Data Analytics MSc Dissertation MTH797P, 2019/20

Statistical Analysis of Extreme Events using Time Series Financial Data:

The Nord Pool Electricity Prices

Leticia Perez Amor, ID 190663449

Supervisor: Dr Wolfram Just

A thesis presented for the degree of
Master of Science in *Data Analytics*

School of Mathematical Sciences
Queen Mary University of London

Declaration of original work

This declaration is made on September 9, 2020.

Student's Declaration: I Leticia Perez Amor 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, **and**
2. using quotation marks "...", **and**
3. explicitly mentioning the source in the text.

Abstract

This paper presents the framework by which electricity markets, particularly the Nord Pool, operates and elucidates how it was created and how it currently works. It uses R and Excel to produce a statistical analysis of electricity prices from the Nord Pool (ranging from January 1999 to January 2007). This analysis has a strong focus on extreme events (ie. the situation in which prices exceed the normal range of daily fluctuations, defined to be at 60EUR/MWh). The paper also takes into account the complex seasonality present in the data to model it using different methods, and produces forecasts. Finally, residuals are analysed and weaknesses of the analysis are highlighted.

Contents

Chapter 1

Introduction

The International Energy Agency (International Energy Agency 2013) emphasised back in 2013 how vital energy has become to our society, playing part in the economy, the environment and how we live. Today 90 per cent of global electricity demand belongs to industry and buildings. In fact, according to the Sustainable Development Scenario, electricity is expected to experience a rise in consumption by 2040 - primarily driven by the spread of electric automobiles (International Energy Agency 2019). Due to its prominence, organisms such as the European Union guarantee the supply of energy to citizens by creating a *"common energy market ruling"* (European Commission 2019). The premise behind this approach is that, by enabling the free flow of electricity to those countries or regions where it is needed the most, Europeans will benefit from trade across frontiers and competition.

The existence of this common market, only commenced relatively recently at European level, and was inspired by the Nordic pooling system. The main basis of the European Union Treaty (free movement of goods, services, capital and people) did not seem to apply to the energy sector. In fact, the European Union's electricity sector in the 1980s was constructed as a regulated monopoly (KU Leuven Energy Institute 2015). Each country had one or several companies that maintained the generation and supply of electricity. The turning point occurred after the Maastricht Treaty was signed in 1992, when the free movement of goods, including energy, was satisfied. Jan Roönnback, Director of Market Coupling of the Nord Pool, stated in August 2019 that the Nord

Pool *"can lay claim to having invented the modern power exchange"*, since they were the first to arrange cross-border power trading (Rönnback 2019).

1.1 Contents of the thesis

The history of the Nord Pool is only recent. As it is explained in *Chapter 2*, the birth of this market started in 1991, when the electricity trading market was deregulated in Norway, and started growing in 1996 with the joint Norwegian-Swedish power exchange foundation. As of today, there are over 300 participants from all over the world trading in this market that is shifting towards a low carbon economy. In pooled markets such as the Nord Pool, energy flows to those areas where it is needed the most, enabling users to benefit from competition. Nonetheless, trading electricity presents a shortcoming: supply has to equal consumption on an immediate basis, as it cannot be stored. Additionally, *Chapter 2* presents the different electricity markets that exist in order to achieve the flexibility needed: day-ahead markets and intra-day markets, and introduces organisms such as the Balance Responsible Parties (BRPs) and the Transmission System Operators (TSOs), which ensure the balance between supply and consumption.

Despite the existence of these mechanisms to ensure balance, prices can vary greatly due to political decisions, meteorological conditions or disturbances occurring in power plants. The paper uses data of electricity prices from the Nord Pool, ranging from January 1999 to January 2007. In *Chapter 3*, exceedance prices are defined and analysed in R in an hourly, weekly and monthly basis. Most of these exceedances are explained by daily rhythms and industrial activity. The paper then moves on to present summary statistics (mean, variance, etc.) of the hourly logarithmic returns. Skewness and kurtosis are both numerically and graphically tested, and an analysis of the cumulative distribution function (CDF) and the probability density function (PDF) is shown. Finally, the paper defines stationarity and how to test it formally, using the Augmented Dickey-Fuller test. *Chapter 3* ends up by defining conditional heteroskedasticity and graphically proves that the data do present it.

Chapter 4 models the data in order to produce accurate forecasts. Seasonal patterns are graphically analysed in the first place, by plotting the average prices and the coefficient of variation across months, days of the week and hours of the day, and also by plotting autocorrelation functions. It is concluded that the data present yearly, monthly, weekly and daily seasonality. Once this is done, the paper moves on to theoretically explaining the models that will be used: ARIMA, ETS (both using LOESS regression), TBATS and seasonal Naïve method. The Naïve method is included as a benchmark model. Once these models are produced, the accuracy of each is measured and one model is chosen (the aim is that the chosen one performs better than the Naïve model). Finally, residuals are checked in case there is correlation, and criticisms and future improvements of the analysis are presented.

1.2 Motivation for this work

Price spikes are a recurrent issue in electricity markets, and it could also be extrapolated more widely to financial markets in general. Academics have been trying to find a good approach to model and predict them. It is valuable to understand how prices behave, being the variance notably germane. Future prices are unclear, and they should be defined using a probability distribution. This entails that statistical methods are the legitimate approach to study them. Once insights about price behaviour are available, they can be used to diminish risk or make more accurate decisions about the future.

Nowadays, forecasting is of vital importance across many sectors, being one of them the electricity sector. The motivation for this work is to study and accurately forecast electricity prices of the Nord Pool using R and Excel. It is important to have access to this information, as power plants can more accurately plan their outputs and engage in successful contracts, and consumers can plan accordingly. Moreover, it could also incentivise governments to subsidy prices at certain times - particularly for low-income families or certain industries - now that electricity is such a vital good.

Chapter 2

The Nord Pool

The Nord Pool is said to "*have invented the modern power exchange*". This section presents its history and current aims, followed by information about the Nordic nations where the Nord Pool operates (ie. Norway, Sweden, Finland and Denmark). It then moves on to describe how electricity markets, particularly the Nord Pool works and how prices are agreed on.

2.1 History

Until 1991, Nordic electricity markets were controlled by vertically integrated publicly owned power companies. This meant that prices and investments were informally regulated. It was in 1991 when the electricity trading market was deregulated in Norway, and only two years later Statnett Market AS established itself as an independent company. Deregulation refers to the fact that the state does not run the power market, and competition is adopted. Statnett Market AS generated a value of NOK 1.55 billion in the first year of operations, which would correspond to over GBP 123 million today (Nord Pool 2019).

Despite the Norwegian financial electricity market being developed without formal regulation, the Securities Trading Act stated that the exchange "*shall be carried out with due consideration to principles of efficiency, neutrality and equal treatment of trading participants, as well as ensuring that the market as such is transparent*". This approach was taken since it would build a more effective market that would further secure the supply (as power capacity can be more efficient when covering large regions). Along these lines, in 1996 a joint Norwegian-Swedish power exchange was founded and renamed as Nord Pool ASA. Each half of the Nord Pool ASA was owned by the Norwegian and Swedish TSOs. By 1998 Finland joined, and in the year 2000 Denmark joined.

In 2001, Nord Pool ASA de-merged into Nord Pool Spot AS that operated two markets: the day-ahead market (ie. Elspot market) and the hour-ahead balancing market (eg. Elbas market). The Elspot market is the largest and most important physical market of the Nord Pool that can trade up to 36 hours in advance. The Elbas market was born because of the need to create a market for trade in contracts as near as possible to the delivery hour. It trades up to 1 hour before the delivery hours, and is where Swedish and Finnish members originally traded, joined later on by Danish and Norwegian market participants. In 2005, the Kontek bidding area was opened in Germany (previously referred as Vattenfall Europe Transmission control area), which is one of the 4 transmission system operators for electricity in the country. This provided the Nord Pool with access to the 50Hrtz Transmission GmbH. The Elbas market (an intra-day market) was licensed in the Netherlands and Belgium in 2011, and in Latvia and Lithuania in 2013. And finally, in 2014, the Nord Pool took exclusive control of the UK market. Ever since, it has continued expanding and in 2019 it launched day-ahead trading for Central and Western Europe, being the first offerer of day-ahead markets across Europe (Nord Pool n.d.).

Moreover, not only Nordic countries take part in the market: there are more than 300 participants (from other European countries and the US) that trade in Nord Pool markets. What originally started as a simple market, whose aim was to develop an efficient exchange to trade power, ended up causing the emergence of a financial market that set the basis for the modern power exchange. Today, the Nord Pool's main aim continues to be to maximise social welfare (Rönnback 2019). But it also recognises the existence of challenges such as climate change, and regulatory and security factors. It is committed to maintain the security in supply while moving towards a low carbon economy. Nonetheless, it must be borne in mind that the introduction of renewable energies presents challenges such as higher risk of blackouts and high initial costs.

2.2 Nordic countries' characteristics

Thanks to the wide governmental support in terms of innovation, Nordic nations present a strong position compared to other countries, since they generate more than 30 per cent of the global production of wind energy technology. In the field of biomass-based generation of power, Nordic countries have a proportion of almost 30 per cent of all the biomass-based heat and power worldwide. Energy innovation is such a crucial activity, that according to Nordregio, the International Research Centre for Regional Development and Planning, it accounts for around 6 per cent of revenues and employment in the area, and the export of energy translates into around 5 to 9 per cent of industrial exports (Nordregio 2019).

Not only Nordic countries generate electricity from renewable resources at a level 4 times higher than that of OECD countries, but also present high levels of consumption of electricity from renewable resources. The share of renewable sources in final energy consumption across these countries in plotted in *Table 2.1.* As it can be observed, Norway and Sweden are in the lead with more than half of electricity consumption coming from renewable sources, followed by Finland and Denmark.

Country	**Share of renewables in final consumption**
Norway	59.5 per cent
Sweden	51.4 per cent
Denmark	33.1 per cent
Finland	42 per cent

Table 2.1: Share of renewable sources in final energy consumption.
Source: IEA, 2016 (International Energy Agency 2016)

The four Nordic nations are small in terms of their population, but the electricity consumption per capita is one of the highest in Europe, especially in

Norway and Finland as can be seen in *Table 2.2*. Metropolitan regions, as expected, present particularly high consumption of electricity. It is worth highlighting the consumption levels of cities such as Stockholm, Västra Götaland and Oslo. These cities present as well a higher share of service sectors. And of course, households also account for a big proportion of consumption, particularly in Norway and Sweden. It must be noted that despite presenting one of the highest consumptions of energy per capita in the world, greenhouse emissions are moderate due to their little reliance on fossil fuels.

Country	**Electricity consumption per capita**
Norway	24.1 MWh/capita
Sweden	13.1 MWh/capita
Denmark	5.8 MWh/capita
Finland	15.8 MWh/capita

Table 2.2: Electricity consumption per capita.
Source: IEA, 2018 (International Energy Agency 2018)

The high consumption of energy is explained apart from the cold climate these countries share, by the varied industrial activities across them. Norway has electricity-intensive industries according to Nordic Energy Research, and the use of electricity is more widespread in the aim to heat spaces and water (Nordic Energy Research 2012). Sweden and Finland count on heat intensive industries and Denmark has a modest energy-intensive industry.

