VILNIUS UNIVERSITY
FACULTY OF MATHEMATICS AND INFORMATICS
MODELLING AND DATA ANALYSIS MASTER'S STUDY PROGRAMME

Master's thesis

Electricity Price Forecasting in Lithuania

Elektros kainos prognozavimas Lietuvoje

Greta Kvietkauskaitė

Supervisor: Habil. dr. Rimas Norvaiša

Vilnius, 2021

Table of contents

Figures

Tables

ABBREVIATIONS

ACF	Autocorrelation Function
AIC	Akaike Information Criterion
ARIMA	Autoregressive Integrated Moving Average
kWh	Kilowatt Hour
MAPE	Mean Absolute Percentage Error
MAE	Mean Absolute Error
MWh	Megawatt Hour
PACF	Partial Autocorrelation Function
PFP	Percentage filter on prices
RMSE	Root Mean Square Error
RFP	Recursive filters on prices
SFP	Standard deviation filter
TFP	Threshold filter on prices
TWh	Terawatt-hours

ABSTRACT

Electricity Price Forecasting in Lithuania

As a result of worldwide deregulation of electricity markets, electricity can be bought and sold like any other commodity. Considering the increase of renewables, decarbonatization, upcoming deregulation of Lithuania, forecasting electricity price is an important topic in Lithuania, which needs to be analysed. So far, only a few researches about electricity price forecasting in Lithuania were done, meaning that further analysis is needed. The short-term electricity price forecasting model can be used by electricity generators, suppliers, traders, and end customers (mostly large customers). This work aims to build a statistical AR-type time series model for short-term electricity price forecasting in Lithuania using external variables. Electricity price displays high volatility, spikes, and double seasonality, making it difficult to achieve high accuracy in forecasting.

Keywords: short-term electricity price forecast; Lithuania’s power market; electricity price; Nord Pool market; ARIMA.

SANTRAUKA

Elektros kainos prognozavimas Lietuvoje

Dėl visame pasaulyje vykdomo elektros rinkų liberalizavimo, elektrą galima pirkti ir parduoti kaip ir visas kitas prekes. Atsižvelgiant į atsinaujinančių energijos šaltinių augimą, šiltnamio efektą sukeliančių dujų kiekio mažinimą pasaulyje ir artėjantį Lietuvos elektros rinkos liberalizavimą, galime teikti, jog elektros energijos kainos prognozavimas yra svarbi tema Lietuvoje. Lietuvoje yra tik keli darbai susiję su elektros kainos prognozavimu, tai tik įrodo, jog ši tema yra svarbi ir turi būti detaliau analizuojama. Elektros kainų prognozavimo modelį gali naudoti elektros energijos gamintojai, tiekėjai, prekybininkai ir galutiniai vartotojai (didžiosios įmonės). Pagrindinis šio darbo tikslas yra sukurti statistinį AR-tipo laiko eilučių modelį trumpalaikėms elektros energijos kainų prognozėms Lietuvoje, naudojant išorinius kintamuosius. Elektros energijos kaina yra nepastovi, turi didelius kainų šuolius, dvigubą sezoniškumą, dėl to yra sunku pasiekti didelį prognozavimo tikslumą.

Raktiniai žodžiai: trumpalaikė elektros kainos prognozė; Lietuvos energetikos sektorius; elektros kaina; Nord Pool prekybos birža; ARIMA.

1. INTRODUCTION

As a result of worldwide deregulation of electricity markets, electricity can be bought and sold like any other commodity. Nevertheless, electricity stands out from other commodities like grain, gold, oil, coal, or natural gas. Electricity is unique due to some distinctive attributes. The most distinct attribute is that electricity is non-storable in large quantities. Electricity generators must generate as much electricity as it is consumed at that exact moment, meaning that the electricity grid always must be balanced and well planned to prevent outages, bottlenecks in the transmission grid, or other issues.

Electricity price forecasting in Lithuania is essential at this time as a result of upcoming Lithuania's deregulation. May 2020, the Law of electricity was amended, and the regulation of retail electricity prices for household consumers will be abandoned in stages by 2023, starting from 1 January 2021 [1]. At the moment in Lithuania, just business clients can choose electricity supplier freely and choose at a fixed price and spot price from Nord Pool Power market, since Lithuania is a part of this Nordic country's electricity market. Total consumption of business clients in Lithuania in Jan-Sep 2020 was app. 72.6%, leaving app. 27.4% [2, 3] to households, which will be free to choose their supplier and price determination model.

I decided to cope with this problem by forecasting short-term electricity prices. Not only due to deregulation of households in Lithuania, but as a result of growing renewables generation and Lithuania's and its neighbouring market decarbonization, which influence short-term price fluctuations and increasing need for a dynamic and flexible market. But also due to the introduction of smart grids, smart metering, a common platform for data collection and exchange [4], which may impact electricity price. All of these up-coming changes in Lithuania are to implement Nacional energy independence strategy 2018 [5] and impact short-term electricity prices. Nevertheless, the Lithuanian electricity market is selected not only due to upcoming changes, but also, due to the lack of statistical research based on Lithuania's electricity market prices.

The precise electricity price forecasting is important for all market participants because many market players depend on electricity price trends. Price is important to electricity generators for adopting strategic and tactical decisions on how much and when to generate and sell. Electricity suppliers, electricity traders, and end customers (vast industrial customers) seeking risk minimization and profit maximization.

My thesis **aims** to build a statistical AR-type time series model for short-term electricity price forecasting in Lithuania.

The main **objectives** that will help me to reach my aim are:

1. To analyse literature that focuses on electricity price forecasting in Lithuania;
2. To analyse literature that focuses on electricity price forecasting at Nord Pool Power market;
3. To collect data and to do a preliminary analysis;
4. To detect a set of candidate explanatory variables that may influence electricity price in Lithuania;
5. Select and build understandable and easily usable short-term electricity price forecasting model for electricity price buyers and sellers;
6. Compare selected models with and without external variables using accuracy measures;
7. Compare selected models with different outlier detection method using accuracy measures;
8. Evaluate the most precise model adequacy.

For my thesis, statistical AR-type time series models were chosen due to quite high accuracy and model simplicity. Moreover, models can be further applied in selling or buying companies of the market because these models are easy to understand and implement in the short-term, and it is easy to incorporate exogenous variables. No short-term electricity price forecasting in Lithuania was done using these types of models. To improve model forecasting, accuracy external variables will be applied, which is also new in Lithuania's market perspective. Various statistical outlier detection methods will be used to deal with electricity price spikes and improve forecasting accuracy. Fourier series with daily, weekly, and yearly seasonality as the external variable will be included to deal with double seasonality.

This thesis is the first analysis of the Lithuanian electricity market that focuses on outlier detection methods.

Chapter 2 will introduce the specialties of electricity price and one of the biggest electricity market Nord Pool. In Chapter 3, papers and researches about Lithuania and the Nord Pool Power market will be analysed. Chapter 4 starts with an overview of the available methodologies to model electricity prices and explains the chosen statistical model. In Chapter 5, preliminary electricity price analysis will be done, important determinants for the electricity price will be worked out; this chapter also presents the development of the models and the results of the analysis. In Chapter 6, final remarks will be given.

In this work, a simple, understandable, and reusable model for electricity sellers and buyers was built. To the author's knowledge, this is the first work that used both outlier detection and double seasonality detection methods. To deal with price spikes, different outlier detection methods were introduced. To deal with the double seasonality, the model with the Fourier series was implemented. Models were built using seven external regressors Calendar days, Coal price, Hydro power in Sweden, Actual load in Lithuania, Natural gas price, Temperature, Wind power in Lithuania. The best fitted ARIMA model was ARIMA(3,1,2) using percentage filter on price outlier detection and Fourier series included as an external variable, together with other seven external variables. This model RSME reached 4.82 and MEA 3.04, which was assumed to be quite accurate for statistical models.

2. ELECTRICITY MARKET

2.1. Electricity price

There are special features of electricity that sets it apart from other commodities. As it was mentioned in an introduction, electricity cannot be stored in a huge quantity. So, the grid's continuous electricity flow is required, and the balance between electricity demand and supply should always be secured. This specific of electricity leads to the requirement of reserve capacity in an electric power system [6, p. 25]. In Lithuania, a short-term reserve was always secured by Kruonis PSHP, but from 2021 this unit will be just one of the market's players to provide balancing services. In contrast, the long-term shortages must be secured with Elektrenai Complex units – 7^{th} and 8^{th} blocks. So, in case of an emergency, these units must be ready to work.

Another distinct feature of electricity as a commodity is the need for the electric energy transmission infrastructure or so-called electric power network. From that point of view, electricity may be considered a network-based commodity [6, p. 25]. The high voltage grids are operated by the transmission system operator (TSO), in the Lithuania we have AB "Litgrid". Despite deregulation and government control reduction, the transmission is kept under government control.

Electricity is one of the most volatile commodities. The daily average electricity spot price change in Lithuania can vary from negative to app. 200 Eur/MWh, while a yearly average price is app. 40 Eur/MWh. Price variations usually relate to an unexpected increase in demand, shortages, over-production from wind farms, or the failures and outages of the transmission infrastructure [3, pp. 1-2], so companies use bidding or hedging strategies to cope with price volatility.

Renewable energy and its volatility due to high dependence on temperature are one of the key variables for price spikes and are becoming more and more relevant due to increasing production from renewables. In Lithuania, net generation from renewables increased by 11.1% in Jan-Sep 2020, in comparison to the same period last year. Nevertheless, that total generation from renewables decreased from 80.1% in Jan-Sep 2019 to 63.9% in Jan-Sep 2020, influenced by the higher generation of gas-fired Elektrenai Complex [2]. Renewables growth is not only Lithuania's but also, and the world's one of the main goals, and it is expected to grow exponentially, despite some delays in 2020, due to Covid-19 [7]. Another important goal in the world is decarbonization. As a result of CO_2 decrease, more renewables will be needed to generate the same amount of electricity.

Electricity price has high volatility, outliers, non-constant mean and variance, seasonal and calendar effects that make electricity price forecasting quite a hard task. Despite electricity characteristics, various external indicators are expected to do an influence and improve electricity price forecasting.

Not only growing renewable energy but also and other drivers are affecting the prices on the market, such as temperature and wind power and its forecast, as well as power plant availability and transmission congestions. Also, it is seen that in the long run, electricity prices on the Nord Pool market are significantly influenced by the water level in the reservoirs of the Norwegian and Swedish hydro power plants [6, p. 18].

Despite electricity volatility and inconsistency, its demand is relatively inelastic. This means that if the electricity price will suddenly spike over 200 Eur/MWh, the electricity demand will not change, and it will stay relatively the same. The demand is highly dependent on unforeseeable factors such as weather or climate. Also, electricity displays seasonal patterns due to economic activity and weather conditions. Seasonality can occur on various levels, including an hourly, daily, weekly, or yearly seasonality [8, p. 53].

