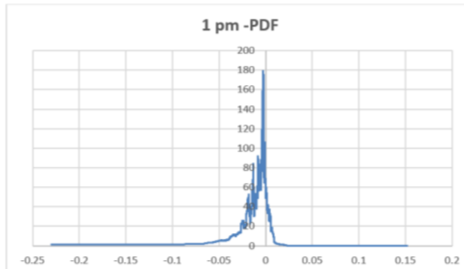


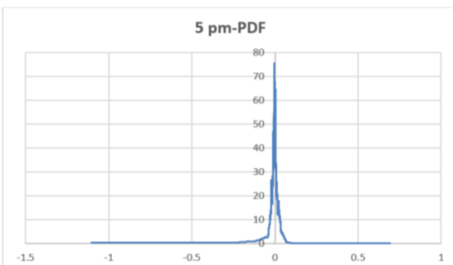


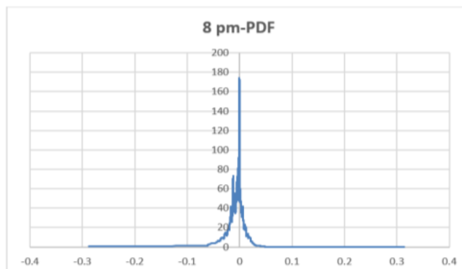


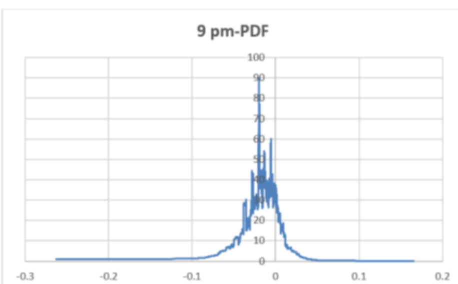












The following are the skewness values for the distributions above:

Time	**Skewness**
whole data set	1.88
1 am	2.96
3 am	-1.83
5 am	2.28
8 am	-3.76
1 pm	-2.03
5 pm	-2.94
8 pm	0.62
9 pm	-0.20

As can be seen from the table, the distribution of the whole data set is highly positively skewed as well as the distributions of 1 am and 5 am. The distributions of 3 am, 8 am, 1 pm & 5 pm are highly negatively skewed. Thus, for these highly skewed distributions we do not have a GBM process as we would need the skewness to be zero. However, at 8 pm the distribution is moderately skewed, so the skewness at this hour is significantly lower when compared to the hours before that. The distribution that shows a strong evidence for a GBM process is for 9 pm as a skewness of -0.20 means that the distribution is approximately symmetric. This supports a GBM process as we observe somewhat a normal distribution whereby the skewness is very close to zero.

4.0 Conclusion

We analysed a data set that is given from 1999 to 2007. A study of a longitudinal period enables us to get a good insight about how the returns behave for each hour of the day and how extreme events such as weather changes affect the prices. When analysing the mean and the variance we found some results that we would expect, such as low prices at midnight. We also found counter-intuitive results such as a negative mean at 5 pm and 6 pm. To get a better insight about the behaviour of the returns, we then looked at the distributions and observed on what hours of the day the extreme events had the strongest impact and it turned out that, surprisingly, at 2 am the extreme events have a strong impact on the electricity price. Thus, when dealing with real life data, we get results that one may not expect and may not be able to explain when solely looking at a given data set.

In conclusion it can be argued that the data set for the whole period, as well as for the periods conditioned on the hour, we have a non-stationary time series. There are ways to transform a non-stationary time series to a stationary time series by differencing the data set and then fitting a model to the data set. However, when applying a model to a data set that requires stationarity to a non-stationary data set, it is likely that we get inaccurate estimators for the fitted model that do not follow required mathematical properties such as asymptotic normality or consistency and thus we would have a poor forecast of the prices. Moreover, none of the data sets can be modelled by Geometric Brownian Motion except the time series for the 9 pm data set, whereby we had a skewness close to zero.