Source	**Norway**	**Sweden**	**Denmark**	**Finland**
Industry	3,956 ktoe	4,380 ktoe	737 ktoe	3,315 ktoe
Residential	3,414 ktoe	3,881 ktoe	848 ktoe	1,936 ktoe
Commercial/Public serv.	2,151 ktoe	2,370 ktoe	920 ktoe	1,518 ktoe
Fishing	18 ktoe	0 ktoe	0 ktoe	0 ktoe
Transport	79 ktoe	213 ktoe	35 ktoe	66 ktoe
Agriculture/Forestry	160 ktoe	98 ktoe	151 ktoe	133 ktoe

Table 2.3: Electricity consumption by sector.
Source: IEA, 2018 (International Energy Agency 2018)

On top of this, the distribution of the population is sparse, which is translated into a higher demand for individual transport and the corresponding rise in energy demand (Nordregio 2019). As it can be observed in *Table 2.3*, industrial consumption is the highest across these Nordic countries. Among these industries, the metallurgic industry in Norway and Sweden stands out. And in fact, these two countries present the highest industrial electricity consumption. It is also worth highlighting other energy-intensive industries such as the pulp and paper sectors, particularly relevant in Finland and Sweden. Finally, it is interesting to look at the amount of energy consumed by Norway in fishing - as Norway is the biggest fishing country in Europe.

Another interesting characteristic is that the power generation system across these countries is very varied, as can be seen in *Table 2.4*.

Source	**Norway**	**Sweden**	**Denmark**	**Finland**	**Total**
Hydro	11,986	5,294	1,348	1,143	19,771
Oil	9,245	9,898	6,190	8,144	33,477
Natural gas	6,093	1,000	2,657	2,179	11,929
Biofuels, Waste	1,834	12,203	4,728	10,158	28,923
Coal	823	2,201	1,571	4,184	8,779
Wind, Solar...	333	1,475	1,348	520	3,676
Nuclear	0	17,145	0	5,939	23,084

Table 2.4: Total primary energy supply by source in 2018 (Units in ktoe).
Source: IEA, 2018 (International Energy Agency 2018)

Hydropower is the main power in Norway, especially in the south of the country, thanks to the geographical suitability to generate this kind of energy. Hydropower controls the north of Sweden as well, but there is also a high presence of nuclear plants. In fact, most urban regions in the south are supplied by them. Denmark mainly utilises thermoelectric generation and power generation based on fossil fuels (Bergman 2003a). Finland also uses power generation based on fossil fuels, particularly in the south, that is supplied with nuclear and thermoelectric energy, the latter being generated from natural gas and biomass. The northern part of Finland presents more hydropower plants, and there is a high use of biomass.

2.3 How does the Nord Pool work?

The financial electricity market refers to "*trading in electricity-related commercial paper and derivatives for which electricity is the underlying commodity*" (Finansinspektionen 2005). An example of the financial electricity market is the existence of electricity derivatives, which enable the prices contracted by the customers to be hedged beforehand, or the existence of a financial market for emissions, whose aim is to cut off carbon dioxide emissions. The conditions for a physical market to become a successful financial market are the existence of sufficient trading volumes, and of a supervisory organ to ensure that prices are genuine. If the underlying price reference is not authentic and liquid, as it happens in some European power markets, then financial trading will not develop. The reference price for the financial Nordic market is the system price determined by the Nord Pool. And, when entering financial contracts, technical conditions (eg. grid congestion) are not taken into account.

In the past, buyers and sellers of electricity negotiated long-term contracts in order to set the price of electricity on a bilateral basis. But currently, electricity can be traded in three different markets: in a power exchange, where market participants submit their respective generation and demand bids and a single price is determined; in bilateral over-the-counter trading, where a generator and a consumer agree on a trade contract and the market price can be the one published by the power exchange as a reference price; and in organised over-the-counter trading, where participants submit their respective generations, and demand bids that are cleared continuously. This last type of trading allows that one market participant may accept bids at two different prices.

Nevertheless, despite the electricity sector being considered as more of a "normal" sector in the last decades, electricity as a commodity presents a limitation: "*generation has to equal consumption (plus grid losses) on an instantaneous basis*". To prevent electricity markets from collapsing, they are designed to deal with this characteristic. This is why, on the supply side, there are different electricity markets depending on the time lag with respect to the moment of delivery. Forward and future markets, like in financial markets, offer contracts to supply or consume an agreed amount of electricity at a stipulated price in the future. These were set up in 1993 in the Nord Pool as one of the first more developed complex products introduced for trading. Electricity is traded the day before the day of delivery in day-ahead markets. And, the intra-day market, trades on the same day of delivery. Since electricity cannot be stored, these different markets allow buyers to guarantee that consumption will be satisfied through long-term contracts. All while ensuring that actual consumption (which is harder to forecast), is satisfied via short-term contracts (KU Leuven Energy Institute 2015). In spite of consumption being planned in advance by market operators and the flexibility the market provides, there are deviations in reality.

In order to correct any deviations that may occur, the figure of the Balance Responsible Parties (BRP) exists. It is an administrative entity that balances supply and demand from generators, suppliers and consumers. Balance Responsible Parties are "*responsible for the market's imbalances or fluctuations before the actual delivery*" (Glowacki Law Firm 2019). However, the definitive liability is under the Transmission System Operator (TSO), responsible for the instantaneous generation-consumption balance. Nordic TSOs use the spot market in order to manage congestion in the short-term. Its price mechanism is a useful instrument to handle potential insufficient transmission capacity (bottlenecks). Before the early commencement of trading, each TSO notifies the Nord Pool Spot of the capacity of interconnectors. Congestion in Sweden, Finland and Denmark is dealt with by the TSOs through counter-trade. In Norway, congestion is dealt with, by splitting the market in a way such that it reflects the limitations in the volume of transmission in the planning stage, and utilising counter trades in the operational one. TSOs can activate the amount of system reserves, also known as the Net Regulation Volume (NRV).

Reserves refer to how power plants may boost or diminish their production depending on the needs of the market. Power grids employ tools such as the frequency, to keep the grid steady when forecasts are not accurate or unexpected events occur. The frequency of the electricity grid should have a nominal value of 50.0 Hz (Fingrid 2016). Maintaining this frequency is vital, and generators must spin at this speed or the system may become unstable. In the situation of over-frequency, grid operators can respond by reducing the output from generators. In the situation of under-frequency, there are three stages. In the primary control stage, generators automatically adjust their output through frequency sensors. After 10 seconds, Automatic Generation Control is activated. If the under-frequency is not corrected within minutes through the previous stages, tertiary frequency control is prompted. It requires the manual adjustment of the output of the power plant by the power grid operator.

In order to achieve market balance, financial settlement of the BRP imbalances by the TSO must take place. This refers to the tariffs that the TSO imposes after the actual delivery to those BRPs that present an unbalanced portfolio. These tariffs are based on the Marginal Incremental Price (MIP), which is the highest price paid by the TSO for ascending activations for a given quarter of an hour, and on the Marginal Decremental Price (MDP) which refers to the lowest price received by the TSO for descending activations for a given quarter of an hour (KU Leuven Energy Institute 2015).

Market balance

Using the Elspot market (day-ahead market that trades one-hour duration power contracts) as an example, how the Nord Pool establishes the price will now be explained. Every day at 10pm TSOs perform a "*guaranteed transfer capacity*" preparing the bidding areas for the next day deals. At noon on the day before (day D-1), market participants present to the Nord Pool their offers or bids for the next 24 hours (starting at 1am on the next day, day D). The minimum contract size is of 0.1 MWh (Vehviläinen and Keppo 2001). Once the bidding for the following day is cleared, the Nord Pool makes cumulative supply and demand curves to balance production and consumption. The price of each hour is, of course, the equilibrium point or the point where both supply and demand cross. If there is sufficient transfer capacity, the equilibrium price will be the effective price (EnergyNet 2017).

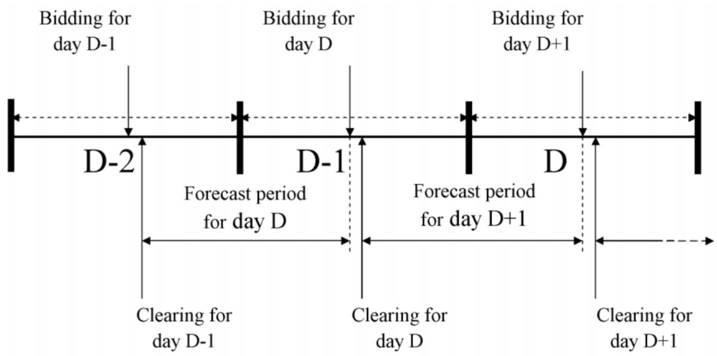


Figure 2.1: Time framework to forecast market prices for the Nordic day-ahead energy market.
Source: Price Forecasting in the Day-Ahead Energy Market by an Iterative Method with Separate Normal Price and Price Spike Frameworks (Sergey Voronin and Jarmo Partanen 2013)

Nevertheless, the situation in reality is usually different, and the Nord Pool must check if there are grid congestions in the power transmission grid, which can be potentially caused by the equilibrium price. In the case these bottlenecks are found, "area spot prices" will have to be estimated so as to encourage trading up to a point below the limited capacity of transmission. These "area spot prices" can also be thought of as different price areas that can be made up of different bidding regions. At 1pm, consumption and prices are published for the following day as can be seen in *Figure 2.1*.

2.4 Market power issue

The main issue in electricity markets is the existence of conflicts of interest, explained by the nature of the market. It is vital that any controversies are identified and disclosed to prevent any possible confidence issue. Lack of confidence in the market could result in a decrease in efficiency and liquidity.
Electricity prices have always cleared despite the denominated supply shocks caused by alterations in the supply of hydropower in Norway and Sweden, which are the biggest producers. In fact, after the creation of the Pool mar-

ket, productivity seemed to increase, and lower electricity prices were achieved (Bergman 2003b). Nonetheless, there are significant spikes in prices that could be explained by the exercise of market power at certain times when supply is scarce. Market power and monopoly are used interchangeably by economists (American Bar Association 2005). Nevertheless, some scholars do make a distinction across both terms. Those firms with "*trivial amounts of market power*" caused by uniqueness across products of competing firms, or because of superiority explained by innovation, should not raise antitrust concerns. Nevertheless, those firms with "*a substantial amount of market power, may be said to have monopoly power or antitrust monopoly power*".

There are two dimensions to the problem of market power. The first one arises when generators do not use all of their available capacity, increasing their profits by decreasing supply. The second one refers to the residual demand left every hour, after all generators except the biggest one, produce at full capacity. In this scenario, the biggest generator is a monopolist in relation to the hourly residual demand, which is inelastic. This means that the market price can be very easily altered by restraining their capacity. In fact, there was a claim that generators were colluding and withdrawing capacity from the Nord Pool to cause an increase in prices. The majority of the worry about market power is related to the withholding of hydropower back in 2002, which was a particularly dry year. Prices were expected to increase due to the decrease in generation, but the surge in prices was much higher than expected.

A more formal way to analyse whether there is market power or not, is by using the Lerner Index. It is a measure of market power that uses the difference between output price and cost, and divides it by the output price. It lies between 0 (perfect competitions) and 1 (strong market power). In the case of the Nord Pool, the degree of market power has been small through the years (Bergman 2003a).

Chapter 3

Empirical analysis

3.1 Data set description

The data set used in this paper corresponds to the day-ahead market operated by the Nord Pool. It contains hourly price data in EUR/MWh, ranging from the 1st January 1999 to the 26th January 2007. In total, there are 70,752 data points. The crucial aspect that differentiates financial time series analysis, such as the one presented in this section, from other time series analysis, is the element of uncertainty inherent to financial analysis (Tsay 2005). Due to this unpredictability element, statistical methods are vital in financial time series analysis. It is relevant to understand how prices fluctuate, and statistical methods are the legitimate approach to investigate them.