Another important event for electricity price in Lithuania is the up-coming deregulation. Deregulation of the electricity market is one of Lithuania's EU commitments. Most countries in the EU and the Nordic region in which Lithuania participates in the electricity market have already deregulated the electricity market for household consumers. Estonia fulfilled these commitments in 2013, and Latvia followed in 2015 [4]. Around 27.4% of Lithuania consumer's electricity prices are vertically integrated, which means that these electricity prices are regulated, and the consumers are offered predetermined tariffs. While in deregulated markets, market participants have more freedom, and they have the option of trading electricity, which leads to market production efficiency and competitiveness [6, p. 18]. Therefore, after Lithuania deregulation, consumers will have many options for choosing an electricity supplier (currently, there are eight possible candidates instead of one [9]). Lithuania is a part of the Nord Pool market, and most of the electricity is bought and sold there.

2.2. Nord Pool power market

The Nordic region has reasonable experience in deregulated electricity markets. In 1993 the market was established in line with Norwegian parliaments to deregulate the market for electrical energy trading. The derivatives and energy markets were separated in 2002 to ascertain the Nord Pool Spot power exchange market. The Nordic power market was fully established in 2000 once the regional electricity markets of Sweden, Norway, Finland, and Denmark merged. Nord Pool Spot currently operates in 13 countries Norway, Denmark, Finland, Estonia, Lithuania, Latvia, UK, Austria, France, the Netherlands, Sweden, Germany, and Belgium. Lithuania entered this market in 2012 [10]. Nord Pool is an initiative owned by different European TSOs participating in the market, including the Lithuanian TSO Litgrid.

Nord Pool Spot main objective is to balance the production with the electricity demand, precisely and at an optimal price, which is, by equilibrium point. The optimal price represents the cost of 1 kWh generation of electricity from the high-priced source requiring a balance of the system. Two different physical operation markets are organized in Nord Pool Spot: Elspot or day-ahead and Elbas or intraday. Elbas market is where up to an hour before the distribution generators and suppliers can upgrade the quantity of electricity traded. This market is crucial if, for example, a nuclear power plant will suddenly stop operating, or strong winds may cause higher wind power generation than expected. The significance of the intraday market is growing as more wind power enters the grid. Wind power generation is incalculable by nature and differs concerning day-ahead contracts. Therefore, produced volume often needs to be offset. While the Elspot market where day-ahead before the delivery producers and suppliers must update the quantity and the price. This way of pricing is called a double auction because both the seller and buyer submit bids. In this market, forecasting and its accuracy are important to minimize the risk and maximize the profits. In this work, I will focus just on day-ahead prices. So, the Elspot market will be discussed and further used in this paper.

To clarify the difference between the day-ahead and intraday market, some more details will be concluded. The day-ahead price forecast for day X is required on day X-1 (official bidding: 12:00 CET). As soon as the 12:00 CET deadline for electricity suppliers and electricity generators or traders to submit orders has passed, all purchase and sell orders are aggregated into two curves for each delivery hour of day X. The system price for each hour of day X is determined by the intersection of the aggregate supply and demand curves, representing all bids and offers for the entire Nordic region, and are published a bit later that day by the system operator day X-1. Hence, actual price data up to 24 hours of day X-1 are available on day X-2. Therefore, when bidding for day X, price data up to hour 24 of day X-1 are considered known. As a result, the actual forecast of day-ahead prices for day X can take place between the clearing hour for day X-1 of day X-2 and the bidding hour for day X of day X-1. A detailed description of how a day-ahead market in the Nordic region works can be found on the Nord Pool website [10].

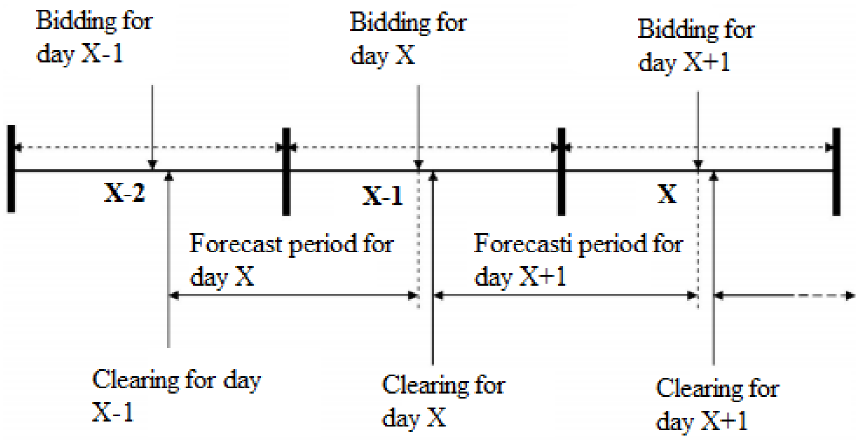


Figure 1. Time framework for Nord Pool system price forecasting in the day-ahead energy market

Nord Pool operates as one market in which supply to a region is aggregated, and generators are transmitted to satisfy the demand with as low price as possible. When the exchange's electricity prices are computed, the purchase tenders are aggregated to a demand curve. The sale offers are aggregated to a supply curve. The spot price is set at the point where the two curves intersect if there's neither market coupling nor market splitting. After the market participants have submitted their orders, an equilibrium between the aggregated supply and demand curves is established for all bidding areas, and a system price is calculated based on the sale and purchase orders. Once the market prices have been announced, the market participants receive a notification of their bids and the following operating day's hourly commitments. All producers are paid according to the calculated bidding area price, and similarly, all consumers pay the same price.

The Baltic and Nordic regions are divided into bidding areas by the relevant TSO to operate the electricity grid's congestions. Bidding areas can have a deficit, surplus, or balance of power. If the transmission capacity between bidding areas is not enough to reach full price convergence across the areas, congestion will lead to bidding areas having different prices. Hydroelectric generation and nuclear producers have relatively high start-up costs and low marginal costs of generation. Gas turbines and oil-fired plants have a comparatively high marginal cost of generation, used typically for peak periods only as a reserve. Wind power or other renewable sources are less predictable than conventional ones, leading to price drops and unexpected variations in wind power generation that may increase the electrical grid's operating costs. The increased deployment of renewable energy sources has made the dynamics of the electricity spot prices even more complex, with more extreme prices that are very difficult to predict.

3. LITERATURE REVIEW

In this Chapter literature review about price forecasting in Lithuania and the Nord Pool market will be made. A brief introduction of what was done, and the main findings of the research will be analysed and compared.

3.1. Previous work review on electricity price forecasting in Lithuania

Focus on Lithuania electricity price forecasting is a rather new branch in the literature because the earliest work that I succeeded in finding was written in 2017. Only four works were found, which proves the lack of statistical electricity price forecasting in Lithuania. Meanwhile, in the world, the number of proceedings papers, articles, and citations on electricity price forecasting has increased significantly over the past 20 years [11]. The literature mainly focuses on spot price forecasting in Lithuania using Nord Pool data for one year. Methods used are various, starting from classical AR, and smoothing methods to deep learning recurrent neural networks method. Literature mainly focuses just on electricity price itself, and no external variables were introduced to the models. I will briefly introduce the findings starting from the oldest to the newest work about electricity price forecasting.

In their report, R. Beigaitė and T. Krilavičius [12] used short-term electricity price forecast using average, seasonal naïve, and exponential smoothing methods were constructed. The data set consists of historical hourly electricity prices (Eur/MWh) from 1 January 2014 to 31 December 2016 using electricity spot price data of Lithuania's price zone in the Nord Pool power market, with no external variables included. Methods were used for short-term day-ahead prognosis of 24 hours. Among the three smoothing methods exponential smoothing method was the most accurate, with minimum values equal to 1.76% MAPE, 0.66 MAE, and 0.83 of RMSE, while the mean values were 16.03%, 6.49, and 8.64, accordingly. Concluding that statistical models do not perform well when capturing electricity price spikes.

R. Beigaitė and T. Krilavičius [13] next year publish another report using more advanced methods. The authors again try to construct a short-term electricity price forecast of Lithuania's price zone in the Nord Pool market by using recurrent neural networks: Elman and Jordan methods. The data set consists of historical hourly electricity prices (Eur/MWh) from 1 January 2016 to 31 December 2017, with no external variables. Forecasting experiments were performed for each day of the year 2017. Data of the year 2016 was used for training, and data of the year 2017 was used for testing to capture seasonality effects, and for the input, features lagged electricity prices were used. Methods were used for short-term day-ahead prognosis of 24 hours. The results showed that the highest average accuracy during forecasting experiments was achieved using Elman neural network with MAPE error equal to 3.55%, 1.12 MAE, and 1.34 RMSE, while mean values were 18.12%, 6.85, and 8.54. From their research, it is visible that in the Lithuania simple statistical exponential smoothing model performs better than the Elman neural network.

A. N. Tat [14] was forecasting Lithuania electricity price using Monte Carlo simulation. The Ornstein-Uhlenbeck process with and without catching price spikes to predict the next day's prices was used. The data set consists of historical hourly electricity prices (Eur/MWh) from 1 June 2017 to 30 November 2017. Data was collected and aggregated from the Nord Pool market data, using Lithuania system price. Results in this research showed that Ornstein-Uhlenbeck process while catching price spikes was better. No forecasting accuracy measures were introduced, and the author used a concise period, so it may be possible that data could not learn yearly seasonality.

In his research, M. Česnavičius [15] constructed the long-term electricity price forecasting model based on univariate ARIMA models using past time series values and error terms. His average monthly electricity price (Eur/MWh) range was from 1 July 2012 to 31 Decem-ber 2019, and a forecast was made for 2020. Data was collected from the Nord Pool market data, using Lithuania system price. Four different ARIMA models were selected during his research: AR (1), ARIMA (1,1,0), ARIMA (1,1,1) and SARIMA (1,1,1). After analysis, the additional fifth weighted SARIMA (1,1,1) model was introduced. The forecasting accuracy was compared using RMSE, MAE, MPE, and MAPE forecasting error statistics. The best long-term forecasting method was AR with 4.13 RMSE and 7.94% MAPE. Since prices for 2020 were forecasted, we can conclude that it was not accurate and prices differ almost by 30 Eur/MWh, because in 2020, extremely low prices in Lithuania were captured due to significant excess of water in hydro reservoirs in Scandinavia and especially in Norway, meaning that external variables should be included and that previous price forecasting using ARIMA models for a long period is not the best fitting model for a long-term forecasting.

To conclude the researches that are done about Lithuania electricity price forecast, it can be said that the best model was not found, and accuracy may be lower because all models used only historical electricity price as the main input. To achieve higher accuracy and to build a better model, external variables must be amended. Also, most of the research included only from half of the year to one year and a half dataset, which is not enough to capture yearly seasonality.

One Lithuanian author master's thesis about electricity price forecasting was found. The forecast was not for Lithuania electricity price, but for France. A. Bagdonov [16] wrote a master thesis about France's electricity market Powernet and used SARIMA-TGARCH and SARFIMA-TGARCH models.

Also, one interesting theoretical model for electricity market price forecasting was done by V. Bobinaite el. at. [17] they analysed structure, price features, main supply and demand indicators, and methods that could be used to forecast electricity price.

After analysing works related to electricity price forecasting in Lithuania, significant improvement can still be made since this theme is not widely analysed yet. The biggest contribution for electricity price forecasting in Lithuania is R. Beigaitė, T. Krilavičius works. They applied both statistical and neural network methods, concluding that the statistical one works better for short-term electricity price forecasting. External variables, a method to detect price spikes and/or seasonality, should be used and implemented.

