Comparison of short-term forecast of hourly day-ahead electricity prices in DK1 using different time series and linear regression models

Master Thesis

BUSINESS SCHOOL
AALBORG UNIVERSITY

Aalborg University Business School (AAUBS)

Written by: Giedre Zukauskaite

Supervisor: Douglas Eduardo Turatti

Date for submission of assignment: 31st of May 2022

Title page

Education: **MSc in Finance**

Place of study: **Aalborg University Business School (AAUBS)**

Author: **Giedre Zukauskaite (20182625)**

Title: **Comparison of short-term forecast of hourly day-ahead electricity prices in DK1 using different time series and linear regression models**

Count of characters: **92.332**

Count of written pages: **38,47**

Abstract

This project is based on forecasting study and is useful finding the forecasting model for day-ahead electricity prices in DK1 from choices of different time series models and linear regression model. The topic is specifically useful in electricity trading since it can help to predict values close to the real prices and order capacities on different interconnectors early to earn higher profits from electricity trading. This study is looking in different Autoregressive and Moving Average models that are made using Maximum Likelihood and Akaike Information Criterion to make models and forecast the day-ahead prices. In addition, linear regression model with fossil fuel prices and CO2 emission prices are used to see if that model can predict prices more accurately since power plants are using a lot of fossil fuel to produce electricity. These models are made with three different datasets: Spot prices from 7 AM, 1 PM, and 9PM. Each of these hours are in different load periods throughout the day, which can deliver interesting results regarding different model accuracy throughout different periods. Study delivers forecasts, which indicate that off-load periods one and two are best predicted by using linear regression. However, peak-load period is best predicted by using AR (5) model. These forecasting models can be useful when made for a few neighbouring grids to buy capacities earlier and earn higher profits. Trading strategy suggested to use with the results from this research is technical momentum strategy that would base trades on models and quantitative data analysis and would focus on day-ahead price movements to help earn higher profit.

Table of Contents

1 Introduction **6**

2 Research Area **11**

2.1 Introduction to research area **11**

2.2 Introduction to problem and motivation **12**

3 Methodology **27**

3.1 Study design **27**

3.2 Approach **27**

3.3 Data sources and collection methods **27**

3.4 Methods/Theories **29**

3.5 Validity **39**

3.6 Reliability **40**

4 Analysis **41**

5 Findings **45**

6 Discussion **55**

6.1 Limitations and future research **55**

7 Conclusion **57**

8 Bibliography **58**

9 Appendices **62**

List of figures and tables

Figure 1 - Pricing of electricity (Winther, 2020)

Figure 2 - Grid development in Denmark (Winther, 2020)

Figure 3 - Price Coupling (Winther, 2020)

Figure 4 - Trading in Energy Sector (Winther, 2020)

Figure 5 - Day-Ahead trading process (Winther, 2020)

Figure 6 - Base load and Peak load in electricity prices (Winther, 2020)

Figure 7 - Imbalances (Winther, 2020)

Figure 8 - Price Cap (Winther, 2020)

Figure 9 - Revenues and costs involved in Energy Trading (Winther, 2020)

Figure 10 - Plots of independent factors

Figure 11 - Plots of dependent variables

Figure 12 - Correlations

Figure 13 - Linear regression models and Time series models testing

Figure 14 - RMSE for forecasts

Figure 15 - Naive model forecast comparison with real prices

Figure 16 - Linear model forecast comparison with real prices

Figure 17 - Time series model forecast comparison with real prices for y (data7AM)

Figure 18 – Time series model forecast comparison with real prices from y1 (data1PM)

Figure 19 - Time series model forecast comparison with real prices for y2 (data9PM)

1 Introduction

This chapter will include introduction to the Energy sector and its relevancy.

Part of the trading market will always involve goods that are essential – commodities. Trading these goods usually involves a lot more than just making bids, for example, transportation, storage, and other requirements are necessary to move these goods. Thus, this study is made to get a better understanding about the electricity trading, especially day-ahead electricity trading, and look for predictive models that could help earn higher profits.

Commodities have always been an essential part of the human allowing to make trades in various forms: people exchanged food, clothes, tools, and other necessities to get by and fulfil the essential needs (Mcleod, 2018). Even though the bartering method has evolved over time, especially through the introduction of cash, it remained to exist until this day and is likely to remain. Electricity became part of the commodities when it was started to be commercially used in 19th century (Winther, 2020). In today's world, electricity has deeply integrated into the daily lives of people and continue to grow in significance, as more people gain access to it. Electricity's versatility provides light in buildings, helps to cook, allows prolonging food life, and boosts the communication amongst people by enabling to use various medias and connection tools, such as internet. The continual growth of need of the electricity creates incentives for this industry to grow this quickly, especially considering that more people are improving their lives by acquiring more appliances, electric cars, and other electrically powered things (IEA, 2021). Because of the growing demand for electricity, the infrastructure of the grids that are handling the physical movement of the electricity are growing larger and continue to be improved to meet the changing environment needs. Grid can be described as all the cable systems that connect producers and consumers of electricity in certain regions (Winther, 2020). Often, grids are divided into certain regions, however, remaining connected, to provide a flexibility of managing electricity flow and a greater balance to the system and reducing the possibility of the power loss (Winther, 2020). To further increase the advantages of the connected grids, interconnectors are established joining neighbouring country grids to allow trading power, often leading to influencing the prices (Winther, 2020). For example, electricity grid in Denmark experienced transformation when two grids: DK1 and DK2 were finally connected with the interconnector "Great Belt" in 1998 joining Scandinavian grid and European grid (Hitachi Energy, u.d.).

The need for energy has driven technologies to capture power in various ways. In the 19th century, innovative technologies were introduced to the market as an alternative to fossil fuel plants, such as wind turbines and solar panels, to explore the possibilities of the electricity generation (Winther, 2020). This inspired numerous new engagements from different fields of studies, for example, for research in ecological and environmental studies to take place in order to find a way to generate electricity without using fossil fuels (Winther, 2020). However, since the demand for electricity kept growing due to the recognized ability to be utilized, the supply of these alternative technologies was unprofitable and could not keep up with the increasing pace and scale of the required amount (Winther, 2020). Other factor, for example, the changes in political and ecological fields, continued to affect electricity markets with "Green Transition", which started in 2016 and was initiated by European Commission to help EU become carbon neutral (Energinet, u.d.). The growing evidence of the significance and the importance of the climate change and mounting pressure from the public has burden politicians to react to it and engage changes that were necessary for the sake of the humankind (European Commission, 2021). This has resulted with the initiatives encouraging a transition for more renewable energy generation and issued climate sensitive regulations in effort to combat climate change, such as carbon emission tax to curb the pollution (European Commission, 2021).

The carbon tax has created a new restraint for the parties that are using fossil fuels to generate power, as companies are given emission's allowance for a year, establishing a cap that could not be exceeded, otherwise, companies would face fines (Energinet, u.d.). However, the emission allowance system has also provided a space for flexibility for the companies using fossil fuel energy, and an additional opportunity to earn more money for the participants that have managed to control their emissions leaving with a surplus. The remaining unused allowances could be sold for the other companies that could not keep in their quotas to purchase the additional permission to avoid fines. Hence, CO2 emissions' allowances could be bought by different companies that expect to exceed the limits of their own allowances in order to avoid fines. As a result, these exchanges become another variable that can affect the electricity prices, especially considering the volatility brought by the energy generated from renewable sources, as wind and solar energy production is heavily dependent from the weather conditions (Winther, 2020). With the growing share of the power that renewables produce, weather conditions are able to drastically influence the price, on some occasions, even entering negative price range, as the generating production is exceeding the demand and consumers are paid to offload the grid. However, on the occasions when the wind is sparse and clouds prevents sunrays from reaching solar panels, plants powered by fossil fuels must increase the production to compensate for the missing gap, creating even more volatile environment, especially considering the fact that those plants emits CO2 and uses the emission allowance (Tuomas Rintamäki, 2017). This leads to one of the main challenges that the renewables are currently facing, which is the conundrum of storing the generated power for the later use when the demand is greater or the conditions are unfavourable to produce, or the threat of over-supplying the energy causing the imbalance in the grind and leading to the blackout (Winther, 2020). Unfortunately, no batteries big enough and efficient enough have been created yet, leading to a problem of letting to waste the cheap energy which cannot be stored until demand increases. Hence, as the prices for the emission allowance for the fossil fuel burning power plants continues to grow and the instability of the renewables remains to contribute to the market volatility, the problem became more mainstream attracting more attention, leading for more various researchers to try tackling this challenge (Colli, 2021). As the prices for carbon emission progressed, power plants opted for less polluting fuel alternative: gas (Micucci, 2022). The natural gas gains more popularity after the technological breakthrough of liquefying them and transforming into Liquefied Natural Gas (LNG) allowing for the easier and more flexible transportation (Micucci, 2022). This gave a significant boost for the gas fuels, as LNG gave the flexibility and allowed to be deliver using ships to any part of the world, while conventional gasses have to be transported through the established piping system constraining to the established network.

One of the more popular existing methods how energy can be stored sustainably is by hydro plants, using the power when it is cheap to pump water into the reservoir, and when it is expensive - releasing the water to power the turbines and generate electricity, and selling the production to the buyers at higher price than originally required to fill the tank, profiting from the difference (Winther, 2020). Even through the hydro plants are considered renewables, they require either a substantial amount of rain to collect enough water to fill the storage, or it must reuse the same water, which requires power to move back to reservoir (Winther, 2020). It is also worth to mention that other alternative methods of energy storing were explored, such as: Gravity storage, Compressed air storage, Thermal storage, and Chemical storage (Winther, 2020). For example, gravity storage was built in Nevada to test it and it can be used if grid needs balancing and too much energy is produced. There, a train that is filled with concrete is taken up the 6 km hill using excess power and if energy is needed to balance the grid later, same train can be used to go down the hill and send created energy back to the grid (Winther, 2020). However, this type of storage requires specific terrain, thus, creates difficulties to build (Winther, 2020). Similarly to Gravity storage, Compressed Air storage requires a specific terrain to be materialized, where underground parts of salt formations can be used to store air, however, it did not get very popular (Winther, 2020). Other methods gained more popularity throughout the world that did not demand unique terrain, such as, Thermal storages (Winther, 2020). However, it is becoming more popular to upgrade power plants to reduce the loss of power and re-use the by-products to minimise the loss. This technique utilizes the hot water collected from the electricity plant to be either reused for district heating, like in Denmark, or used to create electricity by running through pipelines back and forth (Winther, 2020). In Denmark specifically, many plants are called Combined Heat and Power plants (CHP) and it works more efficiently compared to the regular Power plants since the secondary product, hot water, is used in the district heating system to deliver heat to households in larger cities, reducing the lost energy (Winther, 2020). One of the more familiar storing methods is Chemical storages, also widely used in all different electronics as it uses Li-ion batteries that store power and can be charged and be provide mobility (Winther, 2020). Even through this method is known for humanity for some time, it is very difficult to build a large industrial battery to store power at large scale. Researchers are still trying to find a solution to capture and store energy at large quantity (Winther, 2020). Nevertheless, the energy storage technologies are exploring various ways to store the power, especially the captured from the renewables, most of the methods are still in early stages of development, and further improvements are required to make commercially feasible and adopted by the market (Winther, 2020).

As described above, the energy market is highly dynamic environment with the ever-evolving technologies to capture the power for later use. However, as the market continues to drastically develop, the variables that were significant long time ago, for example base fuel price, might be losing their significance as new factors disrupt the conventional formula, emission allowance and renewables, thus, reducing the accuracy of the models developed long time ago. In addition, the volatility of the renewable energy continues to contribute to the challenges of generating an accurate forecast causing the uncertainty and potential imbalance. Since forecasting modelling was known for some time in the financial markets, players form commodity market recognised the potential of it and began the investigation, resulting in a growth in recent studies based on topics to find the model that can forecast electricity prices and make the best possible forecast using historical data (Jannik Schütz Roungkvist, 2020); (Sergey Voronin, 2013); (Lindstrom, 2021). Because the significance of the electricity continues to grow, the market continues to develop and change, especially with the “Green transition” taking place, simple models that were used in the industry early days on can lose their power to predict day-ahead prices opening doors in finding a new model that could perform better. Thus, this thesis is based on the research question introduced in the next chapter.

2 Research Area

This chapter will include introduction to research area, problem, and motivation to the project as well as introduction to the market and trading in the energy sector.

2.1 Introduction to research area

The purpose of this study is to compare a short-time forecasting models: different time series models, linear regression model, and naïve model with factors that have predictable power over the day-ahead electricity price level using historical data in DK1 grid as presented below. This research will try to create a predictive model that will be able to naïve outperform models when predicting day-ahead hourly electricity prices.

As mentioned in introduction, energy market is rapidly changing and introducing new variables, and new theories and models can gain more predictive power compared to the older ones that were created and adapted prior the change of significance of the variables. Power market is sharing some similarities with the financial markets, where different trends and regulations are shaping the trading patterns and new ways of finding predictive power is getting increasingly difficult. Hence, forecasting combination of different models in the same strategy could help to deal with a highly unpredictable market and adjust model's weights to give it more flexibility.

However, since industry have been using the same predictive models for a long time, and the growing influence of new factors, this have resulted in a weakened predictive power of the models resulting in less accurate forecasts and misses on potential profit. Therefore, this research is looking to find out *How can day-ahead hourly day-ahead electricity prices in DK1 be more accurately predicted using other models than the naïve model? Based on comparative forecasting study using linear regression and different time series models.*

To get a better understanding about the DK1 electricity market, one of the sub-questions is going to be used to help to find factors that have predictive power over the change in electricity price level: *Which factors have predictive power over the electricity price levels changes in DK1?*

In addition, second sub-question is going to be used to find out if one of the models: AR, MA or multiple linear regression can outperform naive model: *Can one of the models used in the study outperform naive model?*

Main aim of this study is to find out if any of the models used in the study can perform better than naive model and help to earn better profit when trading electricity in day-ahead market. This research will contribute with finding models or forecasting strategy that have higher accuracy when predicting hourly day-ahead prices to outperform competitors in the industry. Even though electricity trading market is quite different from the financial markets, models used in predicting financial asset prices can be tested in this industry as well (Joscha Schabram, 2021). However, it is important to note that models in asset forecasting are usually made to forecast returns, while in electricity trading returns are usually harder to calculate since traders must pay for capacity of using interconnectors and other fees before knowing how much they earned, which is explained more detailed below. Another important fact is that if trades were not completed properly and any imbalance was made in the markets, then traders can get fines and that would also affect profit negatively. Thus, in the electricity trading, forecasting is usually done to find day-ahead prices, so that the trading capacities can be bought ahead of time and deliver higher earnings. Hence, this study is looking for the model that would be useful when predicting day-ahead prices and benefit traders when making trading strategies.