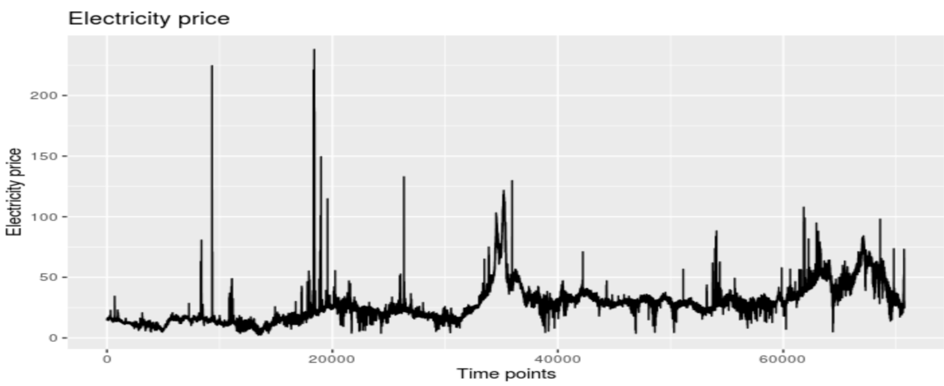


Figure 3.1: Plot of electricity prices January 1999 - January 2007

A time series Y_t is a "*discrete time, continuous state, process where t; t = 1,...,T are certain discrete time points*" (Aas and Dimakos 2014). Generally, points occur at equally spaced intervals. In the data set used in this paper, the intervals are hourly. *Figure 3.1* presents over 70,000 successive hourly prices of the Nord Pool in the aforementioned period. A slim upward trend can be noticed. It can be seen as well, that there are big variations across prices, and these seem to be cyclical.

When analysing financial assets, it is common to focus on the returns rather than the actual prices. There are two main motives that explain the focus on them (Campbell, Lo, and MacKinlay 2012). First of all, returns offer a "*scale-free summary of the financial opportunity*" and second, they offer statistical features that make working with them more appealing than working with prices.

As Daníelsson stated, there are two kinds of returns (Danielsson 2011): simple or arithmetic, and compounded or logarithmic returns. Simple returns are referred to as the percentage change in prices, and formally defined as:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{3.1}$$

Continuously compounded returns are the "*sum of continuously compounded single-period returns*". They are formally defined below, and the logarithms used are natural logarithms.

$$r_t = ln(\frac{P_t}{P_{t-1}}) \tag{3.2}$$

In *Figure 3.2*, the plotted logarithmic returns of the prices can be observed. There are plenty of high and low peaks coinciding with sudden increases in the raw prices, being the pattern along 0. These peaks illustrate extreme changes in prices, ie. extreme events.

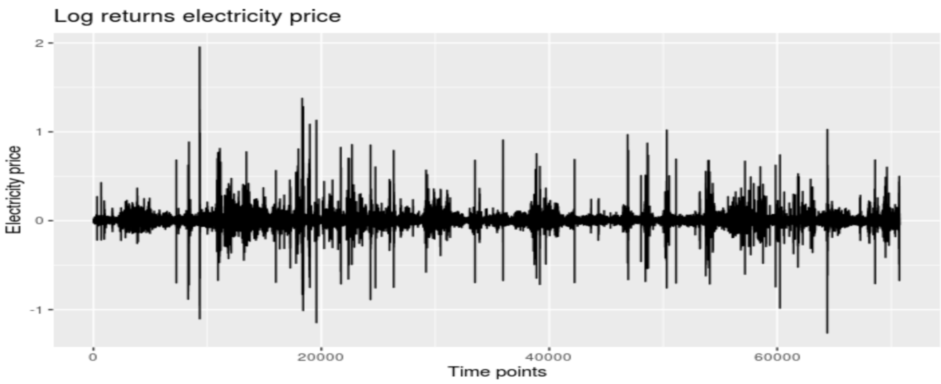


Figure 3.2: Plot of logarithmic returns of electricity prices January 1999 - January 2007

3.2 Extreme events

As it was explained before, it is obvious that due to the way of the market, the electricity market is particularly affected by extraneous factors that cannot be controlled, both in the short and long-term. Extreme events are caused by many factors, one of them being the weather. Bad meteorological conditions such as storms can damage power or distribution lines which may result in power cuts. Infrastructure needs to be repaired, and this results in higher electricity prices. Nonetheless, meteorological conditions such as heavy rain or snow can as well increase the generation of electricity. Since prices are also driven by supply and demand, an increase in supply would bring prices down. Nevertheless, this second example is less common, because the energy mix is greatly varied. Other examples of these factors that affect the electricity market (apart from the weather) are "*production disturbances in power plants, events affecting buyer decisions made by government agencies, and political decisions*" (Finansinspektionen 2005). It should be noted that political circumstances in particular, affect this market to a greater degree than the majority of markets.

World events also explain extreme events, particularly if there is an unsettling situation in countries that produce gas and oil. The latest example at the time of writing this paper, was the negative prices of oil barrels (WTI crude) due to the COVID-19 pandemic. It was the first time in history where oil prices - particularly known for its volatility - turned negative. Lockdown greatly decreased transport use across the planet, and the need to store the commodity in very specific conditions explains this drop. This implied that producers paid buyers to store the commodity.

These external events that greatly affect prices are defined as "extreme events" in statistics. Of course, the economic and social consequences of extreme events are "*a matter of enormous concern*" (Ghil1 et al. 2011). On top of this, electricity markets present a very inelastic demand. Since these price spikes are a recurrent issue in electricity markets, and we could also generalise to financial markets in general, academics have been trying to find a good approach to predict and deal with them. It is vital to do so, to guarantee that producers respond optimally to the pool and engage in successful contracts.

3.2.1 Extreme value theory basics

The aforementioned extreme events can be modelled through the extreme value theory. Extreme value theory started being developed in the 1950s in the field of civil engineering and it "*provided a framework in which an estimate of anticipated forces could be made using historical data*" (Coles 2001). As extreme values are scarce by nature, extrapolation from observed to unobserved levels is needed. In fact, much of the research made on predicting electricity spot prices only focuses on forecasting the next period prices. This limitation is a result of the high volatility present, and how easily can prices spike in short periods. The aim of extreme value analysis is to "*quantify the stochastic behaviour of a process at unusually large - or small - levels*". Its paradigm is the following:

$$M_n = max(x_1, ...x_n) \tag{3.3}$$

Assuming that $x_1, \ldots, x_n$ are observations up to period n. If the conduct of x_i were known, the analogous conduct of M_n would be known. Since the conduct of the observations is not known, it is impossible to obtain M_n. Nevertheless, given certain premises and letting n tend to infinity, models to derive M_n can be accomplished. This is attained by disentangling the randomness in the observations' process.

It is not only important to look at how prices spike, but the time between those spikes. According to novel research, "*the time between spikes has a significant impact on the likelihood of future occurrences*". Results indicate that there is evidence of dependence across the time between extreme events in the markets, which are directly related to the intensity of the stochastic process of price spikes. There are different approaches, one of them being the classical time series approach used by Conejo et al. (Conejo et al. 2005). Other academics focused on the estimation and forecasting of risk measures, proposing hedging strategies such as Vehviläinen et al. (Vehviläinen and Keppo 2001). And finally, others rather than concentrating on forecasting the price trajectory, focused on predicting the likelihood of extreme price events such as Lindsay et al. (K A Lindsay and T M Christensen and A S Hurn 2015).

3.2.2 Exceedances in the data set

Following the example of other papers, such as the aforementioned by Lindsay et al., this paper defines price exceedances as the situation where "*price exceeds a particular threshold value that is chosen to lie substantially outside the normal range of daily fluctuations*". To do so, in *Figure 3.3*, the median and the 10th and 90th percentiles are laid out. To obtain this plot in R, the data have been subset by creating vectors of positions for each hour of the day. That way, the position for values at 00:00 was formed by values starting in 1 (since the data set starts at 00:00), ending at 70,752 (the last data point) and the increment of the sequence was equal to 24 (as a day has 24 hours). Once these vectors of positions were defined, prices at those positions were grouped. That way the prices corresponding to each hour of the day were grouped. For each group of price points, the median was calculated by using the command `median`, and the quantile by using the command `quantile`, and specifying the desired percentile (either 0.1 or 0.9). The defined price spike or exceedance price must be above those lines. Therefore, the exceedance price was defined as 60/MWh, since the 90th percentile is still regarded as "natural".

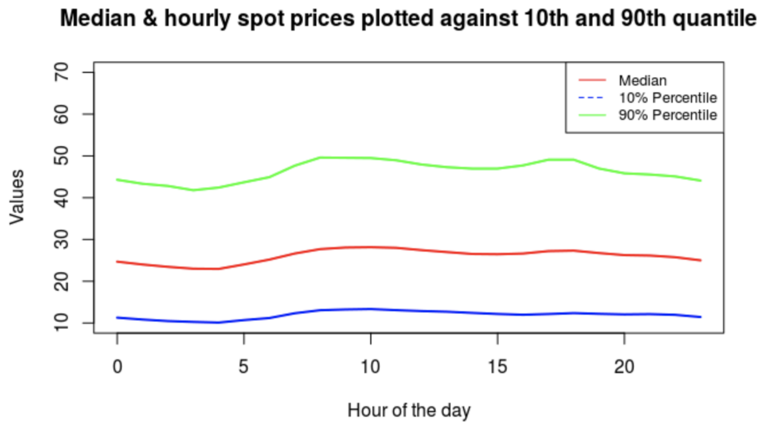


Figure 3.3: Median hourly spot prices plotted against 10th and 90th quantile

The tables below present the number of exceedances above 60/MWh on an intra-hour, intra-day and intra-month basis. The count of these extreme values was done in Excel. First, a column was added indicating either the hour of the day (from 00:00 to 23:00 and back to 00:00 again), the day of the week (24 rows with the value "Monday" followed by 24 rows with the value "Tuesday", and so on) or the month of the year (744 rows with the value "January" (also 744 rows for those months with 31 days), 672 or 696 rows with the value "February" depending on whether it was a leap year or not, and 720 rows with the value "April" (also 720 rows for those months with 30 days)). Once this was done, another column was added where the `IF` formula was used, which pointed out those price values greater than 60. Or, in other words, the bound to designate extreme values. Finally, a pivot table was created and filtered by those cells that contained values greater than 60 (obtained through the `IF` command). This way, the pivot table displayed exclusively the count of those extreme values. The results are highlighted below.

In the intra-hour exceedances table presented below (*Table 3.1*), the number of exceedances above 60/MWh can be observed on an hourly basis. It can be seen that exceedances are much more likely to occur between 7:00-11:00 and 17:00-20:00, than at any other time of the day. At these particular times, over 110 exceedances were found, the largest number of exceedances occurring at 8am. This is potentially explained by consumers' daily rhythms - waking up hours before work, and "at home" hours after work. On the other hand, few exceedances occur from 2:00 to 05:00. This is explained by the fact that the use of electricity at those times is, and has been, consistently low for some hours. This therefore, keeps the price typically low.

Time	Count of exceedances above 60/MWh
00:00	81
01:00	75
02:00	76
03:00	73
04:00	75
05:00	81
06:00	88
07:00	114
08:00	148
09:00	137
10:00	127
11:00	113
12:00	106
13:00	105
14:00	101
15:00	99
16:00	104
17:00	117
18:00	117
19:00	112
20:00	100
21:00	96
22:00	96
23:00	86

Table 3.1: Number of intra-hour exceedances

In the intra-day exceedances table (*Table 3.2*), which plots the number of exceedances on a daily basis, it can be seen that they are much more likely to occur on weekdays, particularly Wednesdays and Thursdays. Both of these days present over 380 exceedances, while weekend days present just over 280. This is consistent with the previous hypothesis about daily rhythms. More exceedances will be present on weekdays, since it is on these days where consumers wake up earlier and mostly at the same time, causing a sharp rise in electricity demand.

Weekday	Count of exceedances above 60/MWh
Monday	371
Tuesday	355
Wednesday	387
Thursday	382
Friday	363
Saturday	282
Sunday	286

Table 3.2: Number of intra-day exceedances

Finally, *Table 3.3* groups exceedances by month. It can be observed that the month that presents the highest number of exceedances is December followed by September, August and January.