3.2. Previous work review on electricity price forecasting in Nord Pool Power market countries

Since works published in Lithuania are minimal, and further improvement needs to be done, I checked electricity price forecast articles, reports, and books from 2000 and further. I concentrated only in Nord Pool market countries and situations in this area. Numerous studies were done, so only a few were analysed. Many different models from statistical to advanced deep learning or hybrid models were used for electricity price forecasting. The Lithuanian market's main difference is that almost all these researches used external variables to improve forecasting accuracy. It can be pointed out that numerous researches were done to detect price spikes, but none of the found research used any kind of outlier's detection or elimination method combined with the forecasting model.

O.A. Karabiber et al. [18] analysed the Danish Nord Pool electricity market. They have presented Non-seasonal ARIMA, Trend and Seasonal Components (TBATS), and Artificial Neural Networks (ANN) methods, using seven exogenous variables: temperature, consumption prognosis, production prognosis, wind prognosis, oil price, natural gas price, and hydro reservoir. They applied models capturing double seasonality effects. Among the three individual models, the best performance in terms of mean error is provided by ARIMA with 7.95 RMSE. They reached significant improvement by 1.6 in RMSE with backward variable elimination when the temperature was excluded from the external regressors, leaving only five external variables. This finding suggests that temperature is not a significant indicator of price forecasting, and double seasonality detection methods should be tried to improve further forecasts. T. Kristiansen [19] also build an autoregressive model based on R. Weron et. at. [20] work, but reduced terms of estimation parameters and modified to include Nordic demand and Danish wind power as exogenous variables with MAPE ranging from 8% to 11%. In this work, electricity spot prices were lagged by 168 hours, assuming that weekly seasonality has a higher impact than daily.

S. Duffner el. at. [21] analysed Germany electricity price using linear regression, ARIMA(X) and (M)GARCH models. They used various external variables the price for CO2-certificates, Crude Oil, temperature, planned wind and solar feed-in, the planned and unplanned non-availabilities for the different kinds of power plants and the total planned power generation for all other power plants that are part of the EEX transparency system. They concluded ARIMAX method works best with 6.6 RMSE and that a model based on 24 individual time series works better than one-time series which includes all consecutive hours because computation time is far less for the former and because hours with high volatilities like the early morning hours do not interfere with other hours.

J. S. Roungkvist et al. [11] analysed high-resolution electricity spot price forecasting for the Danish power market. In the empirical study, hourly data of a one-year period was used to perform a linear regression model. External variables for accuracy improvement were included, such as generation, consumption, price data from the year before the month of interest, the exchange between the price area and the nearby price zones. The final model of the electricity spot price with electricity generation, consumption, and previous prices as explanatory variables provided an R-squared on average of 0.73 and an RMSE of 4.26. An interesting discovery was made that the transmission grid's capacity is not as significant as the electricity generation and the electricity spot price the year before and that the consumption is only statistically significant in one of the four cases, due to strong electricity dependence on the demand. J. Peljo [22] in his master's thesis also applied regression analysis. He concluded that above-average stored energy in water reservoirs and hydro plants harms spot prices and below-average values positively impact them. According to the thesis, rising coal prices and electricity demand also caused the spot price to rise.

B. Amor et al. [23] proposed a new hybrid model k-factor GARMA-LLWNN model using the hourly log-returns of electricity spot price from the Nord Pool market, which was applied for various periods and results was similar to the real values, with MAPE < 1%. It is a great win for a researcher because the model has not included any external variable, and such a high accuracy was achieved. Another more advanced research was done by K. Rahimimoghadam el. at. [24] in their article, neural networks and wavelet transformation were used. Few stages for this work were used. Their research concluded that forecast without temperature is better for weekend data, while during the working day's temperature gives lower MAPE app. 1.22%-2.01%.

Table 1 shows that various models and various explanatory variables are analysed, and it really depends on the researcher. However, as can be seen, that all works use a statistical model as a base model. Also, explanatory variables were checked, concluding that electricity has a high impact on calendar days, hydro power, actual loads, and saying that temperature is not significant to the model. Even though statistical models are represented as a base, in most cases, the hybrid models, while forecasting Nord Pool power marker data, work better, especially new hybrid k-factor GARMA-LLWNN model used by B. Amor et al.

Table 1. Comparison of various models, inputs, and accuracies applied in electricity forecasting in Nord Pool market countries

Reference	**Market**	**Explanatory variables**	**Models**	**Accuracy Measures**
O.A. Karabiber et al. [18]	Nord Pool	Temperature, consumption prognosis, production prognosis, wind prognosis, oil price, natural gas price, and hydro reservoir	Non-Seasonal ARIMA, Trend and Seasonal Components and Artificial Neural Networks	RMSE, MAE
J. S. Roungkvist et al. [11]	Nord Pool	Consumption, production (wind, solar and thermal power), price the year before, transmission line capacity, physical exchange on transmission lines	Multivariable Linear Regression Analysis	MAE, RMSE, and MAPE
S. Voronin [6]	Nord Pool	System demand, hydro power, nuclear power	Numerous models, starting from classical to hybrid	MSE, MAE, and MAPE
S. Duffner el at. [21]	EEX	CO2-certificates, Crude Oil, temperature, wind and solar feed-in, planned and unplanned non-availabilities and power generation	Linear regression, ARIMA(X) and (M)GARCH	MAE, RMSE, and MAPE
T. Kristiansen [19]	Nord Pool	Nordic demand and Danish wind power	AR with exogenous variables	MAPE and WMAE

E. Raviv et al. [25]	Nord Pool	Daily average price	AR and Heterogeneous Autoregressive model (HAR), VAR, Bayesian VAR	MAE, RMSE and MAPE
H. Torró [26]	Nord Pool	Temperature, precipitation, reservoir levels, power load and basis	ARIMAX	MSE
B. Amor et al. [23]	Nord Pool	-	k-factor GARMA-LLWNN model	R^2, MAPE and Logarithmic Loss function
K. Rahimimoghadam el. at. [24]	Nord Pool	Temperature	Wavelet Transform	MAPE

Not only electricity price forecasting but also other problems in the Nord Pool energy system was analysed. O. Knapik [3] discovered principle price drivers in the Elbas market and build three models for electricity price spikes forecasting. P. Spodniak, et al. [27] investigated the Nordic day-ahead and intra-day trading behavior and price. P. Pinson el. at. [28] investigated predictive densities for determining the optimal structure of block bids. In his dissertation, S. Voronin [6], analysed models that could predict price spikes in the Finish Nord Pool Elspot market. One of his findings was that Box-Jenkins models could not to estimate high volatility and spike clustering presented in the original price series and he continues other methods analysis. E. Raviv et al. [25] paper illustrates that the disaggregated hourly electricity prices hold convenient predictive information for the daily average price.

4. EMPIRICAL PART

Despite the diversity of existing models, it is impossible to select the most reliable one. Also, many models were already implemented and applied for electricity price forecasting. From previous analysis were clear that the highest accuracy has hybrid or computational intelligence models. However, they are more complicated than simple statistical models, which can be easily implemented in any market buying or selling company to check the upcoming forecast with quite low accuracy.

Based on Weron [29], V. Bobinaite al.et. [17] models, applied for electricity price forecasting, can be classified into five groups: 1) multi-agent (Nash-Cournot framework, supply function equilibrium, agent-based simulation models, strategic production-cost models); 2) fundamental (parameter-rich fundamental and parsimonious structural models,); 3) reduced-form (jump-diffusion and Markov regime-switching models); 4) statistical (exponential smoothing methods, regression models and AR-type or ARX-type time series models, and GARCH-type models); 5) artificial intelligence (feed-forward neural networks, recurrent neural networks, fuzzy neural networks, support vector machines). All these models aim to characterize the trajectory of the spot price or return across time.

For my thesis, statistical ARIMA time series models were chosen, due to quite high accuracy and model simplicity. Models can be further applied in selling or buying companies of the market because these models are easy to understand and implement in short-term and it is easy to implement and incorporate exogenous variables. No short-term electricity price forecasting in Lithuania was done using these types of models. To improve model forecasting accuracy, external variables will be applied, which is also new in Lithuania's market perspective. Various statistical outlier detection methods will be used to deal with electricity price spikes and improve forecasting accuracy.

4.1. Autoregressive integrated moving average

Box and Jenkins in 1976 introduced the Autoregressive integrated moving average (ARIMA) model, which has become one of the most frequently used and recognized forecasting models due to its simplicity [30]. To use data with the ARIMA model, first of all, data must be stationary. Stationarity proposes that the trend and the seasonality of the data should be eliminated before applying the ARIMA model [18], so they do not depend on the time at which the series is observed [31]. ARIMA models are combined from three main parts, which are seen in the name itself: autoregressive (AR), integral (I), and moving average (MA). The first part AR refers to using the lagged inputs to forecast future data. The integral (I) part refers to the number of differencing because after estimating of the models, the data need to be integrated to reverse the initial differencing. MA is like AR, except that instead of inputs, the past errors are forecasted [31]. It is often impossible to decide the ARIMA model, when the seasonal adjustment order is high, or its diagnostics cannot indicate that time series stationarity after seasonal adjustment. In these cases, the static parameters of ARIMA model are examined. Another constraint of ARIMA approach is that it needs many observations to acknowledge the best fit model [32]. ARIMA is usually seen as ARIMA(p,d,q), p is a representation of a number of autoregressive terms, in other words, a lagged imputation points. AR models assume that Y_t is a linear function of the preceding values, where each observation consists of random component ε and a linear combination of the previous observations, ϕ_1 is a self-regression coefficient, AR(1) can be noted [33]:

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t \tag{1}$$

d is a number of differences used or the integrated process. The differencing order corresponds to the degree of series trend first-order differencing accounts for linear trends, second-order differencing accounts for quadratic trends, etc. Although short-term values may fluctuate with large contingencies around the mean, the level of the series over the long term will remain unchanged [32]. An integrated process is defined:

$$Y_t = Y_{t-1} + \varepsilon_t \tag{2}$$

where ε_t is white noise. Finally, q represents MA order – it is a number of lagged errors. MA orders specify how deviations from the mean for previous values are used to predict current values. MA(1) process is defined as following [33]:

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} \tag{3}$$

Note that ARIMA(p, 0, q) is simply an ARMA(p, q) process [29]. It is common to denote ARMA(1,1), given by

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \tag{4}$$

In the ARIMA model, input variables can be differenced as many times as needed to make the data stationary, and the model can be extended to improve forecast, using ARIMAX, where X here stands for and external variables.