2.2 Introduction to problem and motivation

For better understanding of the research topic, decisions regarding data and modelling, and reasoning of the arguments, it is necessary to present context of the energy market first and lay a foundation on which the study will be built on.

To start with, the short time forecast means that forecasting is completed for a relatively short period of time. Since values for one specific hour from each day are collected in one datasheet, 316 future values are going to be forecasted to find out which model delivers most accurate predictions and make a forecast for that specific hour to compare it with the testing sample.

To get a better understanding of the factors with predictable power for the day-ahead electricity price change, previous studies must be reviewed and analysed. Some scholars have indicated for the demand and supply to be two main factors that have predictive power over day-ahead electricity prices (Monforte, 2000); (Griffin). However, other conducted studies focusing solely on finding the significant factors that can help to predict day-ahead electricity prices also indicated Seasonality as one of the main factors to have an effect on the prices (Alvaro Escribano, 2002); (C.R. Knittel, 2005); (Takashi Kanamura, 2007). In certain countries, high changes of outside temperature during the year can influence electricity prices, as power can be used for warming or cooling buildings (Andrew Henley, 1996). However, since in Denmark temperatures do not deviate that much to extreme heat or cold, and considering the fact that country also have widely integrated CHP plants in larger cities to make electricity production more efficient when warming is needed, prices are not that much affected specifically by the changes in temperature as much as in other countries (Winther, 2020). Furthermore, multiple various studies that were conducted on finding the statistically significant factors that have predictive power over the electricity price level change indicated several main factors with predictive power were indicated as: consumption, production, Interconnectors with neighbouring grids, CO2 emission price, Fossil fuel prices: Coal, Oil, and Natural Gas, and finally, daily seasonality (Carlos Fernández Alvarez, 2021). All in all, there can be many different factors affecting the day-ahead electricity prices, however, this study will include only the prices of the fossil fuel since it affects electricity prices directly as a cost in production.

Power trading can take place in three different markets: Futures, Day-ahead, and Intraday (Winther, 2020). Futures market is used to trade Futures related to energy production, for example, fossil fuels used in the industry (Winther, 2020). Day-ahead market is used to set prices for each hour on next day and usually can be used to trade energy to balance consumption and production (Winther, 2020). Intraday trading is usually used to make last trades to balance out the grids and make sure that consumption and production matches in each (Winther, 2020).

Day at the Electricity trading floor

Electricity trading happens 24-hours a day, meaning that 3 different shifts are required to cover all the trading. Day shift starts around 7 and includes a catch-up with the night shift traders to hear about the situation in the market. Especially, if any issues or downtime on certain grids, interconnectors or plants occurred since it can affect the trading during the day. Then, from 7:30 bid preparations are started to be prepared using the different data regarding the weather, fuel prices, consumption, and production plans to have a biding strategy ready before noon. Afterwards, all the news regarding the situation in the market can affect the strategy up to 12:00, meaning that day-ahead traders must be ready to change the plans and trades if needed. At 11:58 traders submit all the offers and go for lunch since now TSOs must approve all the bids. When traders come back from the lunch, they find results – if their offers were accepted or not. Around 14:15 meteorologist team delivers newest weather forecast, which usually means that any changes must be reviewed by intraday traders, who must balance the market and make more trades to do that. Just before leaving the day shift, around 15:45, traders participate in bidding with brokers. The rest of the time in between usually are taken by reading news, changes in the markets and learning more about each market specifics (Winther, 2020). There are also evening shift and night shift in the energy trading. Night shifts are currently moved to the traders in Singapore, meaning that local traders in Europe does not have to work throughout the night and their colleagues can make sure that grids are balanced throughout the nights.

Electricity Day-ahead prices

Day-ahead hourly electricity prices are always presented to the market at 12:00 for the following day, where each hour has a calculated forecast of the prices set for each hour (Winther, 2020).

However, this market is exposed to more various external factors that often are beyond of control which can affect the price stability (Winther, 2020). Comparing to Intraday market, Day-ahead traders must also account for the fuel prices that are required to power energy production plants and planning the production to the anticipated consumption of the following day (Winther, 2020). However, when a significant portion of power generation comes from the renewables, Day-ahead traders must also be aware of the additional variables that could influence the production capacities, such as wind speed, precipitation, and sunny weather (Winther, 2020). Furthermore, the price coupling with the neighbouring grids is another important factor to be accounted for when looking in the next 24 hours' prices (Winther, 2020). This is important because prices in neighbouring grids can affect prices in DK1 grid by importation or exportation of electricity. While Day-ahead market creates a "benchmark" for the next day, Intraday market plays an important role in ensuring the power supply up to 15 minutes before the energy is consumed to ensure that grid is kept balanced and demand is completely fulfilled (Winther, 2020). These imbalances can be affected by the increase or decrease of consumption or the production, changes in weather, or any technical problems arisen in the grid or power plants.

General principal for calculating the price for the electricity is illustrated in figure 1. This graph indicates that consumption must match with the production and power from cheapest production facility is always used first to make sure that electricity prices are the lowest possible. This is similar to the goods market where the prices are determined by the supply and demand (Winther, 2020). It is important to highlight the fact that the cheapest electricity always comes first, and in the scenarios where it is not possibly to supply, for example reduced production due to change of whether conditions or technical issues, the next more expensive option is chosen. Finally, when the equilibrium between supply demand is established, the price for MWh can be found with the red slope illustrated in figure 1. This slope can move horizontally depending on the demand for electricity prices. First price 'step' always represents the cheapest electricity supplier that is able to fulfil the base need, and the left demand is filled with other 'steps' which are more expensive options depending on the fuel nature (Winther, 2020).

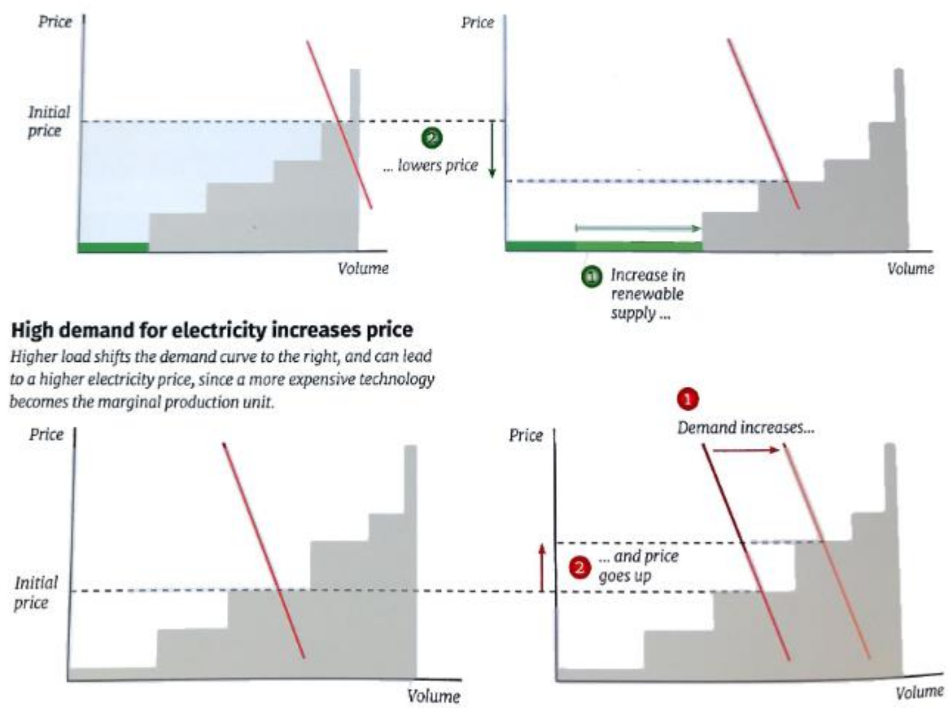


Figure 1 - Pricing of electricity (Winther, 2020)

Even though the pricing strategy is relatively straightforward, it has interesting dynamics. If the higher price of production, for example, coal, is setting the electricity prices, then slight increase in the production methods in the steps below that step, are not affecting the prices at all. However, if the prices increase to the level above coal production, then steps switches and now the new production methods set prices instead of coal.

DK1 grid is a part of Denmark's national grid system joint together with DK2 grid. DK1 grid includes western part if Denmark, while DK2 grid mainly covers the Sjælland region and around the capital of Denmark. The entire Danish grid is managed by Energinet, which is state owned company, and is responsible for maintaining and taking care of the gird itself, as well as for balancing of it 24/7. Denmark's grid is part of European grid and it is connected with neighbouring grids, joining Scandinavian with European markets and can be illustrated in figure 2 (Energinet, u.d.).

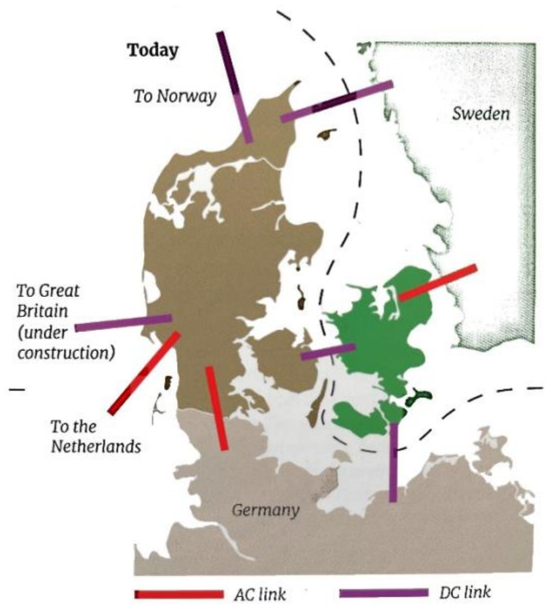


Figure 2 - Grid development in Denmark (Winther, 2020)

New technologies and grid enlargements are allowing for better grid balancing between different countries, as the unification of several grids brings additional security of energy supply despite any transportation or production disruption. This also creates a phenomenon called as "price coupling", where prices between different grids is to some extent similar (Winther, 2020). If the conditions allow for one country to produce abundant amount of electricity from renewables, they can sell certain amount for the neighbouring countries if the generation for buyers using fossil fuels is more expensive than importing. Since neighbouring grids are connected using the interconnector, possibility to share cheaper power helps to lower prices in the neighbouring grid. However, since the demand is increasing within the country that had cheaper power originally, demand is increasing and getting closer to an average price between the two sharing grids, so it affects prices in both grids sharing power.

This results in the situation where two connected grids would have equal or very similar electricity prices, also known as ‘price coupling’, illustrated in figure 3.

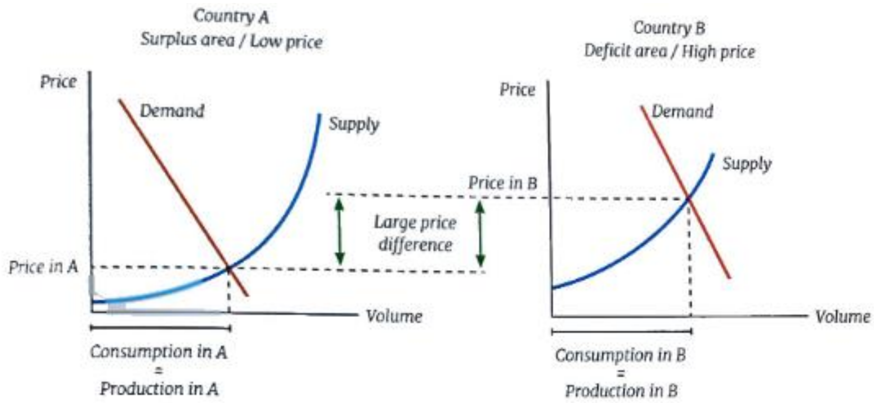


Figure 3 - Price Coupling (Winther, 2020)

Electricity prices measurements are calculated for the megawatt hour (MWh), which is equal to 1000 kilowatt hours (Kwh). This measure means that price is calculated for electricity that is produced in one hour and can be transferred to the consumers (Winther, 2020). As some of the plants are not very efficient and a certain amount of power is lost while being transported via transmission lines of the grid, it was important to establish a way to standardize size of unit that is traded – which was chosen to be MWh (Winther, 2020). However, it is important to note that some of the power plants are less efficient than others, for example, power plants that uses fossil fuel have efficiency of only around 33%, while CHP plants sell not only power that they produce, but also the by-product, hot water, for heating local buildings making them around 66% efficient (Winther, 2020). The 33% efficiency translates into that 3 units of fuel are used to generate 1 unit of power, meaning that 2/3 are wasted energy (Winther, 2020). Researchers are looking for ways for increasing the efficiency of power plants to make sure that used fossil fuel at the plants are utilized as much as possible, considering the fact that fossil fuels are finite resource bound to ran out eventually (Winther, 2020).

However, compared to renewable power production, the efficiency is significantly higher with the exception on that the weather conditions must be right for the sites to produce energy. For instance, for wind turbines, one of the limitations is that they can be used only up to 40 km/h wind speed since this is the limit of production capabilities for the technology right now (Winther, 2020). The faster the blades spins, the more power is produced and can be sold. Though, if the wind speed exceeds the capabilities of the wind turbine, turbines must be shut off to protect the motor shaft from the damage, and wind blades from breaking (Winther, 2020). To overcome the weak wind, turbine blades can be increased to capture more wind through enlarging the rotation diameter and ensuring that the motor is generating power, resulting in spread of the bigger windmills (Winther, 2020). Solar panels also require specific weather conditions to be able to produce power – sunny hours and, also, clean solar panels to make sure that as much power is produced as possible (Winther, 2020).

One of the main issues with using the renewables is that on some occasions it might produce more power that there is a demand for (Winther, 2020). Because of that, owners of wind parks might get paid by the TSO to stop the turbines and not produce power just because that would imbalance the market and create a potential blackout (Winther, 2020). Another challenge arises when renewables produce more electricity than there is a demand for and pushes prices into the negative range, where consumers are paid to consume electricity and offload the grid (Winther, 2020). Currently, market is also structured in a way that investments to build wind parks are beneficial since investors can get paid for investing in renewables, and even though they might not receive that much profit for MWh of produced electricity, grants from government can help to earn profit in the long run (Winther, 2020).