5.0 References

Bergman, L., 2001. Regulation and Competition On the Nordic Power Market. In *Stockholm School of Economics, Stockholm, Sweden. Technical paper presented at 18th World Energy Congress,* pp. 1-3. [Online] Available at: https://pdfs.semanticscholar.org/1864/9cffe78f204586e786d29d5204fabc9502dc.pdf [Accessed 08/07/2018]

Bergman, L., 2002. The Nordic electricity market-continued success or emerging problems?. *Swedish Economic Policy Review, 9(2)*, pp. 51-58. [Online] Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.476.1518&rep=rep1&type=pdf [Accessed 02/07/2018]

Bergman, L., 2003. European Electricity Market Integration: The Nordic Experiences. In *Research Symposium European Electricity Markets*, The Hague. [Online] Available at: https://www.ecn.nl/fileadmin/ecn/units/bs/Symp_Electricity-markets/c1_2-paper.pdf [Accessed 02/07/2018]

Boyle, P.E., 2017/2018. *Lecture 1: Derivatives Basics, Futures/Forwards, Lecture 2: Introduction to Options, Lecture 7: Interest Rates & Swaps,* lecture notes, Financial Derivatives, ECOM026, Queen Mary University of London. [Accessed 28/07/2018]

Erbach, G. , 2016. Understanding electricity markets in the EU. *European Parliamentary Research Service*, PE 593.519, pp. 1-5. [online] Available at: http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/593519/EPRS_BRI(2016)593519_EN.pdf [Accessed 21/07/2018]

Erzgraeber, H., Strozzi, F., Zaldivar, J.M., Touchette, H., Gutierrez, E. and Arrowsmith, D.K., 2008. Time series analysis and long range correlations of Nordic spot electricity market data. *Elsevier*, PACS: 05.45.Tp, 05.40.Jc, pp. 1-5. [Online] Available at: https://pdfs.semanticscholar.org/2b01/bde7404c4fd8fbcb762130cdf757d921ee6b.pdf [Accessed 07/06/2018]

Exchange Utility Ltd Company, 2018. *What factors affect business electricity prices?*. [Online] Available at: https://www.exchangeutility.co.uk/news/business-electricity-prices-what-affects-your-bill-2/ [Accessed 17/07/2018]

Giraitis, L., 2017. *Lecture 1: Time Series and their characteristics,* lecture notes, Time Series Analysis, ECOM014, Queen Mary University of London. [Accessed 20/07/2018]

GoodData Corporation, 2018. *Normality Testing-Skewness and Kurtosis*. [Online] Available at: https://help.gooddata.com/display/doc/Normality+Testing+-+Skewness+and+Kurtosis [Accessed 01/08/2018]

Investopedia, 2018. *Types of Financial Markets and Their Roles* [Online] Available at: https://www.investopedia.com/walkthrough/corporate-finance/1/financial-markets.aspx [Accessed 22/08/2018]

Laerd Statistics, Lund Research Ltd, 2018. *Measures of Central Tendency.* [Online] Available at: https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median.php [Accessed 15/07/2018]

Manikandan, S., 2011. Measures of Dispersion. *Journal of Pharmacology & Pharmacotherapeutics*, 2(4), 315-316. [Online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3198538/ [Accessed 09/07/2018]

Musaddique, S., 2017. 'Cost of Brexit: The impact on business and the economy in 2017 and beyond', *The Independent*, 26 December. [Online] Available at: https://www.independent.co.uk/news/business/news/brexit-economy-sterling-currency-investment-cost-impact-business-financial-banks-insurance-retail-a7695486.html [Accessed 14/08/2018]

Phillips, M., 2017/2018. *Topic 7: Continuous-time stochastic processes,* lecture notes, Foundations of Mathematical Modelling in Finance, MTH771P, Queen Mary University of London. [Accessed 20/07/2018]

Tsay, R.S., 2010. Analysis of Financial Time Series, Wiley, Hoboken. [Online] Available at: ProQuest Ebook Central, pp. 1 & 30. https://ebookcentral-proquest-com.ezproxy.library.qmul.ac.uk/lib/gmul-ebooks/detail.action?docID=565117 [Accessed 10/07/2018]

U.S. Energy Information Administration (EIA), 2018. *Electricity Explained, How Electricity is Delivered to Consumers*. [Online] Available at: https://www.eia.gov/energyexplained/index.php?page=electricity_delivery [Accessed 21/07/2018]