The Box–Jenkins principle consists of five steps: data preparation, model identification, parameter estimation, and diagnostic checking and forecasting [34]. Data preparation will be discussed in further section of *Practical implementation*. To simplify model identification, which means to find p,d,q values, stationarity tests should be applied. The assumption of stationarity is needed for time series analysis. Otherwise, the relationship between two variables would change arbitrarily, and correlations between the two in a regression analysis could not be tracked. To identify the model, we should generate a stationary time series. Kwiatkowski-Phillips-Schmidt-Shin [35] unit root test can be used to identify if the series is stationary and choose the correct d parameter. For p and q parameters, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are used, where ACF is the order of AR(p) and PACF is the order of MA(q). To calculate an autocorrelation of lag k, the correlation between Y_t and Y_{t-k} as follows [33]:

$$ACF\ (k) = \frac{Cov(Y_t\ ,\ Y_{t-k})}{\sqrt{Var(Y_t)Var(Y_{t-k})}} \tag{5}$$

where Y_t is the input, Y_{t-k} is a lagged version of input. While PACF measures the association between Y_t and Y_{t-k} controlling possible effects of linear dependence among values at lags [33]:

$$PACF = \beta_k, \text{ where } \ Y_t = \alpha + \beta_1 Y_{t-1} + \cdots + B_k Y_{t-k} + \varepsilon_t \tag{6}$$

After, we can continue with the further step of Box-Jenkins – parameter estimation. And finally, model diagnostics should be done. The model hypothesis regarding the errors must be fulfilled. The diagnostic statistics and plots of residuals can be used to check the model's goodness of fit. Other plots that can be used are histogram, normal probability plot, and time sequence plot. If it does not look adequate after this, we must come back to the first step. When a model is computed, the second, third, and fourth Box-Jenkins development process steps are no longer repeated, and the selected model will be used for forecasting purposes [36].

4.2. Forecasting

Forecasting involves taking mathematical models to fit on sample data and using them to predict the future. In statistical handling of time series data making predictions is called extrapolation [37]. Electricity price forecasting can be split into 3 categories based on time horizons. Despite that in literature, there is consensus at what point the threshold should be. The categories are as follow [29] [38] [17]:

1. Short-term forecasts. Forecasts from a few minutes up to a few days or a week. The market players mainly use them to maximize profits in the spot markets.
2. Medium-term forecast. Forecasts from a few days to a few months ahead. They might be preferred for balance sheet calculations, risk management and allow the successful negotiations of bilateral contracts between suppliers and consumers.
3. Long-term forecast. Forecasting period can vary from a few months up to a few years. Such forecasts might influence the decisions on transmission expansion and enhancement, generation augmentation, and distribution planning.

Forecasting time series is quite a difficult task, and it needs a lot of various assumptions since there are many different approaches, and aims for modelling [39]. To detect forecast accuracy and how well the model can forecast, various error measurements can be conducted. As it was analysed in the literature part, the most popular as widely used accuracy measures are root mean square error (RMSE), mean absolute error or mean absolute deviation (MAE), mean absolute percentage error (MAPE). They are calculated as following [40] [17] [6]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}t)^2} \tag{7}$$

$$MAE = MAD = \frac{1}{n}\sum_{t=1}^{n}|F_t - A_t| \tag{8}$$

$$MAPE = \sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| x \frac{100}{n} \tag{9}$$

where n is the number of test data, $y_t$is the actual value, and $\hat{y}_t$ is the predicted value.

The MAPE measure works well in load forecasting since load values are significantly higher than zero, but MAPE can be misleading when applied to electricity price accuracy measuring. It is so because when electricity prices are close to zero, MAPE values become very large, regardless of the actual absolute errors, so this particular forecasting error measure has to be treated with caution [29]. As a result of these findings, only RMSE and MAE forecasting accuracy measures will be used.

5. PRACTICAL IMPLEMENTATION

The practical implementation part consists of three main parts: data collection, construction of different forecasting models and finally, forecasting evaluation and results. In the first part, reliable electricity price data and other variables are selected, and a primary statistical analysis is done with general observations about time series. The second part explores these Box-Jenkins steps: model identification, parameter estimation, and diagnostic checking. Finally, the third part is forecast evaluation using different accuracy measures for the best model identification and model adequacy interpretation.

All experiments were computed using statistical package R (*https://www.r-project.org*).

5.1. Preliminary data analysis

After collecting the data, preliminary data analysis is required to understand the data and its patterns better. This will let to find good models that fit our data better. In this work, Lithuania's price zone in Nord Pool Elspot market is analysed. The dataset was collected from the official Nord Pool website [10]. Dataset consists of historical hourly electricity prices (Eur/MWh) from January 1, 2015, to October 2, 2020, in total of 50,448 rows.

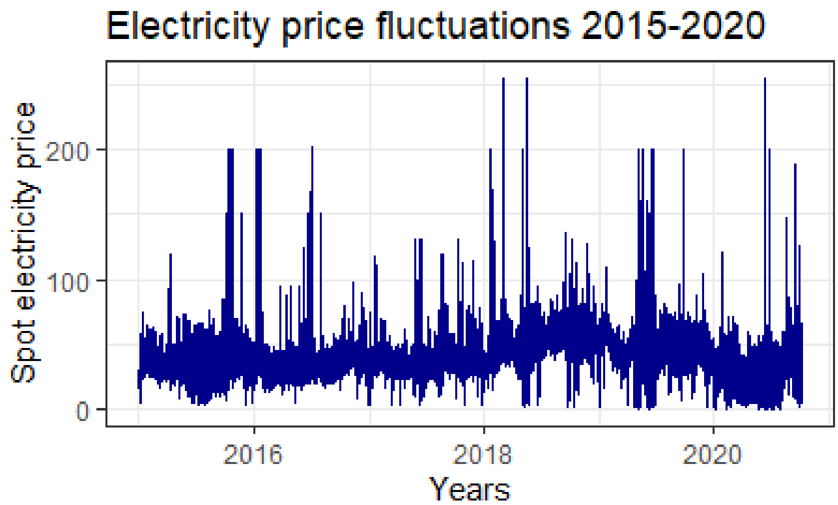


Figure 2. Hourly electricity price in Lithuania from 1 January 2015 to 2 October 2020

From the graph we can see that electricity price in Lithuania in long-term period is quite stationary, yearly mean varies from 31.56 Eur/MWh in 2020 to 50.00 Eur/MWh in 2018, while during this whole period electricity price mean was 40.58 Eur/MWh. Yearly patterns are not visible in the long run, due to price fluctuations. But we can see an increase in extreme prices since 2015, due to in growing renewables, various test runs, and breaks in electricity transmission grids. To see yearly patterns each year data was analysed.

Yearly data looks quite stationary (see Appendix 1), but we can see that electricity price during the winter tends to be higher than during the summer. It can be noted that from 2015 electricity price spikes rose, the most price spikes are seen during the summer period due to test runs and more electricity generated from renewables. Some price spikes are indicated during the winter, most probably due to losses in the transmission grid. Maximum electricity price was detected in 2018 and 2020, reaching 255.03 Eur/MWh. Fist time from 2015, on the 6th of June 2020 at 4 a.m., negative electricity price was detected, reaching -0.09 Eur/MWh. First, negative prices indicate an oversupply of electricity in the grid, which means that the first time during this period in Lithuania was generated more electricity than the market needed. Traditional financial models that only allow non-negative prices cannot take this feature into account [31]. Concerning this, we can assume that 2020 was a record year, and prices dropped due to of a significant excess of water in hydro reservoirs in Scandinavia and especially in Norway. The situation highlighted several aspects: on the one hand, the whole Nord Pool power market region could obtain cheaper electricity. On the other, record amounts of the same water were discharged from reservoirs due to insufficient electricity capacity to take full advantage of the opportunities. The SE4-LT connection also gives the Baltic region access to cheaper energy, but the bandwidth also limits the possibilities of how much cheaper energy we can get. It is likely that in the future, years like 2020 will only increase due to the warming climate, and Scandinavia will increasingly be able to offer cheaper stored hydropower. After the yearly seasonality check, weekly and daily patterns will be disclosed.

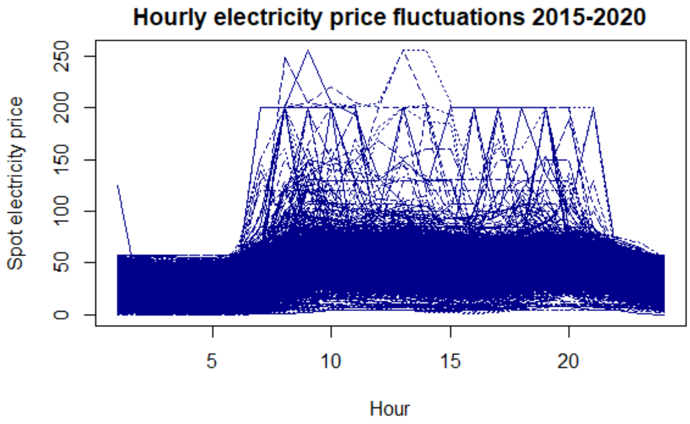


Figure 3. Daily electricity price fluctuations

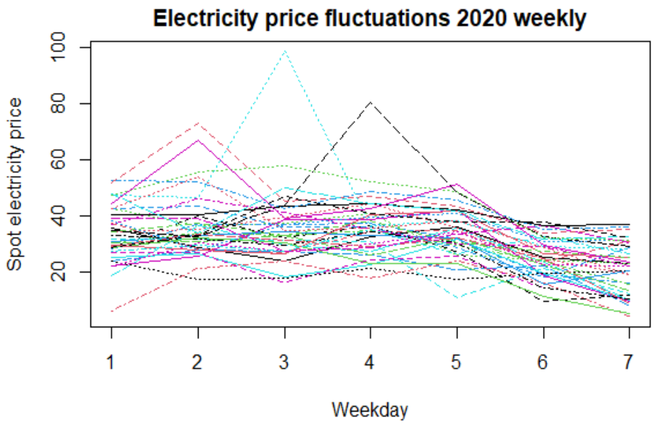


Figure 4. Weekly electricity price fluctuations

In Figure 3, hourly patterns are seen. This time is visible that electricity price tends to be lower at night and higher and more volatile during the day as a cause of higher electricity consumption of any production company at their working hours. The highest price was at 12 a.m. 255.03 Eur/MWh, the lowest negative price was at 4 a.m., reaching -0.09 Eur/MWh. While the highest mean was electricity price was 47.70 Eur/MWh at 8 a.m., which could be related to the production start, when all the machines need more power, while the lowest mean was detected at 3 a.m. with 26.57 Eur/MWh. Similar patterns followed for the same reason are seen and during the week. 2020 data were split, and we can see that during the day, electricity price is more volatile than at the weekends. The highest price was on Thursday, and the lowest on Monday, while the highest mean price was on Wednesday at 41.94 Eur/MWh, and the lowest mean price was on Sunday at 32.61 Eur/MWh.