Month	Count of exceedances above 60/MWh
January	396
February	40
March	44
April	16
May	0
June	1
July	0
August	534
September	631
October	79
November	19
December	667

Table 3.3: Number of intra-month exceedances

Interestingly, most spikes in the summer months occur in the year 2006. According to Lewiner, this is explained by adverse hydro conditions, "*the impact of emissions and a general overreaction of market players and speculators*" (Lewiner 2007). An added problem, were the unexpected plant outages in Sweden, that put pressure on prices. The Nord Pool still presented the lowest mean spot trading price in Europe, but 2006 was a tricky year where the historic price gap with respect to the rest of Europe was eliminated. Furthermore, according to the literature, the intensity of these exceedances does exhibit a temporal dependence.

3.3 Summary statistics

Table 3.4 plots several summary statistics (minimum and maximum values, mean and median) of the logarithmic returns at each hour of the day. It also presents information about the skewness, kurtosis and variance at each hour. These results were obtained in R. To attain them from an hourly perspective, the data have been subset by creating vectors of positions for each hour of the day. That way, the position for values at 00:00 was formed by values starting in 1 (since the data set starts at 00:00), ending at 70,752 (the last data point) and the increment of the sequence was equal to 24 (as a day has 24 hours). Once these vectors of positions were defined, the logarithmic returns of the prices at those positions were grouped, in order to create the vectors corresponding to the logarithmic returns corresponding to each hour of the day. The mean, median, minimum and maximum values were obtained by using the command `summary`, the variance by using the command `var` and the skewness and kurtosis by the commands `skewness` and `kurtosis` respectively.

3.3.1 Minimum, maximum and mean values

Different conclusions can be obtained by looking at *Table 3.4*. First of all, the highest drop in logarithmic returns occurs at 00:00, followed by 08:00 and 17:00. This means that at the aforementioned times, there was most likely a

Time	Min	Max	Mean	Median	Variance	Skewness	Kurtosis
Overall	-1.264	1.955	0.0000067	-0.0029334	0.003491878	1.930427	86.41983
00:00	-1.264	0.0434	-0.03579	-0.02126	0.002927647	-8.095034	127.4326
01:00	-0.564	0.693	-0.025681	-0.014131	0.003010217	2.952077	75.96874
02:00	-0.756	0.145	-0.0178828	-0.0039849	0.00333136	-8.036666	82.91733
03:00	-0.711	0.506	0.004152	0.005652	0.002718644	-1.841327	42.28428
04:00	-0.246	1.02	0.041234	0.025797	0.00491896	4.635324	43.7548
05:00	-0.663	0.856	0.04870	0.03473	0.00522026	2.284948	25.84884
06:00	-0.424	1.955	0.06142	0.04093	0.009349064	6.936157	87.69128
07:00	-0.101	1.13	0.04965	0.02508	0.007659255	5.366954	40.82311
08:00	-1.147	0.589	0.004674	0.002853	0.003642484	-3.770894	71.4212
09:00	-0.883	0.172	0.002181	0.002727	0.001984682	-7.710851	115.9262
10:00	-0.955	0.22	-0.007825	-0.003360	0.001857343	-10.77955	183.7601
11:00	-0.63	0.06	-0.017432	-0.009981	0.001020176	-7.318405	87.33456
12:00	-0.324	0.093	-0.013289	-0.007805	0.0004714429	-4.613383	42.57078
13:00	-0.23	0.152	-0.010609	-0.006369	0.0003677554	-2.048213	18.13248
14:00	-0.238	0.5	-0.005468	-0.004536	0.001046861	4.199396	62.82241
15:00	-0.193	0.809	0.005864	0.001552	0.001863455	6.603826	95.34644
16:00	-0.277	1.249	0.0182720	0.0088300	0.002734446	10.42916	185.4612
17:00	-1.104	0.692	-0.0001489	-0.0006579	0.003248869	-2.947752	87.77065
18:00	-0.83	0.298	-0.016335	-0.009851	0.003205587	-7.067578	78.00707
19:00	-1.011	0.2	-0.01630	-0.01426	0.001671708	-8.035321	152.7783
20:00	-0.287	0.313	-0.004282	-0.003650	0.000580473	0.6134044	29.0255
21:00	-0.263	0.165	-0.0139099	-0.0132510	0.0006825597	-0.2084758	6.117713
22:00	-0.394	0.06	-0.03762	-0.02912	0.001094724	-2.389097	11.3119
23:00	-0.984	0.408	-0.013423	-0.005953	0.003212052	-3.923047	49.93816

Table 3.4: Summary statistics per hour

sudden drop in demand that caused the price to fall dramatically. On the other hand, the highest increase in logarithmic returns occurs at 06:00, followed by 16:00 and 07:00. A sudden increase in demand or decrease in supply, caused the price to jump forcefully. These are unique events, and it is therefore more illustrative to look at the mean of the logarithmic returns. The highest mean is the one corresponding to 6am, followed by 7am and 5am. At these times, the logarithmic returns are the largest on average, which implies that in these one-hour frames, the demand of electricity increased significantly causing the consequent rise in price.

This is explained by the fact that most consumers wake up between 5:00 and 7:00, increasing demand sharply. On the other hand, the lowest mean average returns are observed at 10pm, followed by midnight and 1am. This implies that the prices presented a decreasing jump due to two main factors. After 10pm, many consumers decrease their use: less laptops, washing-machines or dish-washers are in use, and many factories and offices close. A further decrease in use happens later on, since most users are already sleeping. This, combined with the fact that generators still produce energy, drives prices even lower. The overall mean is very close to zero but of course this is not significant for the hourly means - which as it has been seen, range from negative to positive values.

3.3.2 Variance

The variance is more widely known in financial analysis as volatility. It is defined as "*a statistical measure of the dispersion of returns (...). In most cases, the higher the volatility, the riskier the security*" (Kuepper 2020). In simpler words, volatility measures how much a price swings around the mean price, or how variable is the price. As it has been previously stated, the mean overall price is around 0. Therefore, the volatility will measure how much prices peak (ie. how much the price will swing or fluctuate around 0). The overall variance of the logarithmic returns is equal to 0.003491878. As before, this is not very significant. For that reason, variance per hour should be the focus of the analysis.

The smallest variance is present at 1pm and the largest variance at 6am. This means that the market is the most volatile and presents the most price fluctuations at 6am. Again, this could be explained by the fact that consumers may wake up around that time, demanding a lot of electricity particularly in winter, and causing prices to rise. It is true that most consumers wake up around 7:00 and 8:00, but at these times the volatility (ie. the jump in prices above the average) is smaller because the "jump" has already happened (in this case, explained by those waking up at 6am). On the other hand, the smallest volatility is found at 1pm. At this time most users are already sleeping, and have been for a while. Therefore, the demand is consistently low, and it has been for a couple of hours. Consequently, prices that have already been stable for some time, remain stable.

3.3.3 Skewness

Skewness refers to "*the distortion or asymmetry in a symmetrical bell curve, or normal distribution in a set of data*" (Chen 2019), and is formally defined, according to Dr Donald Wheeler (Wheeler 2004), by:

$$a_3 = \sum \frac{(X_i - \overline{X})^3}{ns^3} \tag{3.4}$$

where X_i is the value, $\overline{X}$ is the mean, n is the sample size and s is the standard deviation. A symmetrical data set will present a skewness of 0 because for each data point at distance t there will be a data point at distance -t, balancing it out.

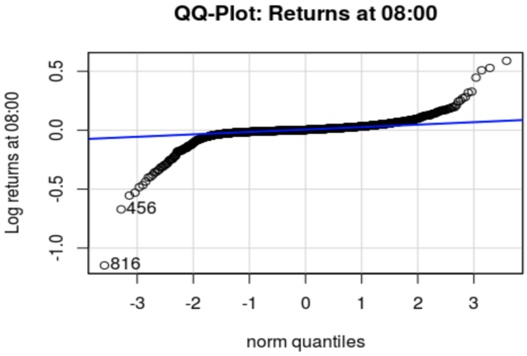
An essential characteristic of normally distributed observations, is that "*they are completely described statistically by the mean and the variance (ie. the first and second moments)*" (Danielsson 2011). This entails that skewness is identical for all normally distributed variables, meaning that it is equal to 0. Negative skewness occurs when the mean is less than the mode. In this case, the left tail will be larger than the right tail. Positive skewness occurs when the mean is greater than the mode. The right tail will be larger than the left tail in this situation.

From the previous explanation, it can be inferred that the existence of spikes produces skewness. The lowest value is present at 21:00, where skewness is equal to -0.20. Since it is the closest to zero, it can be understood that the distribution is roughly symmetric. A graphical proof of it will be shown in the following section. On the other hand, the skewness for the overall dataset is equal to 1.93. When skewness is greater than |1|, the data are highly skewed. And, in fact, the property of extreme skewness is typical in electricity price data (K A Lindsay and T M Christensen and A S Hurn 2015). Furthermore, according to the empirical literature, these spikes display seasonal dependence.

3.3.4 Kurtosis

Kurtosis measures "*the degree of peakedness of a distribution relative to the tails*" (Danielsson 2011). Excessive kurtosis, like the one present in the data, suggests that "*more of the variance is due to infrequent extreme deviations than predicticted by the normal, and is a strong, but not perfect, signal that a return has fat tails*". In line with this, it is important to explain that kurtosis is rejected by some academics, whose reasoning is the following. Despite being a "*measure of the peakedness of a distribution*" according to Dr Donald, "*the central portion of the distribution is virtually ignored (...). While there is a correlation between peakedness and kurtosis, the relationship is an indirect and imperfect one at best*".

By having observed these two measures, it can be concluded that the data is far from normal. Still, there are more formal tests such as the Kolmogorow-Smirnow and the Jarque-Bera tests, and graphical tests such as the use of quantile-quantile plots or sequential moments. Quantile-quantile plots, or qq plots for short, gauge whether data points follow a specific distribution (Ford 2015). By plotting two sets of quantiles against the other, and in the case that both sets of data followed the same distribution, a (pseudo)straight line of dots would be achieved. *Figures 3.8 and 3.9*, present the qq plots for data points at 08:00 and 21:00. The blue line represents the normal prediction, the horizontal axis measures the standard normal and the vertical axis shows the results of the data. These are obtained by using the `qqPlot` and `qqline` commands in R.



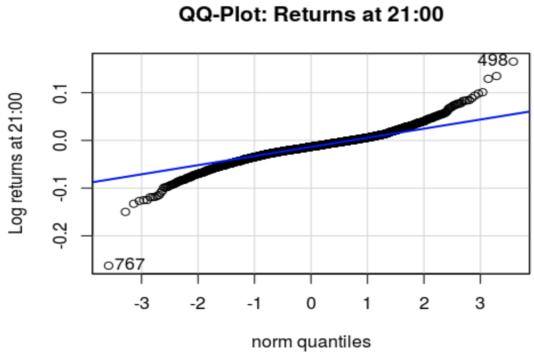


Figure 3.4: qq-plot: Returns at 08:00

Figure 3.5: qq-plot: Returns at 21:00

As it can be observed by the "s" shape in *Figure 3.4*, the data points diverge from the normal prediction. These fat tails that entail non-normal returns, have different effects on how financial data is handled and which conclusions can be obtained. This is mainly caused by the normality assumptions made by the models that are often used, particularly regarding risk handling (where risk is underrated). The closest returns to a normal distribution, as was mentioned in the skewness analysis, are prices at 21:00, as the dots follow an almost straight line. This can be observed by looking at *Figure 3.5*

3.3.5 CDF and PDF

Further analysis of the distribution of the data set will be performed by looking at both the cumulative distribution function and the probability density function. This is done, since high skewness could imply that the mean and variance are not representative and should not be used.