When looking at DK1 grid, it is connected with Norway, where over 91% of produced electricity comes from hydro pants (Statista, u.d.). This means, that when the wind power generation in DK1 is weak, the electricity might be imported from the neighbour grid to secure low prices. Hydro plants are also considered to be renewable way to produce electricity, however, it requires rain to fill up to reservoir, so that turbines can produce power (Winther, 2020). Moreover, it is also worth to mention that France and Finland also generate power through nuclear plants (World Nuclear News, 2022). This method can also be considered as a green way to produce electricity without CO2 emissions, however, many countries are against building of these plants due to the fear of nuclear meltdown and problems with disposing radioactive waste (Winther, 2020).

As more people are concerned over climate change, various initiatives are being taken, such as Green transition, where European Union aims to transform energy sector to be more modern, resource-efficient, and sustainable (European Commission, 2021); (Unicef, 2022); (European Commission, u.d.).

Energy producers have adapted and found another way to earn money – through selling Green certificates, which claims that all the consumption that the company is using per day/week or year is Green energy produced without CO2 emission. This not only provides additional value to the consumers to be able to present that their businesses are environment conscious, but also gives a possibility to earn more profit from the investors that are investing in green power plants. As the renewable energy has additional option of generating profit, it is more appealing for the investors to take part in and can be expected to grow in popularity since the European Union is planning to continue heavily invest in energy sector transformation (European Commission, u.d.). This can expand the demand for the green energy, creating incentives for technological upscaling and lowering the overall costs of manufacturing and development, helping to move countries from fossil fuel to a greener future.

Electricity trading strategies are very much affected by the market itself. Since this commodity is different from all other commodities, because electricity must be consumed at the same time it is produced, making the trading strategies differ from the ones used in traditional financial markets (Investopedia, u.d.). The main difference from the traders' side is that while trading in electricity market, supply and demand must be matched and balanced for it to work, where in financial market that is not something that traders must do. Additionally, this market is relatively restricted for less experienced people in trading because of the complicated nature of the market and strict accesses to the trading platforms, which requires of agreement between TSOs and BRP or trader (European Commision, u.d.). Power trading differentiates form financial market because it can take on three different levels trading the same thing – power, as it is illustrated in the figure 4.

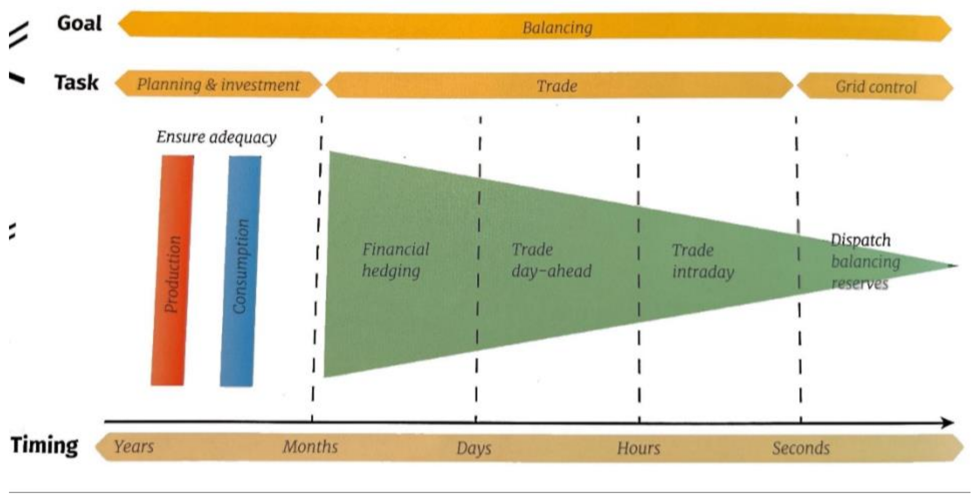


Figure 4 - Trading in Energy Sector (Winther, 2020)

As described earlier, Futures market is used to hedge positions with the fossil fuel and other resources used in the industry, as it takes time for fuels to be bought and transported (Winther, 2020). Then, day-ahead market is used to set prices for the 24 hours next day and balance market with the demand and supply (Winther, 2020). This market is mainly used to hedge high volatility in next day's prices, since different weather conditions and downtime in power plants can affect prices.

Another important aspect to be aware of when trading power is the capacities of the interconnectors. The capacities must be purchased in advance to reserve the ability to transport the power form one grid to the other. Hence, if consumers know how much electricity they will need in every hour next day, they have ability to place orders with the price they are willing to pay per MWh, which is usually lower than the ask price from the producers' side. Then traders can match these demand orders with plants that can deliver needed power with the lowest price spread, meaning that ask and bid prices are very close together. This whole process is illustrated and explained in figure 5.

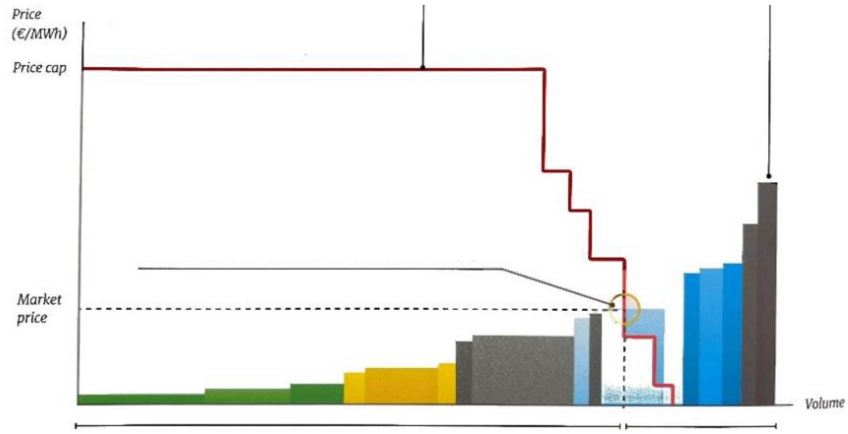


Figure 5 - Day-Ahead trading process (Winther, 2020)

Usually, day-ahead trading is helpful to set baseload of electricity needed for the next day – a set amount of electricity needed in each hour of the day that is equal for all hours and is illustrated in blue in figure 6.

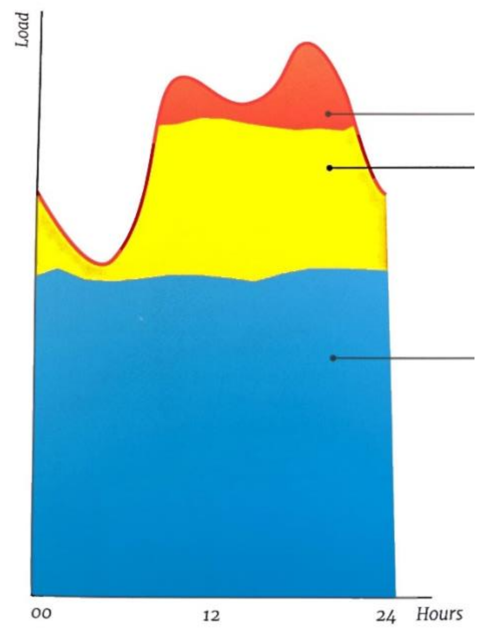


Figure 6 - Base load (blue) and Peak load (red) in electricity prices (Winther, 2020)

Intraday market is designated to keep grids balanced and is traded in real time, meaning that the electricity sometimes must be delivered in 15 minutes (Winther, 2020). Trading in the power market is requiring to be aware of the established rules and work with local Transmission System Operators (TSOs) to be able to send trades and make energy transfers maintaining balanced grid.

Trading companies usually become Balancing Responsible Parties (BRP) in the grids that they want to participate in. This also allows for the trading companies to take producers and consumers as customers and represent them in the trading market trying to sell or buy electricity for the best possible prices. Since electricity market have numerous requirements, there are few stages of balancing before delivery of the electricity to make sure that market participants are able to balance the grid out (Weron R. , 2006). Market order is a way for traders to balance the grid and either take long or short positions in either market, depending on the situation in the grid. Trading can be done in two different forms of products: physical and financial. Trading physical products means that traders are buying real electricity and are responsible for moving it throughout the grid to the customer on time. Financial product means that traders are placing bids, then the electricity is moved by the local TSO without traders needing to do anything else. Each kind of trading has its own advantages and disadvantages. Firstly, physical power takes more planning to execute trade, and TSOs are put in position of uncertainty if traders can balance the grid properly, since they are still responsible to keep grid usable. However, financial products are giving more security to TSOs and some of the countries have preference of selling financial products over the physical ones to keep grid control in TSOs hands instead of different traders. In scenarios where traders create imbalances in the gird and companies are given fine, it causes many problems for the entire grid and can be complex of restoring balance again, so TSOs prefers more security and control as they bear the responsibility of fixing it. Nevertheless, it is also common that the closer trading gets to delivery date, the more physical products are traded instead of the financial ones.

However, sometimes capacity maintenances can happen unexpectedly and cause imbalances, meaning that someone must cover additional costs needed to balance the grid and keep it stable (Winther, 2020). If interconnector or part of the grid gets unusable, then TSOs will pay the compensations to all sides – seller, buyer, and trader. However, if it is caused by the power plant malfunction, then either the producer or BRP are going to be charged for this imbalance. Imbalances are illustrated in figure 7.

Figure 7 - Imbalances (Winther, 2020)

As can be seen in the figure above, depending on the imbalances traders must take long or short position to balance the market out. This task is usually completed by Intraday traders to make sure that no imbalances are left before delivering the electricity to consumers.

When looking closer at the trading itself, capacity purchasing is a huge part of trading electricity. As mentioned previously, traders must buy the capacity to flow energy through the interconnectors and grids, and since it is limited to have maximum capacity, it might create a situation where trader could earn money simply by having the pre-ordered access when the capacity is reached, and no more power can flow preventing any other trades. Interesting fact is that when companies are buying the capacity, they are buying the right to flow that much power, but it is not an obligation. However, when traders nominate flow to the capacity platforms, that's when right becomes obligation, meaning that nominations must be flown through the interconnector, or they might get a fine. Capacities can be bought as yearly, quarterly, monthly, daily, or hourly flows, however, daily and hourly flows are mostly used since companies cannot really predict long term electricity prices that well to be able to reserve the capacity for the whole year. Capacity purchasing is divided into three stages: long term capacity can be bought in auction, and it usually makes around 30% of all possible capacity, which is usually called yearly capacity. Then, monthly capacity takes another 30%. The third stage is capacity sold day before the deliver and includes all the unsold capacity plus the other 30%. If any capacities are still left after these auctions, they are sold again for Intraday traders.

Trading in Electricity markets can be further divided into auction and continuous trading. Auction trading happens on the set timeslots, where all market participants can place their bids for the next day for example, and then bids are matched. Continuous trading is always running and this can include algorithmic trading since bids are placed and matched in real time looking for the best prices.

Trading itself can be done through two main channels: through Exchange or Over the Counter (OTC) trading (Weron R. , 2006). Exchange provides limited flexibility but are well built and are working by the set rules. However, OTC market provides more flexibility, but the deals might be agreed on by email or by phone call making it less secure and not very transparent. Due to regulations and strict rules in electricity trading, most of the company's trades are usually done using Exchanges to keep trades transparent, however, great deals can also be completed using OTC and all the traders have an option to use it when trading electricity.

The purpose of all the traders is in finding the best deals and earning the highest possible profit, disregarding the market in which they are trading in. This can lead to exploitative strategies speculating prices and reflecting in the prices of the electricity. To prevent this from happening, throughout the year's governmental organizations have developed various regulations in effort to manage the price variations at least to some degree. This can be shown as one of the main differences between trading in financial markets and in commodities, in this case, electricity market.

One of the main rules that all traders in energy trading companies must follow is the set price cap for the electricity's price per hour. This is illustrated in the figure 8.

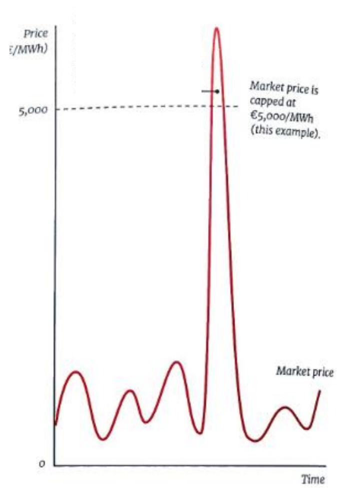


Figure 8 - Price Cap (Winther, 2020)

As presented in the figure above, price cap serves for the end users, since they are to prevent prices from skyrocketing and to protect from paying more than the cap is for one MWh of electricity (Winther, 2020). However, it can also be seen that producers and traders are affected by this regulation negatively and they are losing possible income that could deliver better profits. Other regulations are mostly affecting electricity producers or TSOs, resulting in TSOs being more careful about the trades happening in their markets, and more cautious about the suspicious trades that could cause potential imbalances leading to the fines or potential investigations. All these changes in the regulations ensuring the stability of the grid are affecting the participants in the energy trading business. Companies are required to create department or a person for compliance to make sure, that company is following and checking all the trades before TSOs in order to prevent unwanted issues. Throughout the years, many different cases involved energy trading companies that were sued, and proven guilty had to pay millions for suspicious or unlawful activities. However, if traders are using data and models to help them trade, evidence must be presented of model and tools used in the trading process to prove the innocence and avoid any lawsuit involvements. These additional requirements have limited traders to avoid speculations or any bias based decisions, as this is difficult or cannot be proved, and might get traders in trouble. Hence, more and more tools and models are used in this industry to not only automate tasks, but also to have proof of the trading decision origin.

When talking about profiting from trading in this market, it is important to be aware that it highly depends on the market players position in the market. Producers generally get profit when paid for electricity. Energy trading companies purchase power from the producers in the regions where electricity is produced cheaply and sell it in the regions where the price is higher than original. This also includes covering the transportation cost and profiting from the difference.

When looking in the profits, costs are also important to be aware of. Approximations of revenues and costs are presented in figure 9.

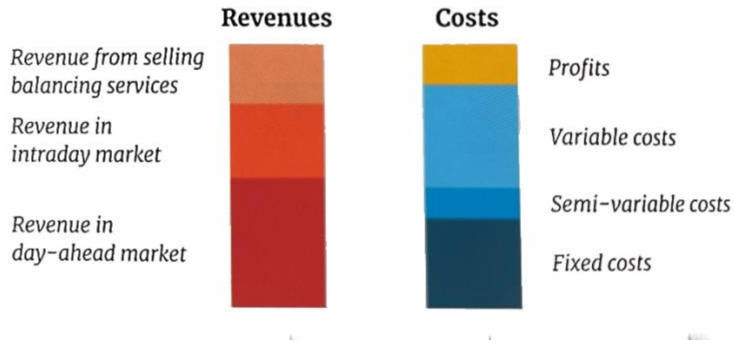


Figure 9 - Revenues and costs involved in Energy Trading (Winther, 2020)

As it is illustrated in the figure 9 above, most of the revenues for trading companies are coming from day-ahead trading, whiles intraday is in the second place, and selling/balancing services are in the third places (Winther, 2020). Due to the ability to earn most of the profit in the day-ahead market, it is chosen to base this study on this specific market and try to find a model that can perform better than naïve strategy. However, it is also important to recognize the main costs that trading companies must pay when they earn revenue. As presented in the figure 9, only a small percentage of the revenues are becoming profit since trading companies must pay fixed costs, for instance participation in the Exchanges fees and, also, pay for capacities to flow electricity, which is part of variable costs (Winther, 2020). Thus, it is important to be able to forecast electricity prices for the next day to get an ability to buy needed capacities cheaper, because the prices of the capacity is increasing nearing the delivery date.