Table 2. Lithuania electricity price: negative price jumps, normal prices, and positive jumps (higher than 100)

	Negative jumps	Normal prices	Positive jumps
Number	1.0	50,113.0	334.0
Frequency	0.0%	99.3%	0.7%

Explanatory variables (for graphical representation see Appendix 2), also referred to as external variables or regressors, are the additional effect data used in the regression models to get a lower forecasting accuracy error. In Chapters 2 and 3, the most informative and widely used features were analysed, and according to their public availability were selected. In this paper, eleven explanatory are considered as external regressors. All variables are from January 1, 2015, to October 2, 2020. Explanatory variables fall into two categories [17]. The first set of explanatory factors is demand-side factors and indicators. Hourly electricity consumption shows people’s life patterns, mechanical systems interacting with weather, cloud cover, timing of sunrise and sunset, water temperatures, and other similar factors. As a result, the calendar days variable was generated from the calendar, where 1 represents working days, and 0 represents non-working days and holidays. The second set of explanatory factors is supply-side factors and indicators. These indicators include changes in fuel prices, production unit availability, temperature, wind speed and hydro flows in some markets. In periods of high demand, the load levels in surrounding areas can significantly impact Lithuanian prices, reflecting the high price of imported energy. As a result, the actual load in Sweden and Finland was included in the model. Due to the high electricity exchange with Sweden [10], Finland was selected as a Nordic countries representative that may have an impact. Data was collected from the Transparency platform [41]. From the Transparency platform, few more variables were selected: Hydro power in Sweden and Lithuania, actual load in Lithuania, and unavailability in the transmission grid. From Intercontinental Exchange [42], Dutch daily natural gas prices and coal prices were taken, assuming that the Netherlands are in the same Nord Pool market, and the fuel prices do not differ significantly. The hourly average temperatures in Lithuania were collected from Lithuania hydrometeorological service [43]. And finally, hourly wind power in Lithuania was included in the primary model. Some of the variables were not hourly, but they were aggregated accordingly to fit into the hourly data set. After graphical representation, unavailability in the transmission grid was eliminated because in 2016, the data was missing, and it was assumed that it could affect forecasting results. It should be noted that when evaluating forecasting accuracy errors, errors that could be in the explanatory variables are omitted.

It is important to highlight that the seasonality of hydro power reservoir levels is also an important factor explaining electricity prices and can be considered an important factor. If reservoir levels are not enough to satisfy demand, electricity prices will probably increase together with power imports [3, p. 16]. Also, a recent significant excess of water in hydro reservoirs in Scandinavia and especially in Norway showed that this indicator influences electricity price not only in Lithuania, but in the whole Nord Pool market region.

5.2. Data preparation

Before applying data to the model, some data preparation steps were made. Since data is from the real world, much preparation needs to be done. Firstly, missing data analysis was done to see if this can impact the model and forecasting accuracy. For this step, different missing value imputation methods have been applied. In total, there was only 0.4% of missing data. Most missing data had coal prices, gas prices, and temperature. After the combinations of missing data in all the variables were done, and no clear dependency was detected. Also, the seasonal patterns of missing data were analysed. Yearly, monthly, weekly, and daily missing data patterns were compared, and some interesting results were found. Despite that, they are not significant to the model. First that most data are missing during the winter period, and early spring and secondly, also it was found out that electricity price and wind power variables have their missing values always at the same hour each year, meaning that there is an impact of time changing in different time zones and some systematic mistake in Nord Pool system due to this reason. Because electricity price was always missing in March when the summertime is introduced, and wind power data is missing in October when wintertime is introduced. A short-term forecasting study by missing values omitting may bias model estimates and is not an option when producing operational forecasts [44], so the different mean imputation methods were applied to all the variables.

After some recent energy sector studies were checked and various missing value imputation methods have been proposed for various indicators forecasting, starting from energy consumption to generation and others. The most used method that I found was a linear interpolation method (LI) K. Zor el. at., 2018 [45], T.Kim el. at., 2019 [46] and S. Jung el. at., 2020 [47]. R.Tawn el. at., 2020 [45], in their research, used various models such as complete case, inverse probability weight, Mean Imputation (MI) and multiple imputation. K.Zor el. at., 2019 [44] used LI, and MI. T.Kim el. at., 2019 [46] used LI, Mode imputation, Imputation Using KNN, Multivariate imputation by chain equation (MICE), and S. Jung el. at., 2020 used more advanced methods. Since electricity data is missing not random, care should be taken to use an appropriate missing data technique. As my goal is to build a simple, usable, and understandable model the LI, MI, and MICE missing value imputation methods will be used. LI is an example of interpolation that exploits a straight line, connecting two data points outside the missing data interval. MI is a simple missing values replacement with the mean variable. This method is simple to implement and preserves all the available information. The last model that will be used in this work is MICE, a fully conditional specification method with classification and regression trees method. This implies that MICE minimizes the problem of finding a joint distribution for all missing data points simultaneously to finding separate conditional distribution for each incomplete data point, making it a very flexible approach [48]. After implementing different missing values techniques, the electricity price mean was checked.

Table 3. Electricity price means using LI, MI, and MICE

	With missing values	LI	MI	MICE
Electricity price mean	40.577	40.575	40.575	40.576

From the first sign, all missing data imputation techniques are quite similar, and final results do not vary significantly. Since the MICE method was closest to the real mean, this method was chosen, but after all model implication, it was detected that slightly better forecasting results were achieved using the MI method, so all further findings will be used with MI missing data imputation method. Missing value imputation in this context should not be significant because the total missing values are less than 1%, so the impact is negligible. Further, missing values rates between 1% and 5% correspond to manageable or flexible sample data. Missing data >5% of the total data require suitable solutions. While missing data of >15% significantly affects the prediction model [46].

After missing value imputation, the correlation heatmap was constructed. From the correlation matrix, we detected that load factors in Sweden and Finland have significantly higher than 0.84. Firstly, the actual load of Finland was eliminated since Finland does not have a direct electricity transmission grid with Lithuania. After removing the actual Finland load factor, a correlation heatmap was generated again, and a -0.72 correlation between Sweden load factor and temperature were detected, so the actual load in Sweden was eliminated. Final table can be seen in Figure 5.

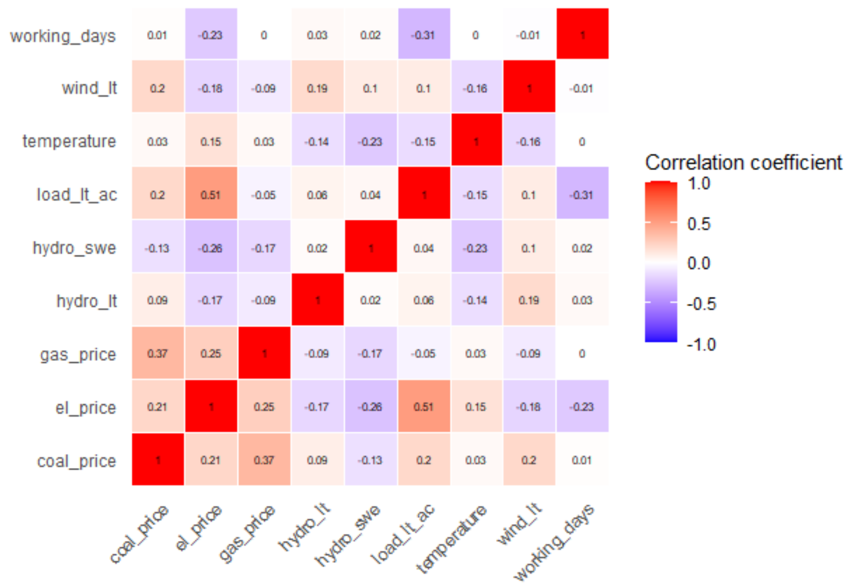


Figure 5. Correlation heatmap of all indicators

After the missing value imputation and correlation check, outlier detection was made. The thesis's novelty and the main contribution to Lithuania price forecasting will be to apply some price spikes detection methods before using the model since none of the Lithuania authors drew attention to this important issue and did not apply this to their models. Other analysed authors also paid little attention and did not apply any outlier detection methods [29] to such an important issue as sharp short-term and hardly predictable extreme values of the electricity price. These outliers usually occur due to accidents at power plants, congestions of the energy transmission grid, climatic anomalies, and growing renewable capacity in Lithuania and the world. To keep the model simple and understandable simple outlier detection methods was proceeded. Four outlier detection methods were compared: Threshold filter on prices (TFP), Standard deviation filter (SFP) on prices, Recursive filters on prices (RFP), and Percentage filter on prices (PFP). Using the TFP method, the threshold of 100 Eur/MWh was set, and to avoid negative data, the prices that were lower than zero were detected. TFP can be formulated as:

$$X_t^0 = \{X_t : |X_t| \geq 100\} \cup \{X_t : |X_t| < 0\} \tag{10}$$

SFP method was calculated as lowing:

$$X_t^0 = \{X_t : |X_t - \bar{X}| \geq 3 \cdot \delta \pm \mu\} \tag{11}$$

RFP method was calculated for every weekly and hourly seasonality. RFP method was calculated as lowing:

$$X_{t,i}^0 = \left\{X_t : \left|X_t - \bar{\bar{X}}_i\right| \geq 3 \cdot \delta_i \pm \mu_i\right\} \tag{12}$$

As the total number of outliers using all the methods are not higher than 1.0%, this threshold was used for PFP calculation:

$$X_t^0 = \{X_t: X_t \leq X_t^{1.0}\} \cup \{X_t: X_t \leq X_t^{99.0}\} \quad (13)$$

Before applying methods, basic outlier detection methods using Box-plot were done. As you can see in Figure 6, various periods were checked. The highest number of price spikes are seen during the working hours, the weekdays, and the summer. While no clear outlier seasonality can be seen in yearly Lithuania electricity data, it just can be remarked that in 2015, 2017, and 2019 electricity extreme points were not that high.

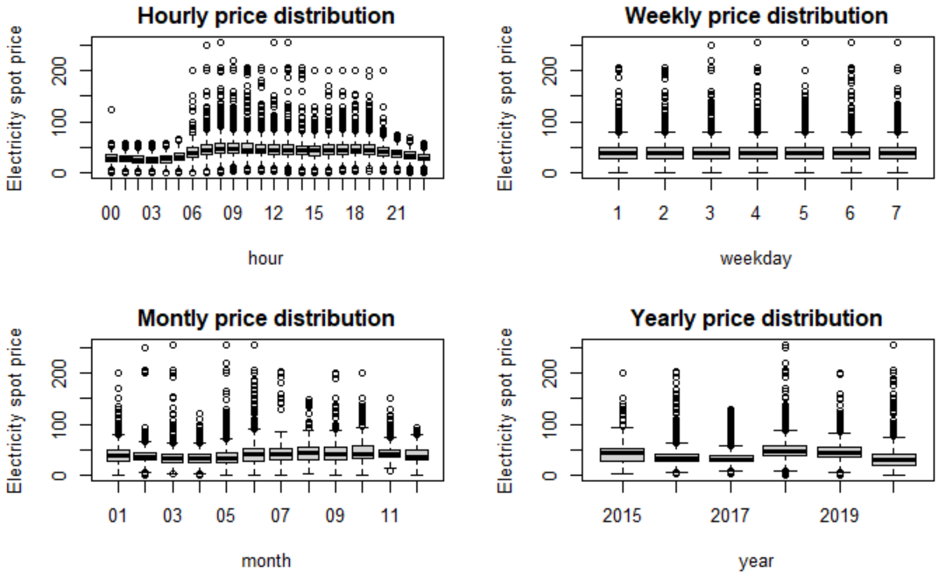


Figure 6. Electricity price outliers during various periods

After outlier detection, outliers should be replaced. For this step, two outlier imputation methods were considered. One was LI, and another moving average interpolation. Both methods were compared, and LI is relatively easy to use, but it demonstrates the limitation of inaccurate prediction when the missing data interval is long because it just draws a straight line between two data points, while moving average interpolation looked much better in this case.