A cumulative distribution function presents the "*probability that the variable takes a value less than or equal to x*". Or, more formally, the cumulative distribution function $F_x(x)$ of a random variable X is defined by:

$$F_x(x) = P(X \leq x) \tag{3.5}$$

It follows the definition of the probability density function of X that

$$P(X \leq x) = \int_{-\infty}^{\infty} f_x(t)dt \tag{3.6}$$

The empirical cumulative distribution function (ECDF) is established on the "*relative frequencies of the observed data*". It is defined as a step function, as it is constant between consecutive order statistics. The ECDF plotted against the CDF is presented below. It was obtained by using the command `ecdf`. As it can be observed, the ECDF is equal to 0 for smaller values than zero, and it is equal to 1 for larger values than zero. Through the ECDF, one can obtain the shape of the frequency distribution. As it can be inferred from *Figure 3.6*, it presents heavy tails.

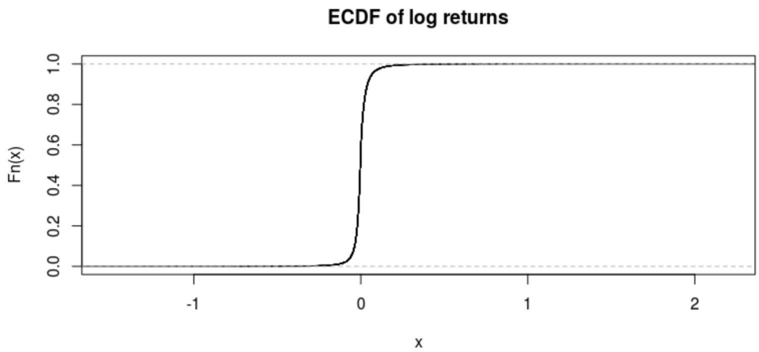


Figure 3.6: ECDF of logarithmic returns

On the other hand, the probability density function is the "*idealised frequency distribution of observable random variables*" (Gentle 2020). Or, in other words, the PDF enables one to obtain the probability for a random variable to appertain to a particular collection of values. Probability density can be formally defined as:

A random variable X is continuous if there is a function f_x such that for all intervals [a,b],

$$P(X \in [a,b]) = \int_a^b f_x(x)dx \tag{3.7}$$

We call the function f_x the probability function of X (or pdf of X).

PDFs have two main properties:

1. $\int_{-\infty}^{\infty} f_x(x)dx = 1$
2. $f_x(x) \geq 0$

for all x.

Dividing the sample into bins is done in order to obtain an overview of the frequency distribution. The most typical representation of this is through a histogram. And, the capacity of the bins dictates how "smooth" the histogram will be. In order to determine the number of bins, the Sturge's rule has been used in R, and 12 bins were determined to be employed. It is important to bear in mind that the larger the bin scope, the smaller the variance and the larger the bias. The PDF is plotted in *Figure 3.7* , and as it can be observed, it presents very long tails (particularly the right tail).

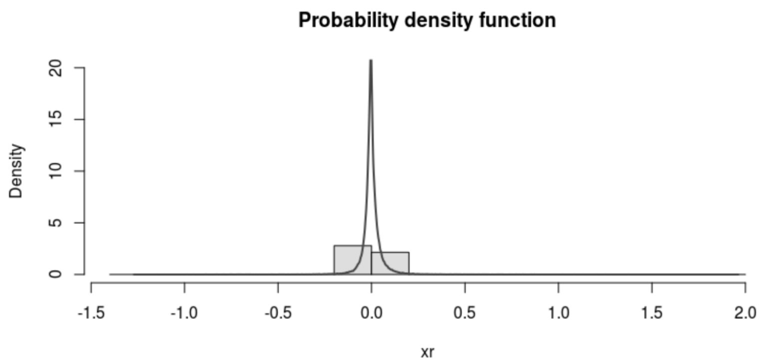


Figure 3.7: PDF

Therefore, as the previous section indicated, the data is highly skewed. Nonetheless, when narrowing down the data set into hourly data points, the skewness decreases. As example of this, are data points at 21:00. This has been proven both numerically and graphically in the last section.

3.4 Stationarity definition and testing

Another important characteristic to look at is stationary. Intuitively, stationarity means that "*the statistical properties of a process generating a time series do not change over time*" according to Palachy (Palachy 2019). Or, in other words, the time series does not present any systematic change in its mean and variance and has no periodic variations.

There are two types of stationarity. It must be borne in mind that strict stationarity does not imply weak stationarity or vice versa. Strict or strong stationarity refers to when the distribution of a time series is the same throughout time. More formally, the time series $X_t, t \in Z$ is said to be strictly stationary if the joint distribution of $(X_{t_1}, X_{t_2}, ..., X_{t_k})$ is the same as that of $(X_{t_1+h}, X_{t_2+h}, ..., X_{t_k+h})$ (Palachy 2019). In other words, strong stationarity means that the joint distribution only depends on the "difference" h, not the time $(t_1, ..., t_k)$.

On the other hand, in weak stationarity only the mean and covariance are invariant throughout time. For this reason, weak stationarity is also known as covariance or mean stationarity. It is formally defined below:

The time series $X_t, t \in Z$ (where Z is the integer set) is stationary if:
1. The first moment or expected value of x_i is fixed, $E[x_i] = \mu$ where μ is independent of i.
2. The second moment or variance of x_i is finite for all t, $E[x_i^2] < \infty$ (which also implies that variance is finite for all t).
3. The cross moment or auto-covariance depends only on the difference $u - v, \forall u, v, a, cov(x_u, x_v) = cov(x_{u+a}, x_{v+a})$.

To sum up, a weak stationary time series presents a constant expected value, finite variance and the auto-covariance only depends on u-v, not on u or v. This can be visually checked by plotting the mean and variance, and seeing whether they are constant over time. As it can be observed in *Figure 3.8*, the mean at each hour (plotted in red) is not constant and therefore does not

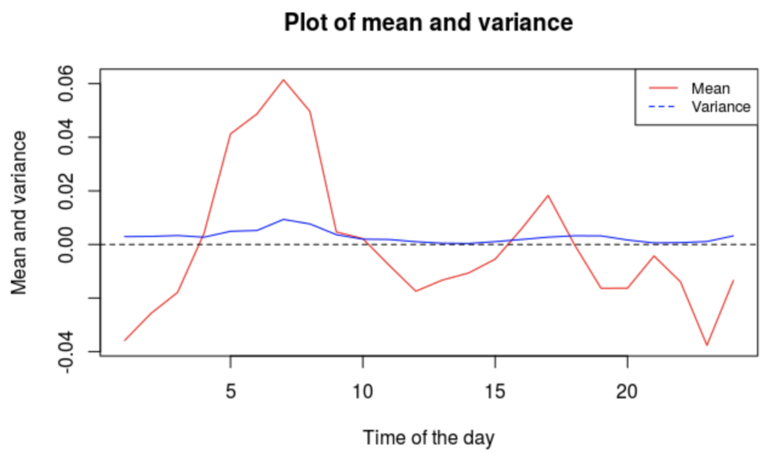


Figure 3.8: Mean and variance plot

respect the first condition stated above regarding stationarity (ie. the first moment is fixed). When looking at the variance at each hour (in blue), it can be seen that it is not constant. This does not respect the second condition stated above. The dotted black line is a straight line at the value 0, achieved by using the command `abline`. It was added to improve the layout and use of the graph.

Non-stationarity can be solved by calculating the difference across successive data points, also known as differencing. The intuition behind this is the following: if the initial data set does not have fixed features over time, the variation from period to period might do. To check more rigorously whether differencing is required in the current data set, there are several tests available. The tests that assess whether a time series is non-stationary and has a unit root are called unit root tests (Dhankar 2019). Unit root tests were pioneered by statisticians David Dickey and Wayne Fuller. The rationale behind these tests is the following. They start off with:

$$y_t = \rho y_{t-1} + m_t \tag{3.8}$$

where m_t is a white noise error term.

By subtracting y_{t-1} from both sides, the following would be otained,

$$\Delta y_t = (\rho - 1)y_{t\text{-}1} + m_t \tag{3.9}$$

where $\delta = \rho$ - 1 and Δ is the first difference operator. The null hypothesis is that $\delta = 0$, which entails that $\rho = 1$ and implies the existence of a unit root and the non-stationarity of the time series.

When performing the Dickey-Fuller test in the data by using the command `adf.test`, the p-value obtained is equal to 0.01. This means that the null hypothesis (stating that the data is non-stationary) needs to be rejected. Nevertheless, it must be noted that stationarity tests are not totally trustworthy, since they assume that variance is constant. This could generate distortions in the results of these tests.

3.5 Heteroskedasticity

Conditional heteroskedasticity can be explained as "*volatility clustering, in which large changes tend to follow large changes, and small changes tend to follow small changes*" (Wang et al. 2005). This can be more intuitively explained when referring to the stock market. In those instances where these markets suffer a considerable decrease, panic is generalised and the automated risk management systems in place short their positions. This further depresses prices. Therefore, what was an initial drop in price, "*leads to significant further downward volatility*" (Auquan 2017).

In the current dataset, heteroskedasticity may be caused by a decrease in production or power outages, initially supported by reserves but causing a higher and higher rise in prices. Therefore, these series are conditional heteroskedastic.

Chapter 4

Modelling time series

Forecasting is of vital importance, not only in the financial sector currently, but has been since 300BC when people would consult the Oracle (Hyndman and Athanasopoulos 2018). The demand for electricity can be relatively easily forecasted since the causes of high demand are understood and there is historical data regarding demand and weather conditions. Nevertheless, when talking about electricity prices, more considerations have to be taken into account and the only accessible information is historical data.

4.1 Complex seasonality

High frequency time series such as cash requests at ATMs and electricity usage, tend to display complex seasonal patterns (Hyndman and Athanasopoulos 2018). In these instances, data may present two or even three seasonalities (weekly, monthly and yearly, for example). This is the case of the data set this paper deals with. To treat this complex seasonality, the `msts` class is used in R, since the `ts` class can only deal with one type of time series seasonality. In order to find the existing seasonalities, the following graphical analysis was performed in Excel.

Figure 4.1 presents the average electricity price at each month of the year (from 1999 until 2006) and the coefficient of variation, which is calculated by dividing the standard deviation by the mean.

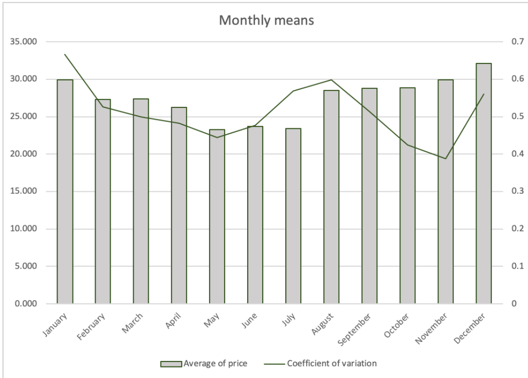


Figure 4.1: Monthly means

This is done in Excel, by grouping all prices by each month of the year, and calculating the average and the standard deviation. The coefficient of variation shows the dispersion around the mean. In those instances where the coefficient of variation is high, the level of dispersion around the mean is higher (Institut national de la statistique et des études économiques 2016). When it is low, then the mean estimation can be considered as precise. When looking at *Figure 4.1*, it can be seen that there is a decrease in average prices in warmer months (from May to July). In fact, standard deviation is the lowest in May. This is explained by the decrease in the use of electricity as the hours of light per day increase, and temperatures rise. On the other hand, standard deviation is the highest in January. This is possibly associated with the holiday season and the decrease in temperatures. From this graph, the monthly seasonal effect can be distinctly observed, so it must be included when modelling.