The main motivation for this project is to see if there is a possibility to find predictive model that could forecast hourly day-ahead electricity prices better to outperform naïve model and predict day-ahead electricity prices accurately. This would be beneficial to use when trading to have an idea of at what level prices can be expected to be next day in each power grid and then buy rights to the capacities on the interconnector for cheaper prices as well as plan the flows beforehand making trading more efficient and easier to earn profit.

3 Methodology

This chapter will include study design, approach, data collection, and finally methods used in analysis.

3.1 Study design

This study is based on Comparative Quantitative study design, which is useful when results from quantitative data analysis are compared and was used in similar type of studies by other researchers as well (Xin Liu, 2019); (Claudia Eliane da Matta, 2021). This approach is useful to compare forecasts and indicate the ones that can outperform naive model and find the best forecasting model for hourly day-ahead electricity prices from different time series models, and linear regression strategies. Because of that, Quantitative study using Econometrics and different financial forecasting models with the historical data can be used to compare the results from the forecasts and help to earn profit when trading in day-ahead electricity market. Since the goal of the study is to forecast prices, and then compare the results to find out which model performs most accurately, historical quantitative data is used extensively to be able to compare different models and conclude the study with the best performing model as well as trading strategy suggestions.

3.2 Approach

In this study deductive approach is used to analyse historical data with different forecasting models to compare forecasts and discover most accurate model to forecast hourly day-ahead prices (Kuada, 2012). Deductive approach is described as: "*application of the systematic steps of the scientific method, while using quantitative properties to research relationship or effect of specific variable*" in (W. Alex Edmonds, 2017, s. 30). This approach is useful when starting with data and examples from previous studies and can be used to compare and improve forecasting models already existing in the market (Kuada, 2012). Hence, data sources and collection methods are described below before looking in the methods and theories used in the study.

3.3 Data sources and collection methods

To carry out this research, data used in this research is Quantitative historical data for the four factors that are believed to have effect on day-ahead electricity prices due to its importance in the process of producing electricity. The prices for fossil fuel are included in the forecasting: Coal, Oil and Natural Gas. In addition, price for CO2 emissions can also be used since it influences prices of electricity and adds up to the production cost. As mentioned previously, paying for CO2 emissions was a part of 2016 Green Transition regulations package and lead to increase in production prices (European Commission, 2021). Even though fossil fuel prices are provided as daily prices in the Investing.com, they were set as prices for that specific hour of that day (Ivesting.com, u.d.). All the data regarding prices of fossil fuel and CO2 is downloaded from Investing.com and are presented in USD (Ivesting.com, u.d.). Dependable variable in this case is Spot Price and it will be downloaded from Energinet's portal and is in EUR (Energinet, u.d.).

All the factors and Spot prices are included in the datasets in a way allowing to make future prediction on previous day's prices. Hence, to forecast the prices for 7AM, prices from the day before at 7AM are used. Meaning that seasonality, which is usually present throughout the day during the Off-peak and On-peak periods are not relevant in this study.

Data is downloaded for a period from 2020-01-01 until 2022-02-28. Despite timeframe being more than two years, the dataset becomes extensive and is able to provide a sufficient information to make a short time forecast. Also, to test the models, only three hours are chosen from each day: 7AM, 1PM, and 9PM. These hours are chosen specifically since days in European trading market are divided in three different load periods: Off-peak 1 (01:00-08:00), On-peak (09:00-20:00), and Off-peak 2 (21:00-24:00). Each hour for this study is chosen from different load periods and helps to not only predict prices, but also find out if any model works better in some of the periods compared to others. Separate dataset is made for each hour including Spot prices and different factors. The chosen hours are also chosen to be in different placements during the periods: 7AM is the second to last hour included in Off-peak 1, 1PM is around the middle of the On-peak period, specifically, in the period where most of the people are at work or school, and 9PM is in the beginning of the Off-peak 2 interval. Different placements in the intervals can help to notice how forecasting models differ between two Off-peak periods and compare it even more.

However, it is important to note that electricity market is very structured and if TSOs or BRPs are aware of huge increase in consumption and production in the next hour, they have a possibility to start moving electricity 15-30 min before the hour starts. This allows more security to the grid, making sure that frequency is stable and sudden changes in the grid would not cause blackout.

3.4 Methods/Theories

All the used tests and theories are presented and explained in the chapters below.

In-Sample and Out-Of-Sample techniques

Out-of-sample technique is widely used in the econometrics and refers to a method of splitting the dataset to training sample and testing sample (Wooldridge, 2019). Researchers can choose how to split the dataset, however, usually training sample takes more than 50% and testing sample takes the rest of the dataset. In this study training sample represents 60% of dataset (474values) and 40% of the dataset is used as testing sample (316 values). This is useful since training sample helps to create the model, while testing sample is used to test the model and make a forecast. Also, to make the model more accurate, refitting technique is used to after each forecast is made to include new forecasted value in the model and make a new forecast, which should increase the accuracy of the model.

Linear Regression

One of the most popular models used in predicting prices is linear regression (Fox, 2016). This model is used when data is distributed in a way that line could be created and drawn through the plot of the dataset. The expression of linear regression is presented in the line 1.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (1)$$

The most common linear regression is Least-Squares regression, and it must follow OLS assumptions. It helps to find the most fitting line, where residual squares are used to find smallest deviations from the line and draw the line in the best possible place, so that data fits it well.

Other researchers also included day-ahead electricity forecasting using linear regression. Jannik Schütz Roungkvist, Peter Enevoldsen and George Xydis (2020) used linear regression as main model in the studies when trying to forecast day-ahead electricity prices for the Danish power market (Jannik Schütz Roungkvist, 2020). Even though they used different independent variables, the model still was used as a simple solution to forecast the prices without including more difficult techniques. Another study that included linear regression was conducted by Nektaria Karakatsani and Derek Bunn (2008) (Nektaria V. Karakatsani, 2008).

General-to-Specific

Before starting forecasts, linear regression model must be created and tested to see if all the factors included in the model have predictive power over day-ahead electricity price in DK1. For this, General-to-specific modelling can be used. It starts with general model with all the four factors introduced previously. Then, individual coefficients are tested, and model is re-estimated if needed. This process is continued until the best fitting model is found. So, the starting model is presented in line 2.

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon_i \qquad (2)$$

Where Y_i is Spot price, X_1 is Oil price, X_2 is Gas price, X_3 is Coal price, and X_4 is CO2 emissions price from the day before. This General-to-Specific modelling is used only in the training phase and not forecasting itself.

OLS assumptions for linear regression

There are seven OLS assumptions that must be accounted for when working with this model.

1. Regression is linear in parameters
2. Error terms have zero population mean
3. Error terms are not correlated with Xs
4. No serial correlations
5. No heteroscedasticity
6. No perfect multicollinearity
7. Error terms are normally distributed

It is important that data follows all these requirements for the model to show expected results.

Normal distribution

The normal distribution is symmetric around the mean and its graph looks like a bell curve (Svetlozar T. Rachev, 2007). This distribution can also be referred to as Gaussian distribution and in this case, test used to identify if data is normally distributed is Jarque-Bera test from the *tseries* R-package. Jarque-Bera test is calculated using formula presented in the line 3.

$$JB = \left[\frac{n-k+1}{6}\right] * [S^2 + (0{,}25 * (C-3)^2)] \qquad (3)$$

Where n is the number of observations in the sample, k is the number of regressors (k=1 if not used in the context of regression), S is the sample skewness, and C is the sample kurtosis (Svetlozar T. Rachev, 2007). This test follows asymptotically Chi-squared distribution with two degrees of freedom and its null hypothesis states that normal distribution is present, while alternative hypothesis states that there is no normal distribution and are presented in lines 4 and 5.

$$H_0: Skewness\ and\ kurtosis\ match\ a\ normal\ distribution \quad (4)$$

$$H_A: Skewness\ and\ kurtosis\ does\ not\ match\ a\ normal\ distribution. \quad (5)$$

If p-value is higher than the 0,05, the null hypothesis is not rejected, and therefore the dataset follows normal distribution. If p-value is lower than the 0.05 significance level, the null hypothesis is rejected, hence, the dataset is not normally distributed (Svetlozar T. Rachev, 2007).

Correlations

Correlation between the values can be distinguished to be positive, negative, or have zero correlation. When correlation is positive, it indicates that variables are moving to the same direction depending on each other, while negative correlation means that they are moving in opposite directions (Fox, 2016). Zero correlation means that no linear relationship is found between the two variables. Since linear regression model is used in this forecasting study, it is important to check for linear relations between the factors. To test the correlation between all the variables included in modelling heat map correlation matrix is used in the study. The chart is made using *corrplot* R Studio package and can visualise the correlations between variables well. However, serial correlation is another important factor to look at.

Breusch-Godfrey test

To find serial correlations between the variables, the Breusch-Godfrey test is used. In this case, it is used as part of multiple regression model testing. The hypotheses for this test are presented in lines 6 and 7.

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_p = 0 \quad (6)$$

$$H_A: \rho_1 \neq \rho_2 \neq \cdots \neq \rho_p \neq 0 \quad (7)$$

Based on these hypotheses, if the sample is large enough, then:

$$LM = nR^2 \sim x^2(\rho) \qquad (8)$$

where n is the original sample size and R^2 is the value calculated using the residuals $e_1, e_2, \dots, e_n$ in the formula line 9.

$$e_i = \alpha_0 + \alpha_1 x_{i1} + \cdots + \alpha_k x_{ik} + \rho_1 e_{i-1} + \cdots + \rho_p e_{i-p} + \delta_i \qquad (9)$$

To use a Breusch-Godfrey test, *bgtest* function from the *lmtest* package is used. If the null hypothesis is rejected, then autocorrelation is present, and it requires correction. If positive autocorrelation is found, then additional lags or independent variables must be added to the model (D. A. Dickey, 1979); (Tsay, 2010). Negative autocorrelation might indicate over differencing in the variables. Finally, seasonal autocorrelation might require addition of dummy variables to the dataset to remove autocorrelations. It delivers similar results to L-jung box test for residuals made for time series model (Farnsworth, 2008).

Heteroscedasticity

Another important assumption to look at is heteroscedasticity in the dataset. Heteroscedasticity means inconsistency of the standard deviation of residuals (Wooldridge, 2019). Meaning that the data is scattered around the plot instead of being close to the best fit line. Heteroscedasticity is usually present in the variables that have large gap between lowest and highest values. To test for heteroscedasticity, White test is used in R Studio, and it has null hypothesis stating that variances of the errors are equal, while alternative hypothesis states that they are not equal. This is also presented in lines 10 and 11.

$$H_0 = \sigma_i^2 = \sigma^2 \qquad (10)$$

$$H_1 = \sigma_i^2 \neq \sigma^2 \qquad (11)$$

It was also explained by Wooldridge as: "*When Var[ε ├|x┤] depends on x, the error term is said to exhibit heteroscedasticity (or non-constant variance)*" (Wooldridge, 2019, s. 51). It is important to be aware of the heteroscedasticity because it can affect results that model shows. It can make coefficients less precise and produce p-values lower than they should be, which can affect the modelling and lead to using factors that are not significant when predicting day-ahead electricity prices. Hence, after running the white test on R studio, p-value can be used to conclude if data is homoscedastic – if p-value is over 0,05, then null hypothesis is not rejected and error terms are homoscedastic, and testing can be continued. However, if p-value is below the significance level of 0,05, then additional t-test is required to change all the heteroscedastic errors to the homoscedastic ones and run the test again. This method is called Heteroscedasticity-robust t-statistics and it was studied by MacKinnon and White in 1985 (James G. MacKinnon, 1985). Formula of the t-test is expressed in line 12.

$$t = \frac{estimate - hypothesized\ value}{standard\ error} \qquad (12)$$

When dealing with heteroscedastic errors, another test is necessary to run to test for significance of the factors after changing the errors to homoscedastic ones. This test is called Wald test and is part of *lmtest* package on R studio. Wald test has null hypothesis testing if betas for each parameter are equal to zero. The null hypothesis and alternative hypothesis are presented in lines 13 and 14.

$$H_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0 \qquad (13)$$

$$H_A = \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4 \neq 0 \qquad (14)$$

If null hypothesis is not rejected, then test concludes that none of the factors are significant when predicting y, however, if p-value is lower than 0,05, then at least one of the factors show significance and have predictive power. It is important to test for heteroscedasticity since variances that differ can affect results and show significance for factors, which have no significance at all (Wooldridge, 2019).

After the significant predictive factors are found out, the new linear regression model can be made with only the factors that showed significance and the whole testing can be repeated to make sure that all the factors included in the linear regression model have predictive power over day-ahead electricity prices. Only the models with statistically significant factors are included in the forecasting study.

Afterwards, the same method as for time series can be used – AIC and BIC criteria can be used to compare the models made using multiple linear regression and find the best fitting model for each dataset. These criteria are presented below.

Information criteria – AIC and BIC

Currently, some of the most popular methods to find the most accurate models are Akaike Information Criterion and Bayesian Information Criterion (Wooldridge, 2019). Expressions of these criteria are presented in lines 15 and 16.

$$AIC_i = -2\log L_i + 2p_i \qquad (15)$$

$$BIC_i = -2\log L_i + p_i \log n \qquad (16)$$

Where n is the number of values in the training dataset, $\log L_i$ is the log-likelihood of the model on the training dataset, and p_i is the number of parameters in the model.

Akaike Information Criterion is helpful estimating how much information model is losing. Of course, less information lost means higher quality model and, thus, the lower AIC is, the better model is to predict the prices.

On the other hand, Bayesian Information Criterion is helpful to find how well the new data set can explain the previous dataset (Fox, 2016). Hence, the lower the BIC is, the better at explaining the dataset the model is.

Time Series Models

The most popular models used in the industry for price forecasting are different time series model. The choice of model depends on the dataset that is used, however, many studies included different variations of ARMA or ARIMA models, since these models are usually used in cases of weak stationarity and with random variables (Weron R. , 2006).