When all these steps were applied, data was separated into the training data set and test data set. The training data set was hourly data from 1 January 2015 to 31 December 2019, containing 43,825 rows, while the test data set was hourly data from 1 January 2020 to 2 October 2020, containing a total of 6,623 rows. It seems simple, but again an important part of electricity price forecasting [6].

5.3.Decomposition

Time series data can be divided into 3 components: trend, seasonality, cyclical decomposition, and noise. The trend is a long-term increase or decrease in data, and it does not have to be linear. If the trend grows linearly, then the time series is linear, and if the trend grows exponentially, then the time series is exponential [31]. The seasonal component is the seasonal factor of time as a month in a year, a day in a week also some cyclical movements that occur outside of seasonality, and there is most likely random noise or unexpected variations that cannot be explained by the model [37]. Mathematically time series can be expressed as a sum or multiplication of these components. When applying additive decomposition, it is presumed that the data is the sum of its components. Multiplicative decomposition assumes that the data are a multiplication of its components, and it is mostly used when the variation in the data increases or decreases along with time. The data used for forecasting in this work do not show a linear relationship in their variation with time because it is quite similar in a long-term period, so an additive decomposition model will be more suitable and will be applied to the data. In Chapters 5.1. and 5.2. multiple seasonality effects were detected. The mathematical formula for additive decomposition can be expressed as the following, where S_t is the seasonal component, T_t is the trend, and cyclical component with noise are defined as reminder component R_t [31]:

$$y_t = S_t + T_t + R_t \tag{14}$$

While multiple seasonality mathematical formula for additive decomposition can be expressed as given below, where D_t is daily seasonality, W_t is weekly seasonality, Y_t is yearly seasonality, all these multiple seasonality's represents a seasonal component, further T_t is the trend and cyclical component with noise are defined as reminder component R_t [31]:

$$y_t = (D_t + W_t + Y_t) + T_t + R_t \tag{15}$$

As a base model in decomposition, only results using TFP will be visualised, other outlier detection methods will be included in the appendix. In can be noted that all result is quite similar and do not change significantly using various methods. Figure 7 presents the application of multiple seasonality decomposition using R software, forecasting package *msts()* function in *R*, where seasonal24 represents daily seasonality D_t, seasonal168 represents weekly seasonality W_t, seasonal8760 represents yearly seasonality Y_t.

In Figure 8, for clear visual seasonality detection, the first five weeks of the training period were used, where daily and weekly patterns are seen. As expected, daily and weekly seasonality are captured to a satisfactory degree, small decrease in daily seasonality due to lower electricity demand at night, decrease in weekly seasonality as seen previously in the weekly electricity price graphs, weekdays price is higher compared to weekends. To deal with seasonality, logarithm or differencing can be applied. In the next section, further steps are made.

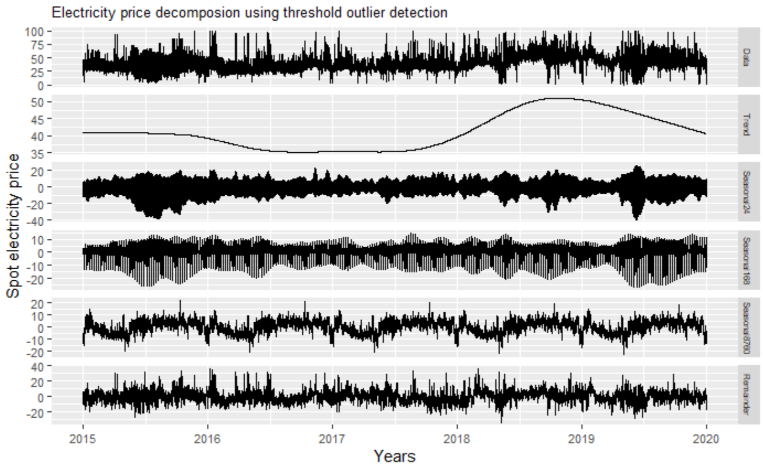


Figure 7. Multiple seasonality electricity price decomposition using TFP

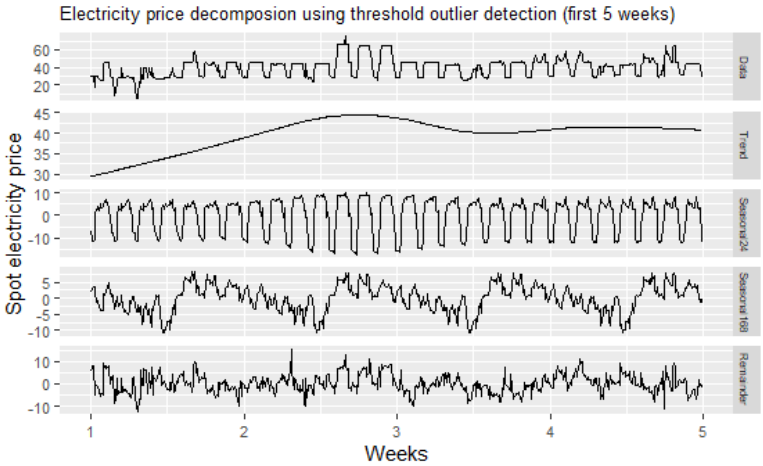


Figure 8. Multiple seasonality electricity price decomposition using TFP for the first five weeks of the training period

5.4. Stationarity testing

To further analysis and appropriate model implication, autocorrelation functions should be inspected to guarantee that time series are stationary, and there is no autocorrelation or structural breaks. For ARIMA(p,d,q) values identification KPSS test to identify differencing (d) number was applied. Autocorrelation Function (ACF) (q) and Partial ACF (p) were used to test if the forecasting model is a good fit or to see if there is more information left in the data that should be incorporated into the model. Autocorrelation is the correlation, the linear relationship, between the function and the delayed version of the function [32]. After ARIMA(p,d,q) values identification, the Augmented Dickey-Fuller Test to check stationarity was applied.

Stationarity tests were applied to both original values (after missing values imputation and outlier detection) and log values of electricity price, but original values with 164 lag were left for further analysis. Most of the studies found also did forecasting for original lagged data series, such as S. Voronin [6], Karabiber el at. [18], R. Beigaitė, T. Krilavičius [12] [13], S. Duffner [21] used logarithmic electricity price form, while S. B. Amor el at. [23] used log-returns.

KPSS test to identify a number of differencing (d) was applied. The first test was implemented on original values, giving p equal to 0.01, with all outlier detection methods applied. The null hypothesis for the KPSS test is that the data are stationary, while the alternative hypothesis assumes that data is not stationarity. Since we got a value lower than 0.05 null hypothesis is rejected, and differencing should be applied. After first differencing with lagged values of daily and weekly data were applied, giving p-value equal to 0.1, so the null hypothesis is not rejected, and first-order differencing (d = 1) should be applied.

Further, ACF (q) will be inspected. As a base electricity price with TFP will be shown, other graphs can be seen in the appendix. Looking at Figure 9, the maximum values are seen at both lags 24 and 48, indicating a daily seasonality. Also, a weekly graph was generated. It is visible that the maximum at both at multiples of 24 and growing pattern at multiples of 168 (7 × 24).

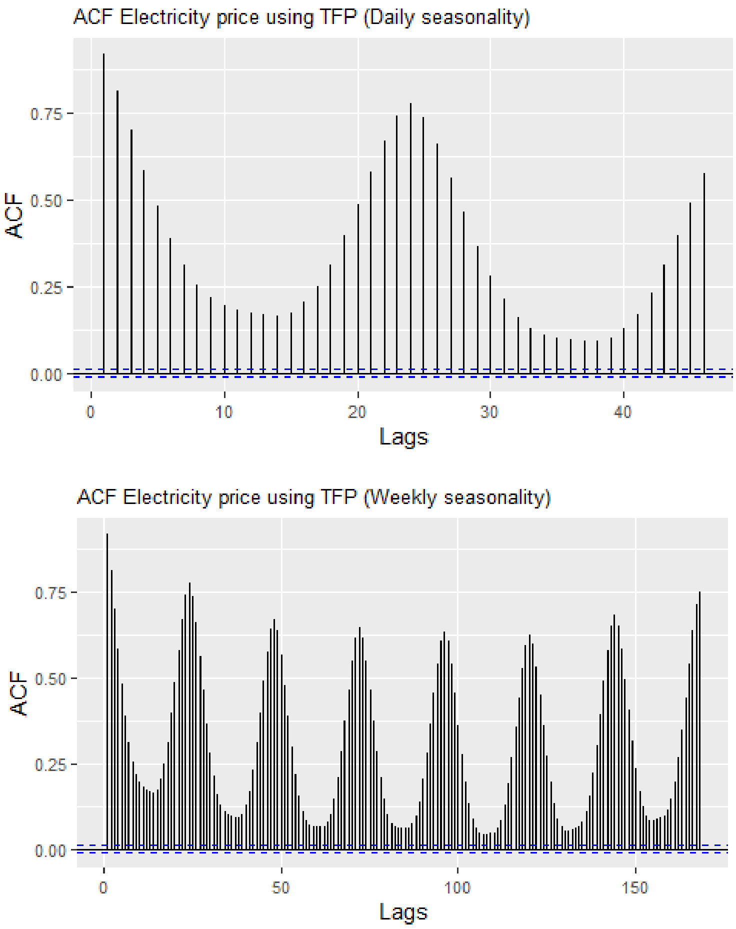


Figure 9. Hourly electricity price daily and weekly autocorrelation

For p indicator detection, Partial ACF graphs and analysis were done. The highest peaks were reached with lag = 0. Finally Augmented Dickey-Fuller (ADF) test was applied on first order differenced data. The null hypothesis of the ADF test is that unit root is present in the time series sample, which means because difference stationary against the alternative hypothesis is that unit root is not present in the time series, which means that data is trend stationary [49]. After applying the test, the null hypothesis was rejected because the p-value is less than 0.05, indicating strong evidence against the null hypothesis. Therefore alternative hypothesis was accepted, meaning that data is trend stationarity.

After ARIMA(p,d,q), values identification and stationarity approval model can be built.

5.5. Model building and forecasting

As mentioned, forecasting has various approaches, and a lot of testing and checking has to be made to find the best forecasting model. In this work, various ARIMA modelling approaches were tested. ARIMA with and without the remaining nine external variables will be analysed. Since most of the works about electricity price forecasting in the Lithuanian market were considering daily seasonality [14] [13], in my work a weekly seasonality with 168 lag, will be analysed. Models will be built by experimenting, and using results got in the previous chapters.

TFP, SFP, RFP, and PFP outlier detection methods will be compared, applying all selected models. The results will be explained by referring to the ARIMA and ARIMAX models with the lowest forecasting accuracy error. In this work, an error will be analysed in two different forms, RMSE and MAE, mentioned in Chapter 4.2.