Figure 4.2 shows daily average prices and daily coefficient of variation. The aforementioned coefficient is notably higher on Sundays, Mondays and Saturdays, indicating the high level of dispersion around the mean. When looking at the average price, there are consistently high prices from Mondays to Thursdays, that decrease on Fridays and contract further as the week approaches its end. This is probably caused by the shut down of the industrial sector during weekends, causing supply to outweigh demand. As demand increases sharply on Mondays once the working week starts again, the average price is the highest. By looking at this plot, the weekly seasonal effect can be observed.

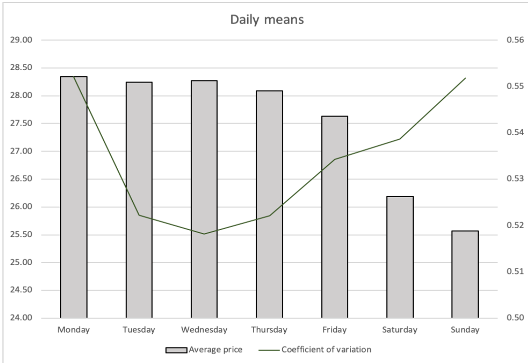


Figure 4.2: Daily means

Lastly, the plot of hourly means on weekdays can be seen in *Figure 4.3*. The highest average price is on Mondays equal to 32.26 EUR/MWh at 08:00, and the lowest on Sundays, equal to 23.28 EUR/MWh at 06:00. This is a difference of 8.98 EUR/MWh between the highest and the lowest hourly prices.

Figure 4.3: Hourly means across weekdays

When looking at the coefficient of variation, it can be seen that it is consistently greater than the mean at night (from 00:00 to 05:00). This indicates that the level of dispersion around the mean is high. The highest price on working days is at 08:00, and it is usually accompanied by a rise in the coefficient of variation. As it can be expected, on weekends the price is lower in the mornings when compared to the rest of the week. There is a clear daily seasonality that must be taken into account.

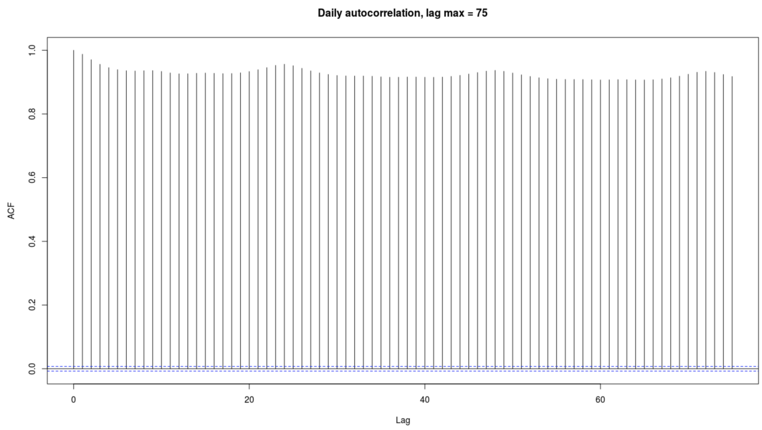


Figure 4.4: ACF

To sum up, it can be assumed that there is a monthly, a weekly and daily seasonality, as it has been explained in the previous analysis. The aforementioned daily seasonality can be seen in *Figure 4.4*, where the Autocorrelation Function (`acf`) is plotted. The daily autocorrelation plot includes 75 lags, to be able to detect the autocorrelations of 3 days. As it can be observed, autocorrelations follow a similar behaviour every 24 hours.

4.2 Price decomposition

The decomposition of prices is plotted below, according to the seasonalities reached in the previous section (daily, weekly and monthly). The plot was produced by using the `msts` command in R. The aim of this decomposition is to be able to separate the deterministic components (trend, T_t and seasonality, S_t) from the random noise, R_t. Formally, decomposition can be written as:

$$y_t = (D_t + W_t + M_t + Y_t) + T_t + R_t \tag{4.1}$$

where $(D_t + W_t + M_t + Y_t)$ is the sum of the seasonalities in y_t.

The random noise should present stable variations and no trend. The first plot in *Figure 4.5*, "data", presents the raw data of electricity prices. It is also defined as the observed component. As it can be observed, it presents increasing jumps in year 2 (2001), year 3 (2002) and year 5 (2004). The second component, the "trend", presents the pattern that the series follows. As it was stated before, the series presents an upward trend, which can be clearly seen in the plot. The trend does not present any pattern, meaning that any seasonal effects have been successfully captured. In the case in which no complex seasonality had been specified, the decompose command would not have outputted a de-seasonalised trend.

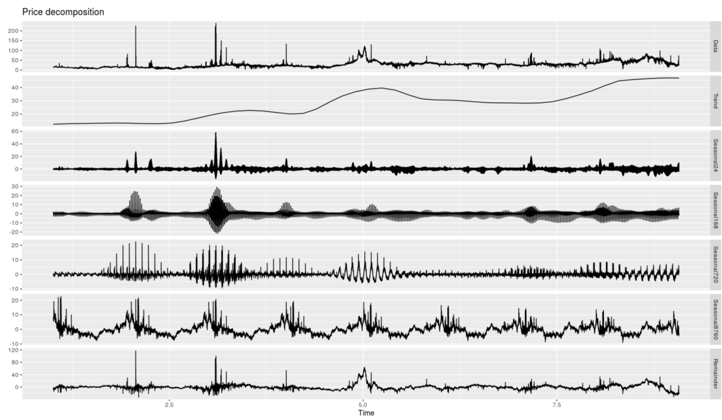


Figure 4.5: Decomposition visualisation

The "seasonal" components present the yearly (Seasonal8760), monthly (Seasonal720), weekly (Seasonal168) and daily (Seasonal24) variations in price. It is calculated by obtaining the average for "*each time unit over all periods, and then centering it around the mean*" (Prastiwi 2019). As it can be seen, yearly and monthly components present a remarkably clear seasonality. Lastly, the "random" component presents what cannot be explained by neither the trend nor the seasonal components.

4.3 Theoretical framework

ARIMA model

When it comes to forecasting, ARIMA models are one of the most popular approaches used. ARIMA models are made up of both an autoregressive (AR) and a moving average (MA) component. Autoregressive models, as it can be inferred by the name, use a regression of the variable against itself. Basically, they explain the present value of the series y_t by using past values $y_{t-1}, y_{t-2}, ..., y_{t-p}$. More formally, an autoregressive process of order p (also referred to as AR(p)) is written as:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + ... + \varphi_p y_{t-p} + z_t \tag{4.2}$$

where z_t is white noise and it is not correlated to y_s for each $s < t$. These models are significantly adaptable when dealing with a wide range of time series sequences.

Moving average models are defined as "*the previous forecast errors*". After a forecast is made, an error term will be outputted, and these error terms may enhance successive forecasts (Kacapyr 2014). Values of y_t can be seen as "*a weighted moving average of the past few forecast errors*". More formally, y_t is a moving average process of order q if:

$$y_t = c + z_t + \Theta_1 z_{t-1} + ... + \Theta_q z_{t-q} \tag{4.3}$$

where z_t is white noise and $\Theta_1, ..., \Theta_q$ are constants.
As it was stated before, these two models can be combined and generalised into an ARIMA (Autoregressive Moving Average) model, which can be written as:

$$(1 - \Phi_1 B - ... - \Phi_p B^p)(1 - B)^d y_t = (1 + \Phi_1 B + ... + \Phi_q B^q) z_t \tag{4.4}$$

where $z_t \sim WN(0, \sigma^2)$, and d is a non-negative integer. The d-term refers to the degree of differencing utilised. In the situation in which d = 0 (usually when the data are already stationary and do not need differencing, as stationarity is a requirement for these models to be implemented), the ARIMA (p,d,q) model is equivalent to an ARMA (p, q).

ARIMA methodology is executed through an iterative process until the most suitable model is found. It is a "*step-by-step process of model identification, specification, estimation, diagnostic and forecast*" (Siluyele and Jere 2016). ARIMA models are particularly useful for short-term forecasting, as they strongly stress recent past means.

ETS (M,N,N) model

Once the data are decomposed, the ETS approach is used in order to make a forecast. ETS refers to simple exponential smoothing with multiplicative errors, and stands for Error, Trend, Seasonal, or ExponenTial Smoothing (Hyndman and Athanasopoulos 2018). The main benefit of this method is that it can deal with any type of seasonality, and seasonality can vary.

Among others, this model is considered a state space model. These models are made up of an equation that details the observed data, and state equations that present how the other components vary as time passes by. It can be formally defined as:

$$y_t = l_{t-1}(1 + \varepsilon_t) \tag{4.5}$$

$$l_t = l_{t-1}(1 + \alpha\varepsilon_t) \tag{4.6}$$

This comes from making the errors relative,

$$\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}} \tag{4.7}$$

LOESS regression

The Seasonal and Trend Decomposition (`stlf` in R) will be used in ARIMA and ETS methods. It uses the decomposed version of the data in order to make forecasts. In order to obtain the trend through the `stlf` method, the

LOESS regression will be used. It merges the clarity of the linear least squares regression with the adaptability of the non-linear regression. It does so, by "*fitting simple models on local subsets of data to create a function that describes the deterministic part of the variation in point-to-point data*" (Najera 2019).

To prevent over and underfitting, the goal is to get the number that minimises the estimation errors. In order to obtain this figure, the `loess.as` function will be used in R. According to R's help file, this command "*fits a local polynomial regression with automatic smoothing parameter selection*". The obtained value is equal to 0.05048568.

TBATS model

The TBATS model was introduced by De Livera, Hyndman and Snyder in 2011. It is an "*exponential smoothing method, which includes Box-Cox transformation, ARMA model for residuals and the Trigonometric Seasonal*" (Orhan Altug Karabiber and George Xydis 2019). All of this enables this method to decrease the number of components in the model, all while dealing with high seasonality.

An advantage of TBATS, is that it does allow seasonality to vary slowly. Furthermore, TBATS will "*acknowledge models with and without trend, Box-Cox transformations and with non-seasonality*" (Skorupa 2019). Then, it will choose the most appropriate model by using the Akaike information criterion (AIC). The main downside is that it takes a lot of time to process.

Seasonal naïve method

A seasonal naïve method will be used as a reference point with respect to the other models (Orhan Altug Karabiber and George Xydis 2019). If any of the other models performs worse than the naïve method, then they will be discarded. This is based on the main assumption of the naïve method, which considers the forecast equal to the last observed value from the same season (Hyndman and Athanasopoulos 2018).

That is, if prices decreased in December 2006, the model will assume that they decrease as well in December 2007, 2008 and so on (for yearly seasonality). This can be formally expressed as:

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)} \tag{4.8}$$

where m stands for the seasonal period and k is the number of full seasons in the forecast period before time T + h.

4.4 Fitting and forecasting

Before performing the forecast, the data set has to be divided into the train and test sets. The test set will then enable the accuracy check to be performed once the forecast is made. The chosen test set has a length of 20 per cent the length of the data set. The forecasting produced by using all 70,752 data points is included in *Appendix A*. In order to obtain a higher accuracy and to guarantee that processes such as the `tbats` work relatively quickly, 2,880 data points were selected, which accounts for 4 months worth of data. As the selected data is a subset of the whole data set, only daily and weekly seasonalities will be included. The train set will have a length of 2,304 and the test set will be composed of 576 data points.

Forecast with SLT and Arima

As it was already mentioned, the daily and weekly seasonalities have been included - as the sample is not long enough to include the monthly and yearly ones. This has been included in the forecast command (`stlf`) by adding the `s.window` specification. Furthermore, the value obtained with the LOESS regression (0.05048568) is included as well by including the `t.window` argument. The selection of the ARIMA method has been stated as well, and the non-stationarity of the data has been solved by differencing once (as d=1). The fitted values in green, can be observed in *Figure 4.6*.