Since time series is one of the most popular models used to forecast returns in financial markets, it also got attention in electricity price forecasting. To start with, Markus Lindström (2021) included ARIMA model in his study and compared it with other models when used for electricity price forecasting (Lindstrom, 2021). In his study, Markus introduced and compared many different forecasting strategies and discussed the popularity of time series in electricity price forecasting. The Rafal Weron (2014) stated that AR-type models are the backbone of time series models used to predict electricity prices (Weron R. , 2014). Similar studies were completed by Rafal Weron and Adam Misiorek (2008) as well, where they compared different types of time series models and tried to find the most accurate one (Misiorek, 2008).

Many other studies, that included time series model comparison with other types of forecasting models when predicting day-ahead electricity prices proves that time series are in fact very popular in the forecasting studies (J. Stuart McMenamin, 2000); (Korpihalkola, 2019); (Misiorek, 2008); (Sergey Voronin, 2013); (Xydis, 2019); (Antonio J. Conejo, 2005). Comparisons are usually including some kind of ARMA or ARIMA models and then another model with a very different forecasting strategy – linear regression, GARCH, SARIMA, Wavelet transform model, or different neural engine models. This confirms that time series models are highly popular and are used in the forecasting almost as a base to compare forecasting methods.

Hence, this study also included different Autoregressive processes together with Moving Average processes, which are presented in the sections below.

Autoregressive process

Autoregressive model is referred to as AR (p) and represents stochastic process. Stochastic means uncertainty and usually is referred to as randomness. This type of model is presented in the formula 17.

$$y_t = c + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \varepsilon_t \qquad (17)$$

Where $\emptyset$ indicate AR coefficients, thus, model AR (1) is expressed as presented in line 18.

$$y_t = c + \emptyset_1 y_{t-1} + \varepsilon_t \qquad (18)$$

To help with creating the right model that included just the lags found in the dataset, Partial Autocorrelation Function (PACF) is made to find the lag autocorrelations that help to create the model depending on the dataset of Autoregressive processes. This model is illustrated in line 19.

$$PACF(y_i y_{i-2}) = \frac{\text{Covariance}(y_i y_{i-2} | y_{i-1})}{\sigma_{y_i | y_{i-1}} \sigma_{y_{i-k} | y_{i-1}}} \qquad (19)$$

The PACF function is useful to use as a guide when creating the AR model depending on the number of the lags in the graph and is called AR lag order.

Moving Average process

Moving Average model presents dependency between observations and residual errors (Tsay, 2010). This model is presented in line 20.

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_p \varepsilon_{t-q} \qquad (20)$$

To find out the MA order, Autocorrelation Function (ACF) is used. The expression of the ACF is presented in the line 21.

$$p_k = \frac{\sum_{t=k+1}^{n} (y_t - \overline{y})(y_{t-k} - \overline{y})}{\sum_{t=1}^{n} (y_t - \overline{y})^2} \qquad (21)$$

, where p_k is ACF with lag k.

The same way as PACF is used to help to find the AR order, ACF function is used to help to find MA order.

Together, AR and MA orders, can be used as ARMA model, which is widely used in the forecasting of prices and returns already. In addition, additional model called ARIMA is part of this group. The ARIMA model includes Integrated (I) part together with AR and MA and represents the difference of raw observations to help time series to reach stationarity (Tsay, 2010). In this case, different models from this group will be used to find the best fitting one.

Random Walk Hypothesis

In the time series models, errors terms must follow Random Walk hypothesis. This hypothesis states that tomorrow's price is equal to today's price (Tsay, 2010). And is expressed as in line 22.

$$P_{t+1} = P_t + \varepsilon_{t+1},\ \varepsilon_{t+1} \sim WN(0, \sigma^2) \qquad (22)$$

In the formula ε_t indicates white noise process and means that error terms are unpredictable white noise shocks. This indicates that variable is not following any trends and moves randomly. Meaning, that day-ahead electricity prices are moving independently to any other factor or variable.

L-jung Box test

Each of the datasets for time series must not only include ACF and PACF functions, but also L-jung Box test is helpful finding out if any type of jointly autocorrelation is present in the data. This model uses squared weighted autocorrelations at specific lag. When this test is applied to the residuals of the model, then degrees of freedom must be equal to m - p (AR order) - q (MA order).

This test is a part of *stats* package in R studio and the formula for L-Jung Box test is presented in line 23.

$$Q(m) = \mathrm{n}(\mathrm{n}+2)\sum_{j=1}^{m}\frac{r_j^2}{n-j} \qquad (23)$$

Reject null hypothesis if

$$Q > x_{1-\alpha,h}^2 \qquad (24)$$

If the null hypothesis is rejected, then some sort of autocorrelation can be found in the data or residuals. The purpose of this test is to test for randomness and make sure that errors in the time series follow Random Walk hypothesis presented above.

Stationarity

Stationarity is another requirement for dataset for the time series models. Usually, stationarity can be searched for in the plots of the variables or by using Augmented Dickey-Fuller (ADF) test presented below (Tsay, 2010).

Augmented Dickey-Fuller test

This test is used to test for stationarity in the test. The hypotheses for the test are presented in lines 25 and 26.

$$H_0: \alpha = 1 \qquad (25)$$

$$H_A: \alpha \neq 1 \qquad (26)$$

From the line:

$$y_t = c + \beta_t + \alpha y_{t-1} + \phi \Delta Y_{t-1} + e_t \qquad (27)$$

Where y_{t-1} is lag 1 of time series and ΔY_{t-1} is first difference of the series at time t-1. Hence, if the null hypothesis is rejected, then series is stationary, and no additional steps are required. However, if null hypothesis is not rejected, then differencing is required to make series stationary (Fuller, 1995).

To run this test *adfTest* command used from the *fUnitRoots* package. If series is not stationary, then a number of differencing required to make it stationary must be determined. The number of differencing can be obtained from the ADF test since number of roots contained in the series is equal to the number of differencing required to make it stationary (Denis Kwiatkowski, 1992). To do that differencing function can be used.

Maximum Likelihood

Maximum likelihood is widely used technique to deriving estimators. However, in the forecasting it is used to find the values of the parameters with the highest likelihood to obtain the data that was observed. The Rafal Weron stated in his book that: "*the ML estimates are almost always the most accurate, followed by regression-type estimates and quantile methods*" (Weron R. , 2006, s. 57). Thus, time series forecasts using Maximum Likelihood are part of the study.

To try different techniques, forecasts will be made loop than is based on Maximum Likelihood, however, *auto.arima* function will be based on AIC criteria to see which method can deliver more accurate forecast. Even though it is important to note that *auto.arima* is not always performing well, so it is important to cross-reference the results.

Time series model selection strategies

To select time series models' orders, different strategies are used. Firstly, ACF and PACF functions are used on each dataset to find the orders from the lags in the charts. Then, a few other models are going to be included. For example, AR (p) and MA (q) models with different orders. Finally, *auto.arima* function is used to automatically detect the model with the lowest AIC from the dataset and use it in the forecasting as well.

Naïve model

In this case a seasonal naïve model can be used to try to follow the same patterns as happened throughout the year to create the forecast. Seasonal forecast means that data from last year's February is used as forecasted prices in this February and so on (Athanasopoulos, 2018). Formula of the naïve model is presented in the line 28.

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)} \qquad (28)$$

where m is the seasonal period, and k is the integer part of $(h-1)/m$. This model uses the observed prices from the previous year as the future prices. It follows similar assumption to Random Walk Hypothesis, just that it uses prices from last season or month instead of the previous day.

Forecast 316 days

It is chosen to produce a forecast for the upcoming 316 days when using these different models and then compare it with the real day-ahead electricity prices. This is set from the training and testing samples, where 40% of the dataset is used for testing the model.

RMSE

For the comparison of the forecasts, RMSE is made for each model and the formula is presented in line 29.

$$RMSE = \sqrt{(o-f)^2} \qquad (29)$$

Where o is observed values, and f is forecasts.

This measure indicates standard deviation of the prediction difference from real price. It is squared and rooted, so that negative values would become positive ones and the mean of all errors could be easily found without dealing with negative and positive numbers. This measure is widely used in regression and forecasting studies to help to compare forecasts made with different models.

All RMSEs are compared in the findings chapter to find the model that is closest to the real day-ahead electricity prices. Then, conclusion can be made regarding the best performing model out of the ones used.

3.5 Validity

In this study, peer reviewed articles and studies that are no older than 30 years old are used as sources to increase the validity of the study. In addition, validity of the study is increased significantly with each additional model added to the study since it helps to identify the best model out of ones used in the study and make it more accurate than just making a single model forecast. Also, all forecasts are checked for errors with different tests and out-of-sample study is used to increase validity of the study.

3.6 Reliability

The study can be replicated by using the appendixes with R Studio code and dataset attached to the report. Also, since historical data is used and all steps of the code are presented and explained, the study can be easily repeated making it more reliable. Because the study does not include any qualitative data, it means that statistical results are identified in the same way removing the bias that are usually found in qualitative study.

4 Analysis

This chapter will include analysis done using methods and theories described above.

To begin analysis, variables from 3 different datasets are plotted to check for stationarity for the linear regression. Plots for 4 factors are presented in figure 10.

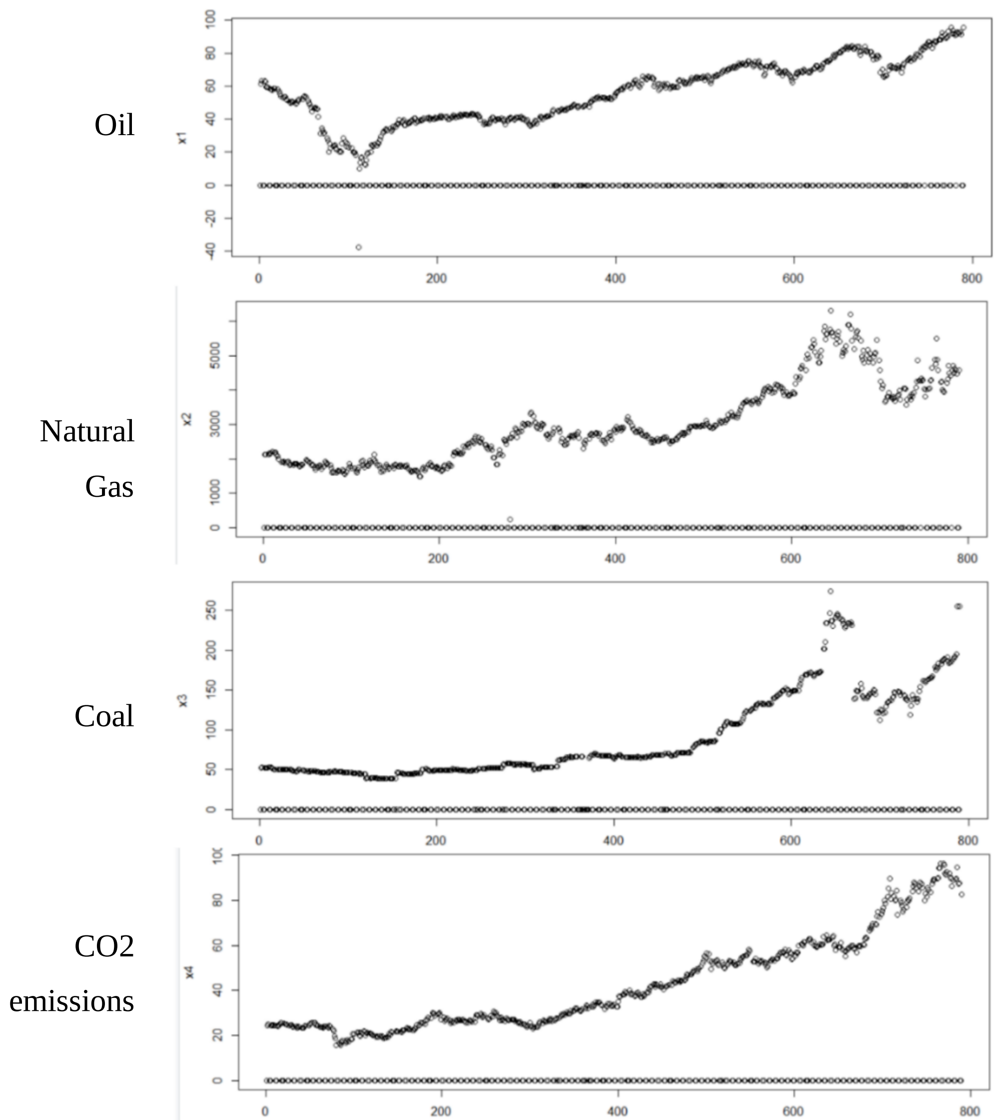


Figure 10 - Plots of independent factors - x1 (Oil), x2 (Gas), x3 (Coal), and x4 (CO2 emissions)

From the plots it can be concluded that factors: x1 (Oil), x2 (Gas), x3 (Coal), and x4 (CO2 emissions) line of best fit can be drawn in each of the plot and thus it can be concluded that these independent factors are linear variables. Following this, plots for spot prices in three different time slots: 7AM, 1PM, and 9PM are illustrated in figure 11.