All the variables in Table 4, after preliminary analysis will be used to build an electricity price forecasting model in Lithuania. After building the model, the backwords feature elimination to external variables will be applied (see Chapter 5.6) to see which variables are significant in this forecast.

Table 4. Explanation of used variable names

Variable	Explanation
working_days,	Calendar days, 0 – non-working days, 1 – working days and holidays
coal_price	Coal price, Eur/t
hydro_lt	Hydro power in Lithuania, MWh
hydro_swe	Hydro power in Sweden
load_lt_ac	Actual load in Lithuania, MWh
gas_price	Natural gas price, Eur/MWh
temperature	Temperature, C
wind_lt	Wind power in Lithuania, MWh

fourier	To incorporate the multiple seasonality, additional Fourier terms are added to the model (using *“auto.arima”*)

Four different types of AR-type models were constructed to define the best-fitting model and forecast, not including the experiments made while conducting these results. To several model selection criteria are calculated to statistically find the model with the best approximation and the smallest forecasting error. For forecasting and model adequacy RMSE with MAE errors will be considered. Their calculation is denoted in Chapter 4.2.

None of tried models had a lower RMSE error than 5, so additional ARIMA models was introduced. Since electricity price data has double seasonality and AR-type models cannot catch these patters, a fit for the seasonality with Fourier series will be used [52]. This data then will be analysed together with and without remaining variables, hoping that this method will improve electricity price forecast in Lithuania. To catch double seasonality *“fourier()”* function from *forecast* package will be used, *“auto.arima()”* function from the same package will be used to make the actual fit along with using additional command *seasonal = FALSE* [53]. This function is an automated algorithm used to choose the best coefficients in an ARIMA model and it is a variation of the Hyndman and Khandakar’s [54] algorithm. It uses unit root tests and minimization of AIC to get to the result. After the *auto.arima* implication the best model with Fourier series as an external regressor, to detect double seasonality, were found: ARIMA(3,1,2). After additional model implication in total 5 models were used to forecast electricity price in Lithuania. All models RMSE and MSE accuracy measures were compared in Table 5.

Table 5. A summary of models forecasting accuracy measures

Outlier detection method	**Models**									
	ARIMA (1,0,0)		ARIMA(0,1,0)		ARIMA(1,1,1)		ARIMA (3,1,1)		ARIMA (3,1,2)	
	Uni	Multi	Uni	Multi	Uni	Multi	Uni	Multi	Uni*	Multi*
Accuracy measure - RMSE										
TFP	5.81	5.28	5.94	5.50	5.89	5.23	5.89	5.22	5.11	5.06
SFP	5.67	5.14	5.79	5.35	5.74	5.09	5.47	5.08	4.97	4.93
RFP	5.81	5.28	5.93	5.49	5.89	5.23	5.88	5.22	5.10	5.06
PFP	5.61	5.09	5.73	5.31	5.70	5.03	5.70	**5.03**	4.86	**4.82**
Accuracy measure - MAE										
TFP	3.36	3.27	3.13	3.11	3.07	3.26	3.07	3.24	3.12	3.10
SFP	3.32	3.23	3.09	3.07	3.03	3.21	3.23	3.20	3.09	3.06
RFP	3.36	3.27	3.13	3.11	3.07	3.25	3.07	3.24	3.12	3.10
PFP	3.30	3.22	3.08	3.05	3.03	3.20	3.03	3.19	3.07	3.04

*With Fourier series an external regressor.

In Table 5, the forecasting errors of both univariate and multivariate models were shown. MAE, which shows the average magnitude of the errors in a set of forecasts, without considering their direction, was lower in univariate series, but since the difference, it is not that significant, the results will be done accordingly to RMSE error. The smallest RMSE were detected in both ARIMA(3,1,1) and additional ARIMA (3,1,2) multivariate models. While All models (except the one using the Fourier series) share relatively similar forecasting errors varying from 5.03 to 5.89 RMSE. We can see that RMSE error in all cases are lower when additional variables are used. To improve these results, additional ARIMA model with Fourier series, which can detect double seasonality, was introduced. ARIMA (3,1,2) with Fourier series as a regressor significantly improved forecast, with RSME reaching 4.82 and MEA 3.04. Meaning that double seasonality while forecasting electricity price is important and improves RMSE forecasting measure by 4.2%. The chosen model graphical representation can be visible in Figure 10.

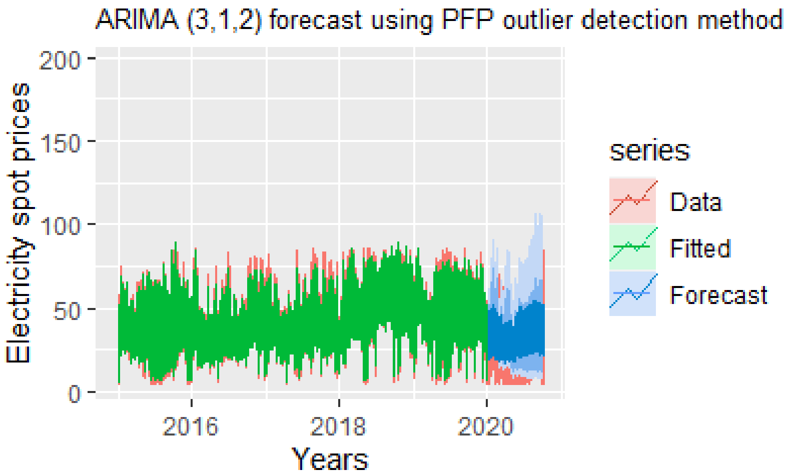


Figure 10. ARIMA (3,1,2) forecast using PFP outlier detection method

5.6. Backwords feature elimination

In Chapter 5.1 more detailed external regressor analysis were made. In Chapter 5.2 for variables selection, linear correlation technique was used. Linear correlation analysis is widely used for feature selection; however, it is a linear technique and often cannot consider the original price's nonlinearities [6]. For significant feature selection, backward feature elimination method was used. This method is used to minimize the number of external variables, optimize the computation time, and minimize the error, for electricity price buyers and sellers, leaving just the most important variables. There are n=9 variables (excluding in *"auto.arima"* used Fourier series and dummies). In the previous chapter, it was concluded that ARIMA(3,1,2) using the PFP outlier detection method model has the lowest accuracy, but despite greater accuracy, due to time-consuming model building, the backward feature elimination was not applied using this model. So, ARIMA(3,1,1) with PFP outlier detection method model, as the second-best model, will be run as a benchmark model for backward feature elimination method. The ARIMA(3,1,1) with PFP outlier detection method forecast will be run with n − 1 = 8 variables and each time a different variable will be eliminated from the forecast and errors will be compared. This will allow us to see if any variables are unnecessary, disruptive, or tremendously important for superior forecasts [18].

The first backward feature elimination attempt (see Table 6) showed that few variables should be eliminated. Surprisingly, the biggest difference in forecasting accuracy measures was when hydro power volumes in Lithuania were excluded, after hydro power elimination, RMSE improved by 2.40% and MAE by 2.98%. There is a slight improvement in both RMSE and MAE maximum error, respectively by 3.48% and 4.02%. Improvement in standard deviation is a bit higher, reaching 6.14% in RMSE and 6.24% in MAE. The variable that has the greatest positive effect on forecasting accuracy is the actual load in Lithuania.

Table 6. ARIMA with nine regressors

Excluded Variable	**RMSE**				**MAE**			
	Mean	Min	Max	Std Dev	Mean	Min	Max	Std Dev
Benchmark ARIMA	**8.254**	**5.028**	**11.481**	**4.56**	**6.137**	**3.190**	**9.083**	**4.17**
Calendar days	8.264	5.030	11.498	4.57	6.139	3.191	9.087	4.17
Coal price	8.277	5.028	11.527	4.60	6.157	3.190	9.125	4.20
Hydro power in Lithuania,	8.056	5.029	11.082	4.28	5.954	3.190	8.718	3.91
Hydro power in Sweden	8.205	5.028	11.383	4.49	6.087	3.191	8.983	4.10
Actual load in Lithuania,	10.494	5.393	15.595	7.21	7.995	3.261	12.728	6.69
Natural gas price	8.398	5.028	11.768	4.77	6.265	3.190	9.340	4.35
Temperature	8.175	5.042	11.309	4.43	6.084	3.192	8.975	4.09
Wind power in Lithuania	8.213	5.040	11.385	4.49	6.065	3.189	8.942	4.07

After hydro power in Lithuania variable elimination (see Appendix 5), errors were rechecked, and no significant improvement with any other variable elimination was recorded. It can be assumed that the remaining variables contribute to the model. Even after hydro power in Lithuania variable removal forecasting accuracy errors remained pretty much the same, meaning that also other factors play a role that is not captured in this dataset, like congestion, hydro power reservoir levels in Lithuania and its neighbouring countries, different bidding strategies by the market participants or countries policies and actions related to energy sectors and that may have an impact to electricity price.

5.7. Results and Evaluation

Considering that electricity price display features such as high volatility, spikes, double seasonality, and numerous methods that can be used in electricity price forecasting, make it difficult to achieve high accuracy in forecasting.

In this paper, various statistical ARIMA models were analysed. The efficiency and usefulness of statistical analysis in financial markets are often questioned. The methods stand a better chance in energy markets because of the seasonality prevailing in electricity price processes during normal, non-spiky periods. Even statistical models do not perform well with price spikes, price spikes should be captured using an adequate model. However, the literature does not agree, or outlier detection should, or should not be included in the models [29]. In this paperwork to deal with high electricity price spikes, four outlier detection methods were used: PFP, RFP, SFP, and TFP. To catch the double seasonality model with Fourier series was introduced. After numerous attempts, the best ARIMA model was ARIMA(3,1,2) using a percentage filter on price outlier detection with Fourier series included as an external variable, together with other variables. This model RSME reached 4.82 and MEA 3.04. This means that double seasonality while forecasting electricity price is important and improves forecasting accuracy measure RMSE by 4.2%, compared to the second-best model with 168 lag, ARIMA(3,1,1) with external variables, using PFP outlier detection method. Seven external regressors were included in the model: Calendar days, Coal price, Hydro power in Sweden, Actual load in Lithuania, Natural gas price, Temperature, Wind power in Lithuania.

Despite various external variables implemented and various models tested, forecasting accuracy remained quite high, and it did not outperform R. Beigaitė and T. Krilavičius [13] [12] researches. For hourly electricity price forecasting in the Lithuania, they used average, seasonal naïve, and exponential smoothing methods, where minimum forecasting errors was reached with exponential smoothing equal to 1.76% MAPE, 0.66 MAE and 0.83 of RMSE and for the next research, they used Elman and Jordan neural network methods, where minimum forecasting error was achieved using Elman neural network with MAPE error equal to 3.55%, 1.12 MAE and 1.34 RMSE. But comparing to Nord Pool market researches, 4.82 RMSE is quite good accuracy. O.A. Karabiber et al. [18] lowest forecasting error was conducted using ARIMA with 7.95 of RMSE. S. Duffner el. at. [21] the best working method was ARIMAX with 6.6. While the lowest accuracy was reached by B. Amor et al. [23] proposed a new hybrid model k-factor GARMA-LLWNN model with MAPE < 1%. To sum up, 4.82 RMSE accuracy is quite good, and the model is adequate for further usage and it outperforms some of the works done with the same models, due to seasonality and outlier detection methods. Electricity price can be forecasted for the next 168h and it can be used for electricity sellers and buyers in the Nord Pool power market. Nevertheless, despite these findings, there is still plenty of room for improvement in Lithuania's electricity price forecasting.