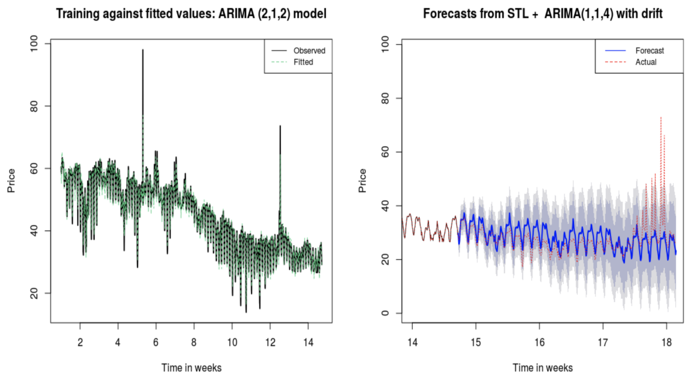


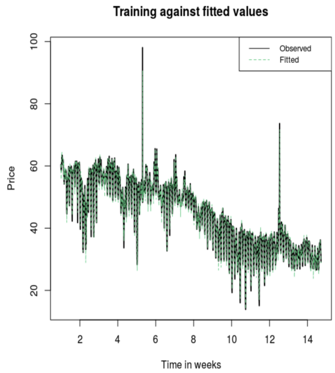
Figure 4.6: Forecast using SLT and ARIMA (2,1,2) against real values

Figure 4.7: Training against fitted values using SLT and ARIMA

As it can be observed, the model seems to be quite accurate. Furthermore, when comparing the forecast produced with the observed values (the test set), the model seems to be relatively accurate, as can be observed in *Figure 4.7*, and confidence intervals are not too big.

Forecast with SLT and ETS

The same arguments have been included when building the ETS model, by using the `t.window`, `s.window` and `method` specifications. The fitted values are plotted along with the observed values in *Figure 4.8*. Once again, the model seems to be accurate. When comparing the forecast produced with the observed values (the test set), *Figure 4.9* is outputted. As it can be seen, the confidence intervals at 80 and 95 per cent are much wider than the ones outputted by ARIMA. Just by looking at these plots, it can already be observed that the ARIMA forecast is more accurate.



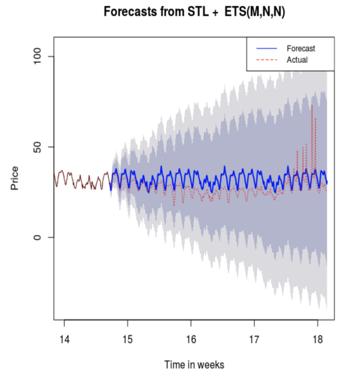
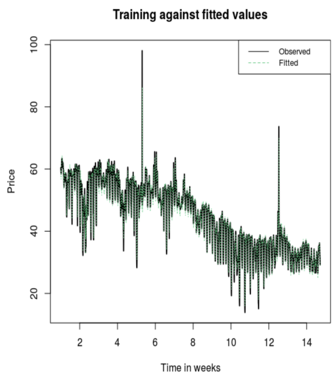


Figure 4.8: Training against fitted values using SLT and ETS

Figure 4.9: Forecast using SLT and ETS against real values

Forecast with TBATS

The main downside from the TBATS method (`tbats` function) is that processing is very slow.



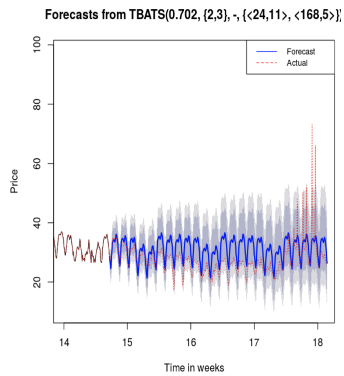


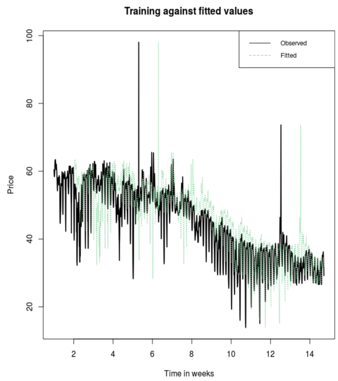
Figure 4.10: Fitted values against real values using TBATS

Figure 4.11: Forecast against observed values using TBATS method

For that reason, the number of data points chosen to perform these forecasts was reduced. Otherwise, R would be interrupted unexpectedly. The fitted values can be observed in *Figure 4.10*. The model seems to accurately fit the observed values. The forecast produced by TBATS is plotted against the actual values in *Figure 4.11*. As it can be seen, the forecast follows a similar pattern to the one followed by the observed values and the confidence intervals seem to be relatively small.

Forecast with seasonal naïve method

Finally, as it was previously explained, the seasonal naïve method (`snaive` method in R) was used as a benchmark model. The fitted values obtained by using this method are plotted in *Figure 4.12*.



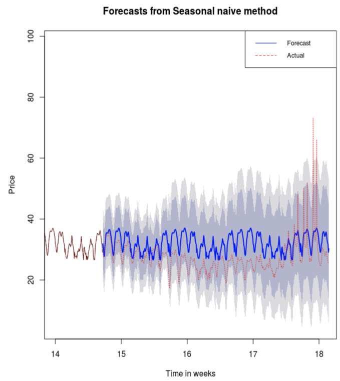


Figure 4.12: Training against fitted values using naïve method

Figure 4.13: Forecast against observed values using naïve method

As it can be observed, the model fits the observed data relatively well. When comparing the forecast produced with the observed values, *Figure 4.13* presents how accurate the forecast is.

4.4.1 Evaluation of forecasts

Now that the forecasts have been produced, it is relevant to look at which model has more predictive power, by using the `accuracy` command in R which compares the forecasted values to the observed values. This command "*returns a range of summary measures for the forecast accuracy*" according to the R library. It provides measures such as the mean error, root mean squared error and mean absolute scaled error.

Those indicators that will be used in this paper, will now be formally defined.

- The Mean Absolute Error (MAE) calculates the mean of the absolute values of the errors. It is defined as:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t| \tag{4.9}$$

- The Root Mean Squared Error (RMSE), calculates the unit root of the square of the error. It is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}e_t^2} \tag{4.10}$$

- The Mean Absolute Percentage Error (MAPE) measures the accuracy of the forecast. It is defined as:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}|\frac{A_t - F_t}{A_t}| * 100 \tag{4.11}$$

where A represents the observed values, and F the forecasted ones. In this occassion, the MAPE will be presented as a percentage.

Model	MAE	RMSE	MAPE
ARIMA	3.057064	4.518536	10.79683
ETS	5.467870	6.303011	20.59352
TBATS	4.561776	5.590668	17.0143
Naïve	5.342726	6.272032	20.33706

Table 4.1: Predictive power across models

Table 4.1 presents the error evaluation for each method. As it can be observed, the ETS model presents a very high mean average percentage error. This entails that it is not an accurate model, despite being a more sophisticated model. The Naïve method, despite being used as a benchmark, performs relatively well with respect to it. Both the ARIMA and TBATS models perform better than the benchmark model, being the ARIMA model superior in terms of accuracy. Therefore, it can be concluded that the more accurate forecast is the first one.

4.4.2 Residuals examination

Now that the best model for forecasting is chosen, it is vital to check whether the residuals present a random pattern. Residuals can be thought of as the difference between observed data points and the model's fitted values.

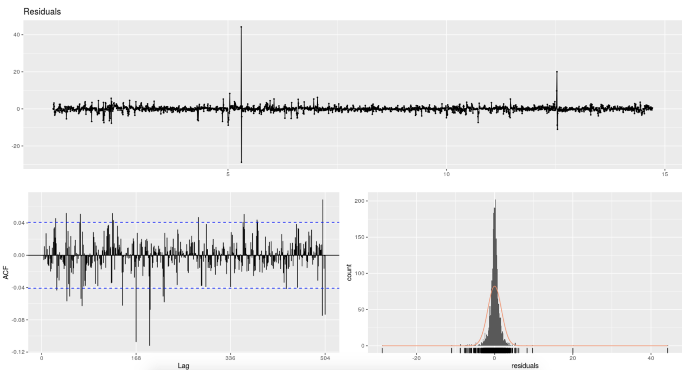


Figure 4.14: Residual analysis

Good residuals should have zero mean and be uncorrelated. The mean of the residuals is equal to -0.001617934, meaning that the forecast is biased. The second condition stems from the need to capture all of the relevant information for the forecast to perform well. *Figure 4.14*, was obtained by using the `checkresiduals` command in R. As it can be observed in the top plot, residuals present no seasonal behaviour. This indicates that the multiple seasonalities included in the ARIMA model (weekly and daily), capture all the seasonal behaviours in the specified time period. By looking at the bottom left plot, it can be seen that the residuals do present some correlation. This entails, as it was explained before, that some information is missing from the model. The series presents a slightly similar distribution to the normal as it can be seen in the bottom right plot. Nevertheless, the right tail is too long.

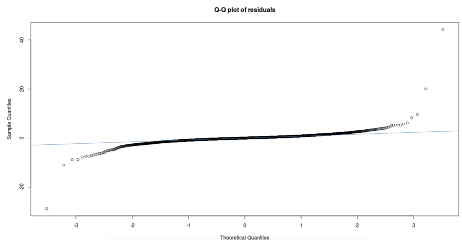


Figure 4.15: qq plot of residuals

Figure 4.15 presents a qq plot that was obtained by using the `qqnorm` and `qqline` commands, and applying them to the residuals. It shows that the residuals do not follow a completely normal behaviour. This could be already noticed in the histogram in *Figure 4.14*. The residuals should form a pseudo-straight line. Nevertheless, they form a slight s-shaped line which means that there are heavy tails. Therefore, it can be concluded that the model can be further improved.

4.4.3 Weaknesses of the analysis

Given the results of the residual analysis, there is still scope to improve the model. More sophisticated models could be used, such as the GARCH model. External regressors could be used to create a dynamic model as well. Temperature data could be included (as electricity demand could be negatively correlated with temperatures, consequently increasing prices), or amount of rainfall (as electricity supply could be positively correlated with this variable, consequently decreasing prices). The aim of these improvements is to decrease the Mean Absolute Percentage Error below 10 per cent - which is considered an “excellent” percentage error, and to obtain totally uncorrelated residuals.

Chapter 5

Conclusion

The role of energy in our daily lives and in the industrial sector will not cease to gain importance in the coming years. Nevertheless, electricity counts on a limitation explained by its nature: "*supply has to equal demand on an instantaneous basis*". Thanks to pooled markets such as the Nord Pool, and organisms such as the BRPs and TSOs, its supply at affordable prices to a wide range of consumers is guaranteed. Nevertheless, extreme weather conditions, the potential existence of market power, or controversial political decisions may cause prices to fluctuate. Price fluctuations took place in a striking manner in the Nord Pool in summer 2006, caused by unexpected plant outages in Sweden, and in 2002, which was a particularly dry year. Prices were expected to rise, but not as much as they did. Nevertheless, the market was resilient and prices stabilised.

The paper has studied these extreme price increases. Extreme prices were determined to be those higher than 60EUR/MWh, after an analysis of the median and 10^{th} and 90^{th} percentile was performed. Further examination of exceedance prices in an hourly, weekly and monthly basis is presented. The paper also statistically analyses the logarithmic returns of prices at different times of the day. The main takeaways are that electricity prices are highly related to daily rhythms and industrial activity. This is explained by the fact that volatility rises between 5am to 8 am, and reaches its lowest point at 1pm. And, average prices are much higher on weekdays than on weekends. Electricity prices also follow a highly skewed distribution, and seem to be non-stationary and conditionally heteroskedastic.