7AM

1PM

9PM

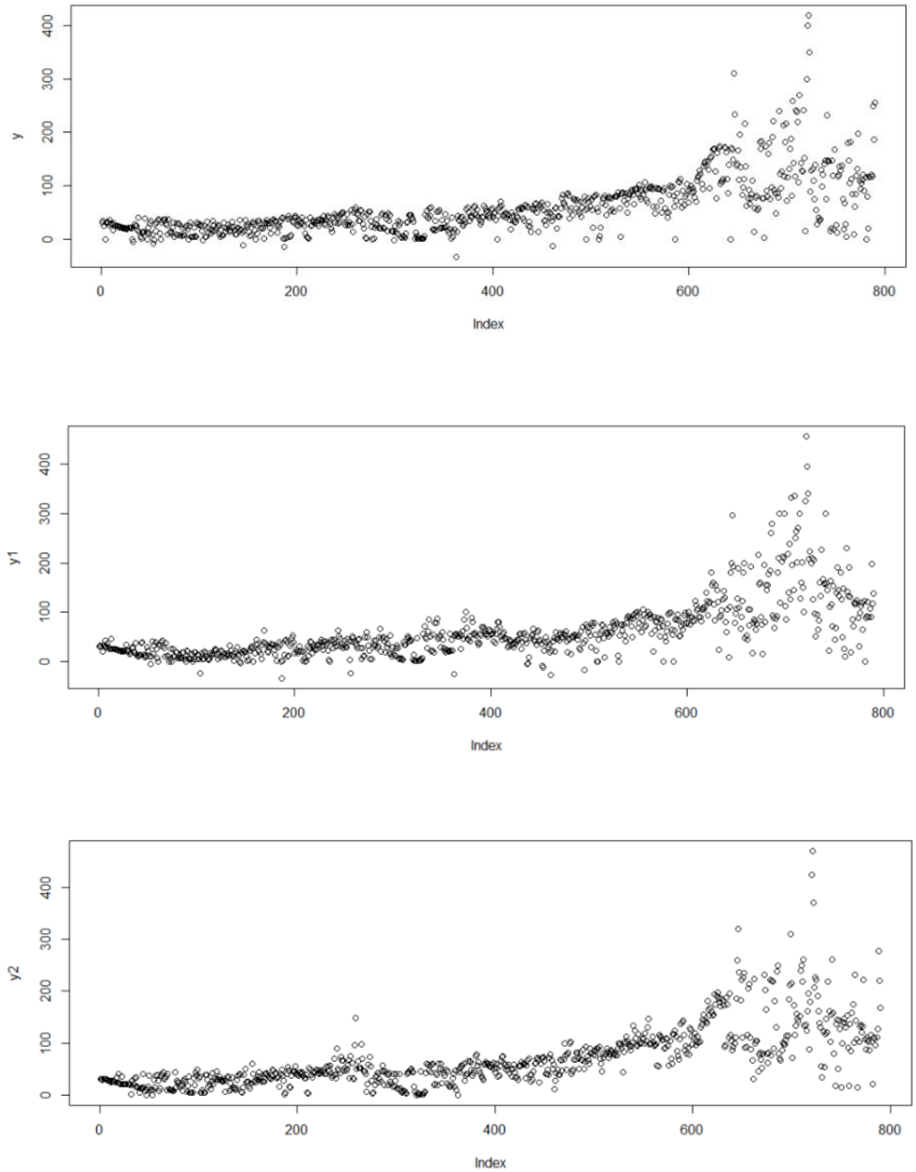


Figure 11 - Plots of dependent variables – y (7AM), y1 (1PM), and y2 (9PM)

Form these plots it can be concluded that line can be drawn in at least a part of the dataset, however, interestingly, variables get very volatile in the last third of dataset and it gets more difficult to place the best fit line in there.

In addition, correlation matrix is made for each dataset to see if there are any correlations between dependent and independent variables. The correlation matrixes can be found in figure 12.

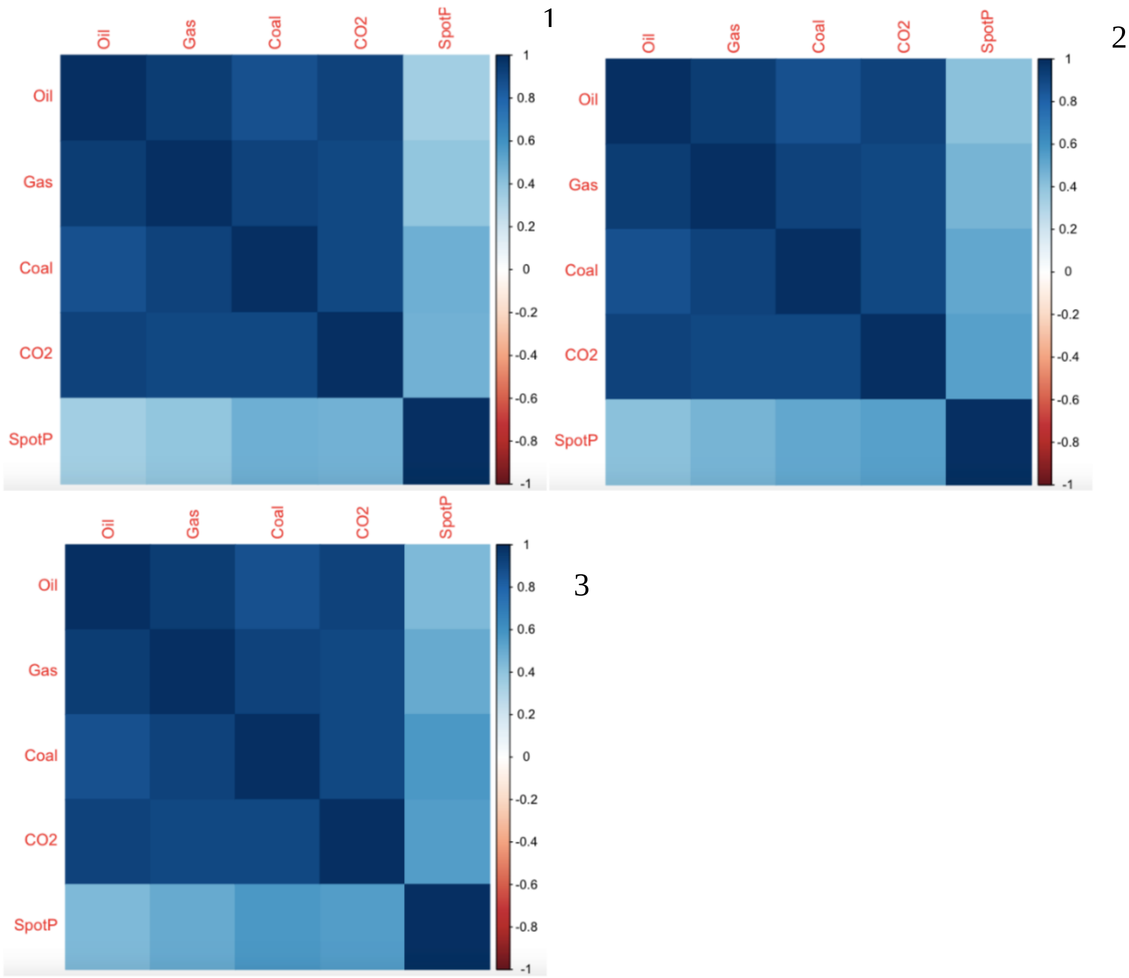


Figure 12 - Correlations – 1 (7AM), 2 (1PM), 3 (9PM)

Correlation matrix indicates that independent factors used in the linear regression model are correlated with each other but not as much correlated with the Spot Price. These results confirm that all fossil fuel prices are very much connected, since to transport coal, oil and gas are used and similarly in mining, machinery uses fuel to be able to extract the resources making the prices correlated with each other. CO2 emissions are also correlated to the fossil fuel because CO2 emissions are produced when using fossil fuel in the power plants, hence, the price for emissions are moving together with the prices of the fossil fuel.

Afterwards, linear regression model is made that has independent factors as predictors to test their significance on predicting dependant variable. During this testing process, other specifications are tested for: normal distribution using Jarque-Bera test, and White test for heteroscedasticity. If needed, t-test can be used to test for significance of factors after the errors are changed from heteroscedastic to homoscedastic ones. Moreover, Wald test is used to make sure that at least one of the factors is indeed significant when predicting day-ahead price. Breusch-Godfrey test was also made using order of 365 since variables are daily. If p-values are rejected, then actions to fix the dataset must be taken.

If any of the models include factors that show no significance, re-estimation of the model to remove insignificant factors and make a new model with only factors that have p-value lower than 0,05 is required. In the forecasting phase, only the significant factors that have predictive power over day-ahead price must be used for the best results.

To test the general models, these names were given to each model: model – linear regression model with four factors and Spot price from dataset 7AM. The model1 - linear regression model with four factors and Spot price from dataset 1PM. Then, model2 - linear regression model with four factors and Spot price from dataset 9PM. Since Gas prices had p-value higher than 0,05, re-estimated models were created: model → modela, model1 → model1a, model2 → model2a.

After testing for linear regression is completed, testing for time series models can be used to construct the models for forecasting. The ACF and PACF functions can be made together with Ljung-Box test to check for autocorrelations and ADF test for stationarity.

After the general testing is done, forecasting can be started. The dataset is divided into two subsets of training sample, which consists 60% of all data, and then testing sample, which consists of the rest of 40%. To start with, the loop is created to make a 316-steps (40% testing sample) forecast for the same hour each day. Then, linear regression and different time series models are used to make different forecasts and to compare the results, RMSE is used as main tool to compare the forecasts' accuracy.

5 Findings

This chapter will include findings from analysis above.

Before presenting the forecasts from different models, it is important to note that only three factors showed predictive power over day-ahead electricity prices in linear regression: Oil, Coal, and CO2 emission prices. The results from the tests made to create a model with only significant factors are presented in figure 13.

Linear Models

p-values	7AM		1PM		9PM	
	model	modela	model1	model1a	model2	model2a
Jarque Bera	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
White test	0.000000132	0.000000112	0.0000000482	0.000000857	0.00000000372	0.00000384
Wald test	< 2.2e-16 ***	< 2.2e-16 ***	< 2.2e-16 ***	< 2.2e-16 ***	< 2.2e-16 ***	< 2.2e-16 ***
BG test	0,03657		0,007179		0,01726	
BG test with dummy	0,03999		0,01145		0,02183	
BG test with higher order	0,0521		0,05057		0,05095	

Time Series models

p-values	7AM	1PM	9PM
Box test	< 2.2e-16	< 2.2e-16	< 2.2e-16
ADF test	0,01 (p-value is smaller than printed)	0,1176 (p-value is smaller than printed)	0,01 (p-value is smaller than printed)

Figure 13 - Linear regression models and Time series models testing

As it can be seen in the figure 13, all the models throughout the testing had p-value lower than 0,05. Meaning that all six models included in the testing have no normal distribution, however, since normal distribution is not a requirement when working with multiple factors, testing can still be continued. As mentioned previously, it is not a requirement for multiple linear regression since the variables are added to the model, the more chances there are that not all the data is normally distributed (Wooldridge, 2019). Then, because White test null hypothesis is rejected, heteroscedastic errors are present in the dataset and additional t-test is required to change heteroscedastic errors to homoscedastic ones. Finally, Wald test is used to test if any of the factors are significant when predicting day-ahead electricity prices. Since all of the p-values are lower than 0,05 and null hypothesis is rejected, at least one of the factors is significant when predicting the prices. It is also important to note that Breusch-Godfrey tests for each model has p-value lower thank 0,05 significance level meaning that corrections are required. In this case, order of the model can be increased to make sure that p-values are higher, or dummy variable for seasonality can be added to the dataset to see if it increases Breusch-Godfrey test's results. Values of 1 marking weekends and 0 marking weekdays were used to try to see if it could improve the dataset. Seasonality was chosen since it is often mentioned as factor that could affect the prices of electricity (Xydis, 2019). However, dummy variable showed very little of improvement and increase in orders were more useful. Dataset for 7AM data needed 387 degrees of freedom, 1PM required 403, and 9PM dataset – 391 to get p-value higher than 0,05. Meaning that each of the datasets require quite a large part of the dataset to be allowed to vary without restraining constraints.

It is important to note that re-estimation was necessary since Gas prices showed no significance when predicting day-ahead electricity prices, hence, new models with only factors were created. In addition, AIC and BIC were used to check if new models with three factors were more accurate than the models with all four factors. All the AIC and BIC were lower for the new models, meaning that these models are more accurate and loses less information. Thus, only the model with these three factors is used in forecasting. Code used for this study can be found in Appendix B.

When looking at time series, L-jung Box tests had p-values lower than 0,05 significance level and indicates some sort of autocorrelation between the variables. Even though many different tests with different lag values were tried, it still had a p-value lower than 0,05. As for the ADF test, p-values can be found in the figure 13, where it indicated that p-values are even lower than were printed in each test, meaning that null hypothesis can be rejected, and time series show stationarity.

After running the code, all the forecasts are collected in one table and presented in the figure 14 below.

Data	Model	RMSE	Data	Model	RMSE	Data	Model	RMSE
	lm	88,45		lm	96,75424		lm	95,0287
	ar1	207,1301		ar1	91,05972		ar1	102,2015
	ar2	206,5462		ar2	90,94478		ar2	101,2449
	ar3	204,0942		ar3	90,28005		ar3	103,6896
	ar4	202,8028		ar4	89,26157		ar4	100,3471
	ar5	201,8234		ar5	88,57478		ar5	98,0145
	ar6	202,1354		ar6	90,36486		ar6	99,9682
	ar7	205,0604		ar7	89,75136		ar7	102,7484
	ar8	204,7492		ar8	89,70738		ar8	103,0462
	ar9	204,4321		ar9	90,45013		ar9	103,1143
y	ma1	216,8373	y1	ma1	97,72569	y2	ma1	118,6552
	ma2	212,4333		ma2	94,23674		ma2	108,1542
	ma3	212,1977		ma3	93,50209		ma3	106,612
	ma4	210,5897		ma4	93,00225		ma4	108,9271
	ma5	209,7442		ma5	92,81834		ma5	107,6549
	ma6	210,6338		ma6	92,36795		ma6	108,4957
	ma7	206,7749		ma7	92,97754		ma7	107,4843
	ma8	204,0168		ma8	90,15353		ma8	106,9257
	ma9	202,2388		ma9	89,14824		ma9	103,6456
	best model	205,9742		best model	90,87429		best model	98,0145
	Naive mode	96,39		Naive mode	100,9627		Naive mode	104,2735

Figure 14 - RMSE for forecasts where lm (linear regression model), best model (model from auto.arima function) and Naïve model for naïve model. Green colour indicates two lowest RMSEs for each dataset.

As indicated with the green colour in the figure, lowest RMSE from each group of models is found. In the y (7AM) dataset, linear model has the lowest RMSE and is best at predicting the day-ahead prices. This confirms that in the morning, when people are waking up and starting to use the energy, mostly fossil fuels are affecting the day-ahead electricity prices since they are used to start many power plants and catch up with the demand. It is important to mention that Naïve model has second lowest RMSE and performs slightly better than other models used in the study. From the two different time series model strategies, the ar5 and ma9 with Maximum Likelihood delivered lower RMSE than the model made with highest AIC.

When looking at the y1 (1PM) dataset, it can be concluded that time series, especially the ar5 model, is the best from the ones used in the comparison. It has the lowest RMSE, proving the point that different load times in the day have different 'best models' to forecast the prices. As seen in the figure, linear model has the highest RMSE from the four green lines, opposite to the hour used before. It is important to note that these results indicate how prices during the On-peak periods are less likely be affected by fuel prices. It might happen due to the increase in energy made from the renewables, since solar energy can only be produced throughout the day. This is something that future studies should test to see if these factors can help to predict electricity prices in the On-peak period.

Finally, 9pm dataset delivers similar results to the ones seen in the 7am, where Off-load hours can be predicted using linear model and are a lot more affected by the fossil fuel prices, while On-peak time delivers better results with using the time series models. Here, Maximum Likelihood and AIC delivers the same results and confirms that ar5 model is the most fitting from the ones used in the forecasting. In this specific case, AR and linear model delivers similar RMSEs and indicates that Off-peak 2 interval is a mix between Off-peak 1 and On-peak forecasts.

Comparison of RMSEs for all the different models used in the study helps to identify differences between On-peak and Off-peak periods and hours throughout the day, meaning that a simple time series or linear model is not enough to accurately predict day-ahead electricity prices. These results provide another interesting fact, RMSE delivers a mean error, meaning that some of the intervals in 316-steps forecast can have higher accuracy than others. Therefore, it is difficult to make the conclusion on the forecasts' accuracy just from the RMSE. However, it might be useful to make shorter forecasts with only 30 or 100 steps or try to divide it into intervals. In addition, plots with the forecasts can be used to see how forecast and real data differ. To better visualise the results of forecasting, plots of all the forecasts are presented below.

Naïve model forecasts are illustrated in the figure 15.

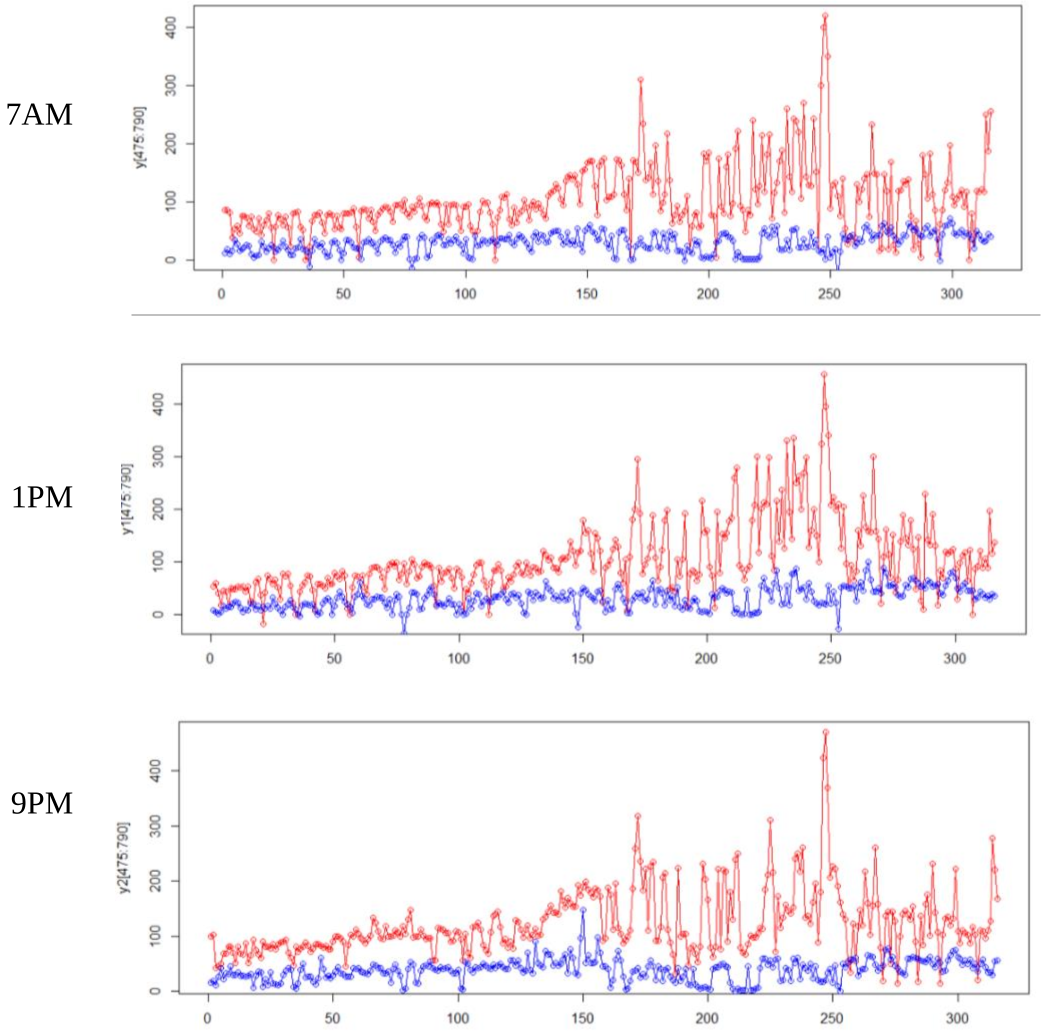


Figure 15 - Naive model forecast comparison with real prices - top chart is for y, middle for y1 and bottom one presents y1. Red colour indicates real prices and blue colour - forecasts

As seen in the figure 15, forecast is quite close to real values in the beginning, however, when volatility of the prices increases, the forecast becomes less accurate. Thus, Naïve model is outperformed significantly by other models used in this study in all the datasets.

Then, plots for this comparison of linear regression model made with 7am dataset is presented in the figure 16.

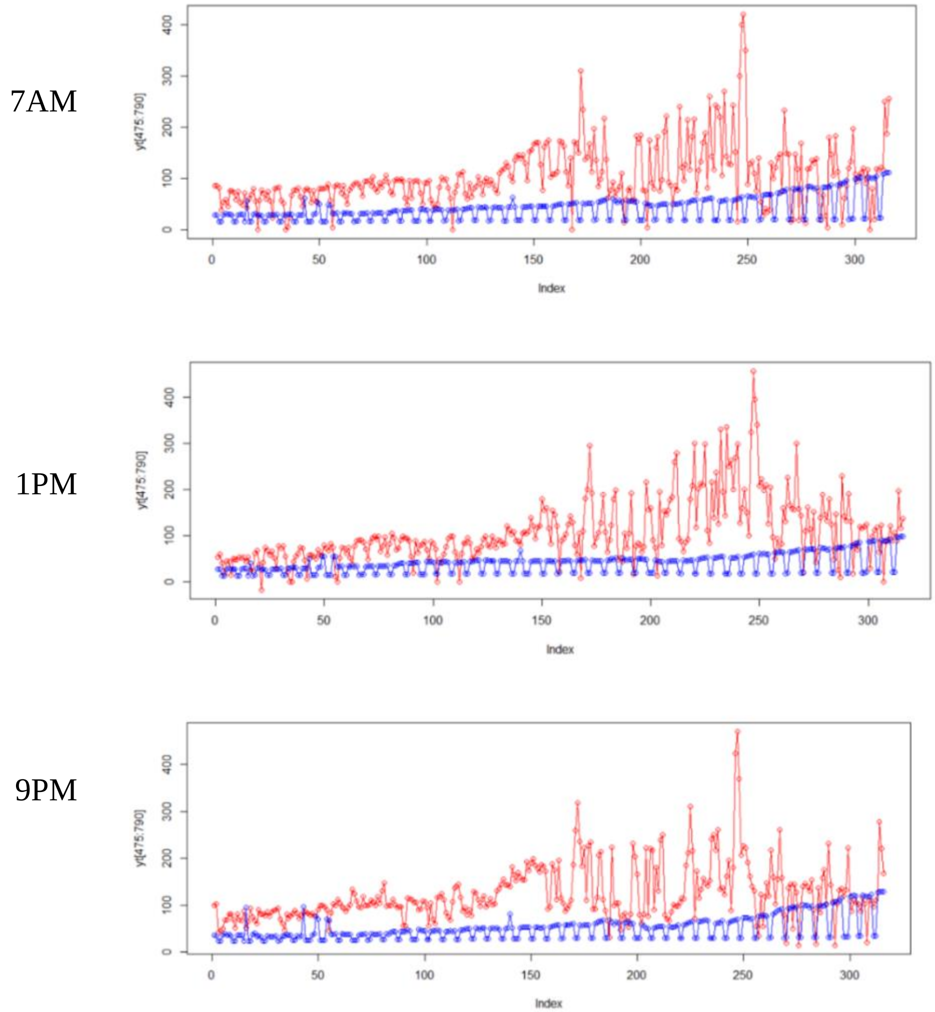


Figure 16 - Linear model forecast comparison with real prices – top chart is for y, middle for y1 and bottom one presents y1. Red colour indicates real prices and blue colour - forecasts

From the figure 16 it can be interpreted that linear regression forecast was close to the actual prices in the beginning, but as market got more and more volatile due to the events in Europe and World, the forecast was not able to predict these spikes that well anymore. Nevertheless, in many places it shows similar trends of going up or down, but unfortunately is relatively far away from the actual day-ahead electricity price. Forecast from other hours are very similar and shows the same trend of getting less accurate throughout the time. Then, plots for time series forecasts made with 7AM dataset are presented in figure 17 below.

AR (5)

MA (9)

Best.model

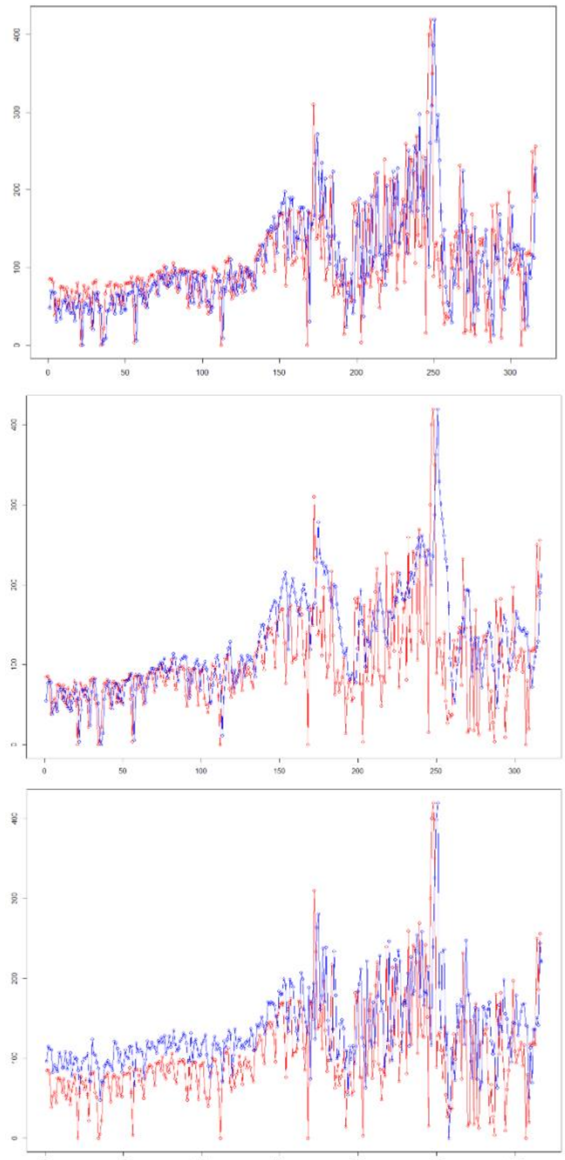


Figure 17 - time series model forecast comparison with real prices for y (data7AM), – 1(AR model 2 (MA model), 3 (best.model). Red colour indicates real prices and blue colour - forecasts

Plot in the top indicates AR (5) model forecast. In the middle, ARMA (1.1) plot is presented. Plot in the bottom is made for MA (9) forecast. Plots indicate that MA (9) and ARMA (1.1) models predicted higher prices than they were in at least some periods. The finding that is very important is that from plots it can be seen how different models fit different periods better than others. For example, ARMA (1.1) model fits the first 100 prices very well. Then, AR (1) and MA (9) models are more accurate when fitting the real prices from 100 to 316 predictions. Nevertheless, it is important to compare forecasts made with 1PM dataset as well. These forecasts are presented in figure 18.

AR (5)

MA (9)

Best.model

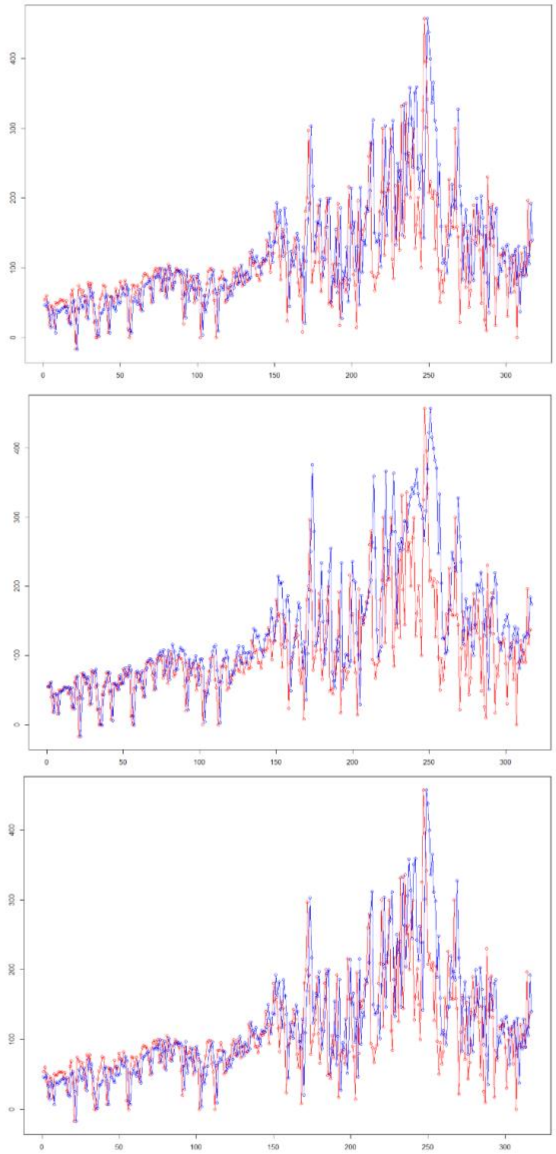
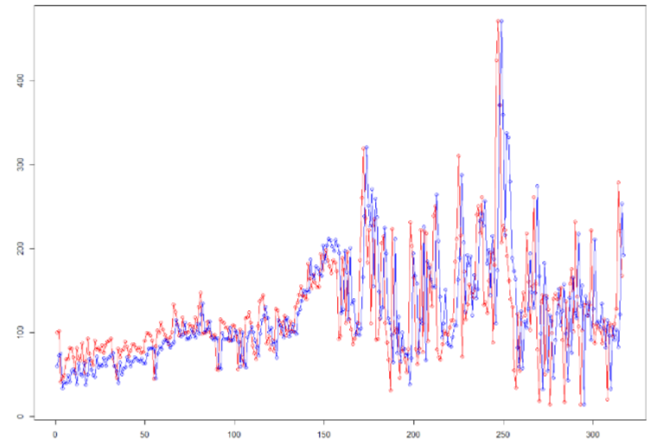


Figure 18 - time series model forecast comparison with real prices from y1 (data1PM) dataset – 1(AR model 2 (MA model), 3 (best.model). Red colour indicates real prices and blue colour - forecasts

First plot indicates AR (5) model forecast. In the middle, ARMA (1.1) plot is presented. Last plot is made for MA (9) forecast. These plots confirm the results seen in the figure 18. Again, ARMA (1.1) is better at forecasting values in the beginning, where volatility is not that high. However, other two models get more accurate afterwards. Finally, last dataset including 9 PM data forecasts are presented in the figure 19 below.