Finally, the paper seeked to find an accurate model in order to predict electricity prices. Four methods were used: ARIMA, ETS (both using LOESS regression), TBATS and Seasonal Naïve method. This last method was included to be used as a benchmark. Those models performing worse than that given by the Naïve method would be rejected. As it was stated before, the TBATS model takes long to process data so in order to guarantee that R would not abnormally crash, 2,880 data points were selected (corresponding to 4 months worth of electricity prices). After assessing their accuracy, the most precise model was the ARIMA model - it presented the lowest MAE, RMSE and MAPE. Residuals of this model were analysed, and the analysis was relatively satisfactory. Residuals presented some correlation and their mean was different from zero. Given the results of the residual analysis, there is scope to improve the models. More sophisticated methods could be used, and external regressions could be included to create a dynamic model (using temperature data or amount of rainfall, which are potentially correlated with electricity prices). The aim of further dynamic analysis and forecasting is to decrease the MAPE below 10 per cent and to obtain uncorrelated residuals.

It is important for governments and producers to have access to forecasting information, and to study exceedances. What are the drivers of high prices on both the supply and demand side, and how often these happen is an invaluable piece of information. This way, power plants can more accurately plan their outputs and engage in fruitful contracts, and consumers can plan accordingly. Furthermore, due to the increase in importance of electricity - which will continue growing - having access to this information could incentivise governments to subsidy prices at certain times, particularly for low-income families or certain industrial sectors. This would be an efficient way of planning and supporting these collectives.

Appendix A

Modelling with the complete data set

The forecasts that will be presented in this Appendix, do contain all the 70,752 data points available. As before, 80 per cent of the available data points, that is 56,600 data points will be part of the training set, while the remaining will make up the test set.

Forecast with ARIMA and ETS

The same procedure as before was followed to obtain the forecast with the ARIMA and ETS methods. In this instance, the value obtained from the LOESS regression is equal to 0.05005558.

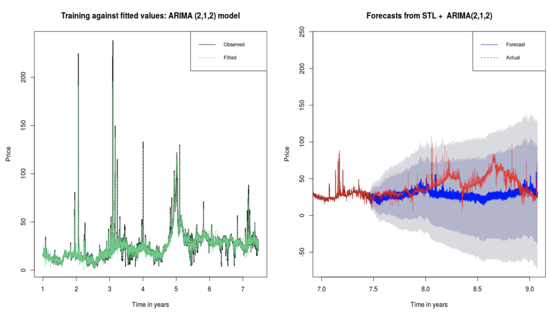


Figure A.1: Training against fitted values using SLT and ARIMA (2,1,2) and Forecast using SLT and ARIMA (2,1,2) against real values

By using the aforementioned `stlf` command, and specifying the daily, weekly, monthly and yearly seasonalities, and indicating the desired length of the forecast (14,150) the following plots were produced. As it can be observed, it seems that the ARIMA forecast is better than the one obtained by ETS, as the latter presents wider confidence intervals.

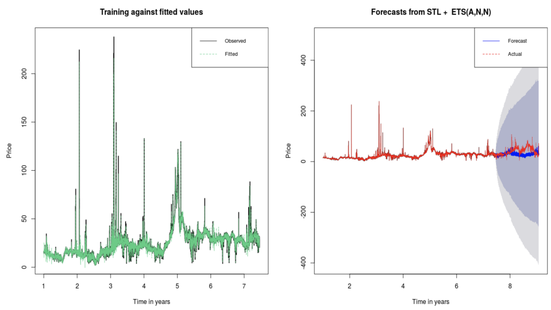


Figure A.2: Training against fitted values using SLT and ETS and Forecast using SLT and ETS against real values

Forecast with TBATS

It is important to highlight that the yearly seasonality has not been included in this model, as the main downside of the TBATS method is that processing is very slow. For that reason, the training set has been shortened to over 20,000 data points.

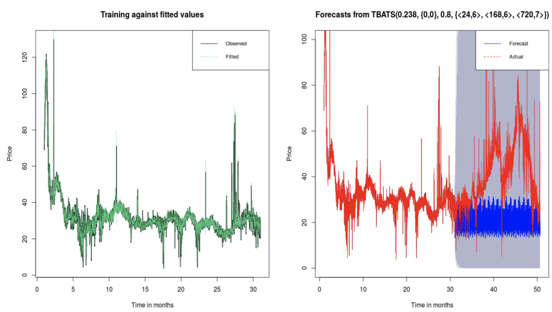


Figure A.3: Fitted values against real values and Forecast against observed values using TBATS method

As it can be observed, the forecast presents wide confidence intervals, and the forecasted values do not seem to be very accurate.

Forecast with seasonal Naïve method

The forecast obtained by the benchmark seasonal Naïve method is plotted below. It is used as benchmark in order to reject those models performing worse than it does.

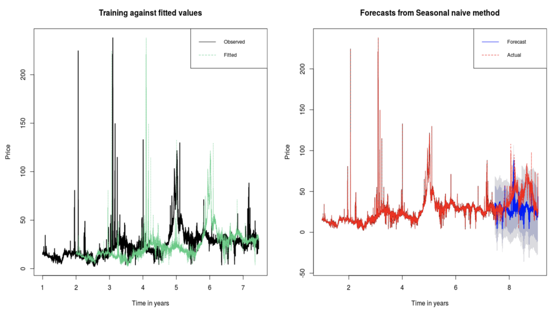


Figure A.4: Training against fitted values using naïve method and Forecast against observed values using naïve method

Evaluation

Table A.1 presents the errors for each model. The ETS model stands out in accuracy, followed by the ARIMA model. Nevertheless, the Mean Absolute Percentage Error is well above 20 per cent, which is the accepted upper bound for MAPE. It can therefore be concluded that there is scope to improve the models.

Model	MAE	RMSE	MAPE
ARIMA	13.92748	18.44300	29.21889
ETS	13.53223	17.95182	28.47161
TBATS	10.90786	13.64158	29.38031
Naïve	14.39790	18.38815	30.81749

Table A.1: Predictive power across models

Bibliography

Coles, Stuart (2001). *An Introduction to Statistical Modeling of Extreme Values.*

Vehviläinen, Iivo and Jussi Keppo (2001). *Managing electricity market price risk.*

Bergman, Lars (2003a). *European Electricity Market Integration: The Nordic Experiences.*

— (2003b). "The Swedish Electricity Market: Current Issues". In:

Wheeler, Donald J. (2004). *Advanced Topics in Statistical Process Control.*

American Bar Association (2005). *Market Power Handbook.*

Conejo, Antonio J. et al. (2005). *Forecasting electricity prices for a day-ahead pool-basedelectric energy market.*

Finansinspektionen (2005). "The Financial Electricty Market". In: *Report 2005:6.*

Tsay, Ruey S. (2005). *Analysis of Financial Time Series.*

Wang, W. et al. (2005). *Testing and modelling autoregressive conditional heteroskedasticity of streamflow processes.*

Lewiner, Colette (2007). *European Energy Markets Observatory (2007).*

Danielsson, Jon (2011). *Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in R and MATLAB.*

Ghil1, M. et al. (2011). *Extreme events: dynamics, statistics and prediction.*

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay (2012). *The Econometrics of Financial Markets.*

Nordic Energy Research (2012). *The Nordics: How much energy does a Nordic citizen use?* URL: https://www.nordicenergy.org/figure/how-much-energy-does-a-nordic-citizen-use/.

International Energy Agency (2013). *When it Comes to Our Need for Electricity, Reliability is Essential*. URL: https://energycentral.com/c/ec/when-it-comes-our-need-electricity-reliability-essential.

Sergey Voronin and Jarmo Partanen (2013). *Price Forecasting in the Day-Ahead Energy Market by an Iterative Method with Separate Normal Price and Price Spike Frameworks*.

Aas, Kjersti and Xeni K. Dimakos (2014). *A Guide to Basic Econometric Techniques*.

Kacapyr, Elia (2014). *A Guide to Basic Econometric Techniques*.

Ford, Clay (2015). *Understanding Q-Q Plots*. URL: https://data.library.virginia.edu/understanding-q-q-plots/.

K A Lindsay and T M Christensen and A S Hurn (2015). *It never rains but it pours: Modelling the persistence of spikes in electricity prices*.

KU Leuven Energy Institute (2015). "The current electricity market design in Europe". In: *EI-FACT SHEET 2015-01*.

Fingrid (2016). *Reserves and balancing power*. URL: https://www.fingrid.fi/en/electricity-market/reserves_and_balancing/.

Institut national de la statistique et des études économiques (2016). *Coefficient of variation*. URL: https://www.insee.fr/en/metadonnees/definition.

Siluyele, Ian and Stanley Jere (2016). *Using Box-Jenkins Models to Forecast Mobile Cellular Subscription*.

Auquan (2017). *Time Series Analysis for Financial Data VI— GARCH model and predicting SPX returns*. URL: https://medium.com/auquan/time-series-analysis-for-finance-arch-garch-models-822f87f1d755.

EnergyNet (2017). *Regulation A: Principles for the electricity market*.

Hyndman, Rob J and George Athanasopoulos (2018). *Forecasting:Principles and Practice*. URL: https://otexts.com/fpp2/accuracy.html#.

International Energy Agency (2018). *Data and statistics*. URL: https://www.iea.org/data-and-statistics.

Chen, James (2019). *Skewness*. URL: https://www.investopedia.com/terms/s/skewness.asp.

Dhankar, Raj S. (2019). *Capital Markets and Investment Decision Making*.

European Commission (2019). *Electricity market design.* URL: https://ec.europa.eu/energy/topics/markets-and-consumers/market-legislation/electricity-market-design_en.

Glowacki Law Firm (2019). "Balance Responsible Parties (BRPs)". In: *European Union Emissions Trading Scheme.*

International Energy Agency (2019). *World Energy Outlook 2019.* URL: https://www.iea.org/reports/world-energy-outlook-2019/electricity#abstract.

Najera, Jorge (2019). *Time Series Forecast using the STL model.* URL: https://rpubs.com/AlgoritmaAcademy/multiseasonality.

Nord Pool (2019). *Nord Pool History.* URL: https://www.nordpoolgroup.com/About-us/History/.

Nordregio (2019). *Nordics top world energy consumption.* URL: https://archive.nordregio.se/en/Metameny/About-Nordregio/Journal-of-Nordregio/Journal-of-Nordregio-2010/Journal-of-Nordregio-no-3-2010/Nordics-top-wor/index.html.

Orhan Altug Karabiber and George Xydis (2019). *Electricity Price Forecasting in the Danish Day-Ahead Market Using the TBATS, ANN and ARIMA.*

Palachy, Shay (2019). *Stationarity in time series analysis.* URL: https://towardsdatascience.com/stationarity-in-time-series-analysis-90c94f27322.

Prastiwi, Ajend (2019). *Forecasting Time Series with Multiple Seasonalities.* URL: https://rpubs.com/AlgoritmaAcademy/multiseasonality.

Rönnback, Jan (2019). "Future Thinking - Empowering the Market". In: *Nord Pool Whitepaper.*

Skorupa, Grzegorz (2019). *Forecasting Time Series with Multiple Seasonalities using TBATS in Python.* URL: https://medium.com/intive-developers/forecasting-time-series-with-multiple-seasonalities-using-tbats-in-python-398a00ac0e8a.

Gentle, James (2020). *Statistical Analysis of Financial Data.*

Kuepper, Justin (2020). *Volatility Definition.* URL: https://www.investopedia.com/terms/v/volatility.asp.

Nord Pool (n.d.). *The Power Market.* URL: https://www.nordpoolgroup.com/the-power-market/.