AR (5)



MA (9)

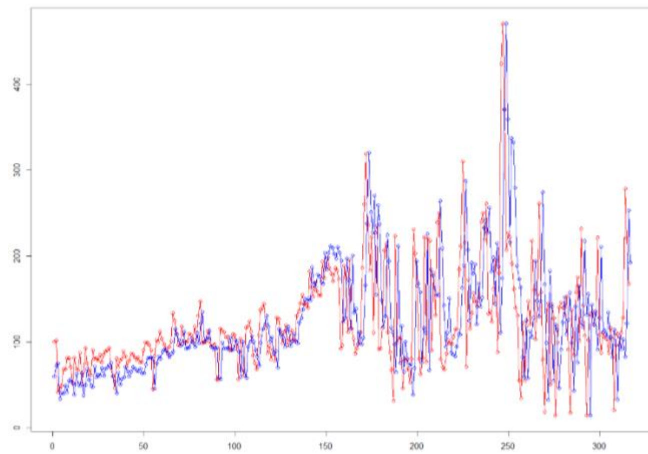


Figure 19 - time series model forecast comparison with real prices for y2 (data9PM) – 1 (AR (5) model 2 (MA model). Red colour indicates real prices and blue colour - forecasts

Again, models are presented from the top: AR (5) and MA (9). It included only two plots because two methods indicated the same AR (5) model. As it can be seen in the plots, all three models for this dataset were not very accurate in the first 50 prediction. However, when prices got more volatile, predictions got a lot more accurate and were even able to predict large spikes of the prices.

Overall, time series models showed significantly more accuracy in predicting the movement of the prices, however, on some occasions the spikes were predicted too early or too late, increasing the errors of the prediction. Since most of the electricity is traded throughout the On-Peak period from 9:00 to 20:00, the model that can best predict the prices in that period is the most beneficial to look into more. From all the models included in this study, AR (5) model was the most accurate predicting 1PM dataset and could be used to plan the future capacity purchasing to increase profits.

Applicability of research

To use this study in real life, forecasting models must be used not only in one, DK1, market, but in all the neighbouring markets as well. After creating the forecast and comparing them, plans for capacity purchasing can be made, if price of the electricity is going to be lower in DK1 than in Germany, then interconnector’s capacity can be bought early, and cheap energy from DK1 can be transferred and sold at the German market delivering better profits. In the case where DK1 has lower prices than any of the neighbouring markets, traders should go long in Denmark and short in the market with higher prices (if any capacity for the interconnectors is available), and then if neighbouring markets has lower prices than Denmark, then traders should go short in Denmark and long in other markets. Since commodity trading has many more costs involved, for example, transportation and handling, usually the sooner these things can be ordered, the higher profit can be earned from the trade. Meaning, that traders who find accurate forecasting models can earn a lot higher profits compared to the ones that just makes trades without it. Since Trading Strategies usually contain three main steps: planning, placing orders and executing them, this research’s results can help in the planning process to prepare for the upcoming day-ahead electricity prices and build a trading plan. This is also leaning towards technical trading strategy since it includes models that affect trading choices. One of the main advantage of this strategy is that it cannot be affected that much by bias of the trader and focuses on the results from the quantitative analysis and model results. It also can be described as Momentum trading strategy since it looks at the changes in electricity prices and their movement to use it to earn profit. However, results reveal that changes in the market or events that cannot be quantified can still affect the day-ahead electricity price movement, hence, this study could be a great addition or tool to help to make trading choices, remembering that news and regulations happening in the market still must be considered when trading.

6 Discussion

This chapter will include discussion with limitations and possible future research ideas.

When looking at the results from the different forecasts it can be seen that electricity trading has a lot more technicalities attached to trading: it follows seasons not only throughout the year, but daily ones as well depending on the consumption of the electricity. In addition, new technologies and legislations, due to the new initiatives regarding the Green transition and other changes in the market, have effect on how electricity price behaves. Renewable energy has increased volatility of the electricity prices, while new regulations or events like war in Ukraine can affect prices of the fossil-fuel increasing the prices of the electricity production and making producers to earn less profit than expected since electricity prices have ceiling set by government. Even though it is difficult to quantify political regulations and news, it might be useful to try to include it in the forecasting study to see if these factors can indeed help to increase accuracy of the predictive model.

6.1 Limitations and future research

Due to the limit of time, only some of the forecasting models were included in the study. However, from the finding of the project it can be seen that depending on the time of the day or year, different models showed higher accuracy. Hence, Forecasting Combinations can be a great tool to try to forecast day-ahead prices. Forecasting Combinations are based on combining different models with weights for each and then making forecast with it (Timmermann, 2018). It is a flexible strategy that could benefit the everchanging electricity market and give traders a possibility to adjust the weight of the models depending on the situation in the market.

Another possible addition to the analysis could be adding forecast made with GARCH model, since last 100 Spot prices showed high volatility due to changes in the market. The GARCH model is more useful in case where prices are very volatile, where data does not follow the line but is more clustered (Tsay, 2010).

Furthermore, smaller intervals can be used to create forecasts. For example, 30 or 100 days could deliver a more accurate prediction comparing with the 316 one.

Moreover, other types of forecasting models from other research could be included in the study as well: Wavelet Transform model, factor model, and others. With more different models to compare, validity of this study and its findings could be increased even more.

In addition, it is important to note that project included more programming than anticipated and since researcher had lack of experience in this field, it could have reflected in the outcome.

7 Conclusion

This chapter will include conclusion for the whole research.

Overall, day-ahead electricity prices forecasting is quite a complicated process since it can be affected by many factors, which can differ depending on the time of the day or year: production, consumption, weather conditions, fossil fuel prices, export and import of energy as well as seasonality. Moreover, it was confirmed in the study that different load periods throughout the day can have different models predicting the prices. All of these changes and additional volatility in the market due to the new technologies in the electricity production as well as political changes in EU makes prices even more difficult to predict. To keep electricity prices fairly low more and more regulations are released constraining energy traders and making it more difficult to plan capacities' purchasing. Hence, energy industry is in need of a model that can predict day-ahead electricity prices accurately to help to earn higher earnings.

As presented in findings chapter, depending on the load time of the day, most accurate models from the ones included in the study differed significantly. However, the linear model was able to outperform naïve and time series models in the forecasts for 7AM, while time series showed more accuracy in the forecasts made for 1PM and 9PM over other models included in the study.

All in all, these forecasting models can be used as a guide of where prices can be expected to move and buy capacities depending on that to earn higher profit in the end. Technical momentum strategy can be built from using results from this research to trade following price movements and its changes to buy cheaper capacities early and earn higher profits when trading in neighbouring markets.

8 Bibliography

Alvaro Escribano, J. I. (2002, June). Modeling electricity prices: international evidence. *SSRN*.

Andrew Henley, J. P. (1996). *Energy Pricing and Temperature Interaction: British Experimental Evidence.* University of Wales, Aberystwyth, Department of Economics.

Antonio J. Conejo, M. A. (2005, May). Day-Ahead Electricity Price Forecasting Using the. *IEEE TRANSACTIONS ON POWER SYSTEMS, 20*(2).

Athanasopoulos, R. J. (2018). *Forecasting: Principles and Practice* (2nd ed.). Melbourne, Australia: OTexts. Retrieved from otexts.com: https://otexts.com/fpp2/simple-methods.html

C.R. Knittel, M. R. (2005). An Empirical Examination of Restructured Electricity Prices. *Energy Economics, 27*, 791-817.

Carlos Fernández Alvarez, G. M. (2021). *What is behind soaring energy prices and what happens next?* IEA.

Claudia Eliane da Matta, N. M. (2021). *A comparative study of forecasting methods using real-life econometric series data.* SciELO Brasil.

Colli, F. (2021). A transition for the citizens?: Ensuring public participation in the European Green Deal. *European Policy Brief, 68*.

D. A. Dickey, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association, 74*(366), 427-431.

Denis Kwiatkowski, P. C. (1992, October-December). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics, 54*(1-3), 159-178.

Energinet. (n.d.). *energinet.dk*. Retrieved February 2022, from energinet.dk: https://energinet.dk/

European Commision. (n.d.). *energy.ec.europa.eu*. Retrieved May 2022, from energy.ec.europa.eu: https://energy.ec.europa.eu/topics/markets-and-consumers/market-legislation/electricity-market-design_en

European Commission. (2021). *ec.europa.eu*. Retrieved from ec.europa.eu: https://ec.europa.eu/clima/citizens/citizen-support-climate-action_en

European Commission. (n.d.). *ec.europa.eu*. Retrieved February 2022, from ec.europa.eu: https://ec.europa.eu/reform-support/what-we-do/green-transition_en

European Commission. (n.d.). *ec.europa.eu*. Retrieved February 2022, from ec.europa.eu: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_17

Farnsworth, G. V. (2008). *Econometrics in R.*

Fox, J. (2016). *Applied Regression Analysis & Generalized Linear Models* (3rd ed.). SAGE.

Fuller, W. A. (1995). *Introduction to Statistical Time Series* (2nd ed.). Wiley & Sons.

Griffin, J. M. (n.d.). The effects of higher prices on electricity consumption. *The Bell Journal of Economics and Management Science, 5*(2), 515-539.

Hitachi Energy. (n.d.). *hitachienergy.com*. Retrieved February 2022, from hitachienergy.com: https://www.hitachienergy.com/case-studies/the-great-belt-link-in-denmark

IEA. (2021, July 15). *IEA.org*. Retrieved from IEA.org: https://www.iea.org/news/global-electricity-demand-is-growing-faster-than-renewables-driving-strong-increase-in-generation-from-fossil-fuels

Investopedia. (n.d.). *investopedia.com*. Retrieved April 2022, from investopedia.com: https://www.investopedia.com/ask/answers/021615/whats-difference-between-commodity-and-product.asp

Ivesting.com. (n.d.). *Ivesting.com*. Retrieved February 2022, from Ivesting.com: https://www.investing.com/

J. Stuart McMenamin, P. F. (2000). Statistical Approaches to Electricity Price Forecasting. *Itron*.

James G. MacKinnon, H. W. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics, 29*(3), 305-325.

Jannik Schütz Roungkvist, P. E. (2020). High-Resolution Electricity Spot Price Forecast for the. *Sustainability*.

Joscha Schabram, X. V. (2021). *A new age for energy and commodity trading.* McKincey & Company.

Korpihalkola, J. (2019). *Forecasting Nord Pool's Day-ahead hourly spot prices in Finland.*

Kuada, J. (2012). *Research Methodology. A project Guide for University Students* (1st ed.).

Lindstrom, M. (2021). *Forecasting day-ahead electricity prices in Sweden.* Umeå.

Mcleod, S. (2018). *Maslow's Hierarchy of Needs.*

Micucci, L. (2022). *gasprocessingnews.com*. Retrieved from gasprocessingnews.com: http://www.gasprocessingnews.com/features/202104/process-technologies-for-lng-production.aspx

Misiorek, R. W. (2008). Forecasting spot electricity prices: A comparison of parametric and. *Elsevier*, 744-763.

Monforte, J. S. (2000). Statistical Approaches to Electricity Price Forecasting. In A. E. Faruqui, *Pricing in Competitive Electricity Markets. Topics in Regulatory Economics and Policy Series* (Vol. 36, pp. 249-263). Springer Science+Business Media New York.

Nektaria V. Karakatsani, D. W. (2008). Forecasting electricity prices: The impact of fundamentals and. *Elsevier*, 764-785.

Penelope L Peterson, E. L. (2010). *International encyclopedia of education.* (3rd ed.). Oxford, UK : Academic Press.

Sergey Voronin, J. P. (2013). A hybrid electricity price forecasting model for the Nordic. *INTERNATIONAL TRANSACTIONS ON ELECTRICAL ENERGY SYSTEMS*.

Statista. (n.d.). *Statista.com*. Retrieved February 2022, from Statista.com: https://www.statista.com/

Svetlozar T. Rachev, S. M. (2007). *Financial Ecnometrics. From Basics to Advanced Modellig Techniques.* John Wiley & Sons Inc.

Takashi Kanamura, K. O. (2007). A structural model for electricity prices with spikes: Measurement of spike risk and optimal policies for hydropower plant operation. *Energy Economics, 29*(5), 1010-1032.

Timmermann, A. (2018). Forecasting Methods in Finance. *Annual Review of Financial Economics, 10*, 449-479.

Tsay, R. S. (2010). *Analysis of Financial Time Series* (3rd ed.). Wiley.

Tuomas Rintamäki, A. S. (2017). Does Renewable Energy Generation Decrease the Volatility of Electricity Prices? An Analysis of Denmark and Germany. *Energy Economics, 62*.

Unicef. (2022, March 1). *unicef.org*. Retrieved from unicef.org: https://www.unicef.org/northmacedonia/press-releases/new-survey-shows-society-concerned-about-climate-change-and-its-impact-children

W. Alex Edmonds, T. D. (2017). *An Applied Guide to Research Designs. Quantitative, Qualitative and Mixed Methods* (2nd ed.). SAGE.

Weron, R. (2006). *Modelling and Forecasting Electricity Loads and Prices. A Statistical Approach.* John Wiley & Sons, Ltd.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting, 30*(4), 1030-1081.

Winther, C. D. (2020). *Visual Guide to the Power Grid Inside the greatest machine in the world* (1st Edition ed.).

Wooldridge, J. (2019). *Introductory Econometrics* (7th ed.). Cengage Learning, Inc.

World Nuclear News. (2022). *world-nuclear-news.org*. Retrieved from world-nuclear-news.org: https://www.world-nuclear-news.org/

Xin Liu, Z. Z. (2019). A comparative study of the data-driven day-aheaad hourly provincial load forecasting methods: From classical data mining to deep learning. *Elsevier - Renewable and Sustainable Energy Review, 119*.

Xydis, O. A. (2019). Electricity Price Forecasting in the Danish Day-Ahead Market Using the TBATS, ANN and ARIMA Methods. *Energies*.

9 Appendices

Appendix A – Excel with data

Appendix B – R studio code