As I already pointed out, numerous variations for the modelling have been tried but did not yield significant improvements and error do not vary a lot, so most possibly also other factors play a role which is not captured in this dataset like congestion, hydro power reservoir levels in Lithuania and its neighbouring countries, different bidding strategies by the market participants or countries policies and actions related to energy sectors and that may have an impact to electricity price. For example, most resent Astravets Nuclear Power Plant commercial start and strict countries policies about them. From November 2020, the commercial flow of electricity from Belarus was cut off when the Astravets Nuclear Power Plant became operational. Market participants are currently still trying to assess the ultimate effect of that effect on prices in the region. We will probably see the ultimate Astravo effect in more time when conditions are available for relatively more historical data than in 2021 onwards.

6. CONCLUSION

Eight objectives were determined and examined in this research, and they can be listed as below:

1. To analyse literature that focuses on electricity price forecasting in Lithuania;
2. To analyse literature that focuses on electricity price forecasting at Nord Pool Power market;
3. To collect data and to do a preliminary analysis;
4. To detect a set of candidate explanatory variables that may influence electricity price in Lithuania;
5. Select and build understandable and easily usable short-term electricity price forecasting model for electricity price buyers and sellers;
6. Compare selected models with and without external variables using accuracy measures;
7. Compare selected models with different outlier detection method using accuracy measures;
8. Evaluate the most precise model adequacy.

All objectives were successfully implemented. The literature on electricity price forecasting in Lithuania was reviewed, concluding that only a few pieces of research were done with a huge space for improvement. In this paper, various improvements for electricity price in Lithuania forecasting were selected, such as external variables imputation, outlier detection methods implementation. The second objective was to analyse literature that focuses on the Nord Pool power market electricity price forecasting, concluding that various models and techniques can be used to deal with various problems, which impact electricity price forecasting.

Another quite time-consuming goal was data collection. Data were collected, aggregated, analysed, and prepared for the model. Analysis of Lithuania's price area showed that there are double seasonality patterns. Furthermore, prices tend to be lower in winter-spring months, at night and on weekends. Based on Lithuania energy market characteristics reviewed in Chapter 2.1 and existing literature Chapter 3, important regressors for the electricity price have been worked out, leaving us with 11 variables: Calendar days, Coal price, Hydro power in Lithuania, Hydro power in Sweden, Actual load in Lithuania, Actual load in Sweden, Actual load in Finland, Natural gas price, Temperature, Unavailability in transmission Grid, Wind power in Lithuania. After linear correlation and backward feature elimination, only seven variables are left, assuming that Unavailability in the transmission grid, Actual load in Sweden, Actual load in Finland, and Hydro power in Lithuania are not significant for the model.

For the fifth objective implementation, the statistical ARIMA model was chosen. This model was chosen due to a few reasons. First, there was no other work for short-term electricity price forecasting in Lithuania using ARIMA models, and in the author's opinion, this type of model is fundamental to analyse, before implementing another type, hybrid or more advanced, models. Second, this model was chosen due to reusability for electricity price buyers and sellers in the Elspot Nord Pool power market because the model is easily understandable, and parameters can be selected or dropped according to its understanding and needs. And finally, because in various cases, statistical models outperform more advanced models [13] [50] and give better results, so the author wanted to focus and analyse statistical model performance.

In Chapter 5.5 and 5.6 selected models, univariate and multivariate, with different outlier detection methods were compared. The prices are forecasted for the Lithuania area in the Nord Pool power market. The best fitted ARIMA model was ARIMA(3,1,2) using percentage filter on price outlier detection and Fourier series included as an external variable, together with other seven external variables. This model RSME reached 4.82 and MEA 3.04. In comparison to similar researches in the Nord Pool market Karabiber et al. [18] and S. Duffner el. at. [21], my model performs a bit better, probably as a result of price spikes detection and double seasonality detection with Fourier series. To sum up, 4.82 RMSE accuracy is quite good and the model is adequate for further usage. Electricity price can be forecasted for the next 168h and it can be used for electricity sellers and buyers in the Nord Pool Power market.

Nevertheless, despite these findings, there is still plenty of room for improvement in Lithuania's electricity price forecasting. In further researches, more significant variables should be tested, including congestion, hydro power reservoir levels in Lithuania and its neighbouring countries, also more neighbouring countries data should be included because the Lithuania market is small and dependent on other, bigger markets. More advanced or hybrid models, which can cope with double seasonality, should be used, assuming it could provide higher forecasting accuracy. In further works, political and environmental aspects should be considered, such as regulations toward Belarus, global warming, or decarbonization.

Considering the increase of renewables, decarbonatization, upcoming deregulation of the Lithuania market, and the merging of the different European markets (referring to recent Netherlands, Germany, France entrance to Nord Pool market), the forecasting of electricity price will remain an important topic in future and will experience much development.

BIBLIOGRAPHY

[1] Republic of Lithuania Law on Electricity, 20 07 2000. [Online]. Available: https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/TAIS.106350/LwPhzaCnNY. [Accessed 15 10 2020].

[2] Lithuania electricity transmission system operator litgrid, [Online]. Available: https://www.litgrid.eu/. [Accessed 15 11 2020].

[3] O. Knapik, Modeling and forecasting electricity price jumps in the Nord, Aarhus V, Denmark: Center for Research in Econometric Analysis of Time Series, 2017.

[4] Ministry of Energy of the Republic of Lithuania, 08 05 2020. [Online]. Available: https://enmin.lrv.lt/en/news/the-seimas-approves-liberalisation-of-the-electricity-market-presenting-new-opportunities-for-electricity-consumers. [Accessed 30 11 2020].

[5] National energy independence strategy 2018 by Ministery of Energy of the Republic of Lithuania, 2018. [Online]. Available: https://enmin.lrv.lt/uploads/enmin/documents/files/Nacionaline%20energetines%20nepriklausomybes%20strategija_2018_LT.pdf. [Accessed 15 10 2020].

[6] S. Voronin, “Price spikes forecasting in a competitive day-ahead energy market,” Acta Universitatis Lappeenrantaensis 530, Lappeenranta, Finland, 2013.

[7] “World Wind Energy Association,” [Online]. Available: https://wwindea.org/. [Accessed 12 12 2020].

[8] R. Bourbonnais and S. Méritet, “Electricity Spot Price Modelling: Univariate Time Series Approach,” in *The Econometrics of Energy Systems*, Chippenham and Eastbourne, Antony Rowe Ltd, 2007, pp. 51-75.

[9] Pasirink tiekėją, [Online]. Available: https://pasirinkitetiekeja.lt/rinkos-tiekejai. [Accessed 13 12 2020].

[10] Nord Pool, [Online]. Available: https://www.nordpoolgroup.com/#. [Accessed 13 12 2020].

[11] J. S. Roungkvis, P. Enevoldsen and G. Xydis, “High-Resolution Electricity Spot Price Forecast for the Danish Power Market,” *Sustainability*, vol. doi:10.3390/su12104267, no. 12, 4267, p. 19, 2020.

[12] R. Beigaitė and T. Krilavičius, "Electricity price forecasting for Nord Pool data," Baltic Institute of Advanced Technology, Vytautas Magnus University, Lithuania, 2017.

[13] R. Baigaitė and T. Krilavičius, “Electricity Price Forecasting for Nord Pool Data Using Recurrent Neural Networks,” Vytautas Magnus University, Baltic Institute of Advanced Technology, Lithuania, 2018.

[14] A. N. Tat, “Electricity price forecasting using Monte Carlo simulation: the case of Lithuania,” *EKONOMIKA*, vol. doi: https://doi.org/10.15388/Ekon.2018.1.11780, no. Vol. 97(1), pp. 76-86, 2018.

[15] M. Česnavičius, “Lithuanian electricity market price forecasting model based on univariate time series analysis,” *ENERGETIKA*, vol. T. 66. , no. Nr. 1, p. 39–46, 2020.

[16] A. Bogdanov , “Elektros kainų modeliavimas tiesioginėje rinkoje,” Master's thesis, Vilnius, Lithuania, 2007 .

[17] V. Bobinaite, I. Konstantinavičiūtė and V. Lekavičius, “Theoretical model for electricity market price forecasting,” *ECONOMICS AND MANAGEMENT*, vol. 17(3), no. doi: 10.5755/j01.em.17.3.2119, pp. 944-951, 2012.

[18] O. A. Karabiber and G. Xydis, “Electricity Price Forecasting in the Danish Day-Ahead,” *Energies*, vol. doi:10.3390/en12050928, pp. 12-928, 2019.

[19] T. Kristiansen, “Forecasting Nord Pool day-ahead prices with an autoregressive model,” *Energy Policy,* vol. 49, no. https://doi.org/10.1016/j.enpol.2012.06.028, pp. 328-332, 2012.

[20] R. Weron and A. Misiorek, “Forecasting sort electricity prices with time series models,” in *The European Electricity Market EEM-05,* Lodz, Poland, 2005.

[21] S. Duffner , “Forecasting German Day-Ahead Electricity Prices using multivariate Time Series Models,” Master Thesis, Norwegian School of Economics , Norway, 2012 .

[22] J. Peljo, “Futures Pricing in the Nordic Electricity Market,” Master Thesis, Aalto University School of Business, 2013.

[23] B. Amor, H. Boubaker and L. Belkacem, "Electricity price, k- factor GARMA, LLWNN, Hybrid, Forecasting," *Journal of Forecasting,* vol. Special Issue Article, no. 10.1002/for.2544, pp. 1-38, 2018.

[24] K. Rahimimoghadam, H. Hosseinian and M. Saravi, “Day-Ahead Price Forecasting of Electricity Market Using Neural Networks and Wavelet Transform,” *Electrical and Electronic Engineering,* vol. 8(2), no. doi:10.5923/j.eee.20180802.02, pp. 37-52, 2018.

[25] E. Raviv, K. Bouwman and D. v. Dijk1, “Forecasting day-ahead electricity prices: utilizing hourly prices,” Econometric Institute, Erasmus University Rotterdam, Rotterdam, 2013.

[26] H. Torró, “Forecasting Weekly Electricity Prices at Nord Pool,” University of Valencia , Valencia, Spain, 2007.

[27] P. Spodniak, K. Ollikkac and S. Honkapuro, “The relevance of wholesale electricity market places: the Nordic case,” Working Paper No.631, Dublin, Ireland, 2019.

[28] P. Pinson, H. A. Nielsen and H. Madsen, “Predictive Densities for Day-Ahead Electricity Prices Using Time-Adaptive Quantile Regression,” *energies,* Vols. 7(9):5523-5547, no. DOI: 10.3390/en7095523, pp. 5523-5547, 2014.

[29] R. Weron, “Electricity price forecasting: A review of the state-of-the-art with a look into the future,” *International Journal of Forecasting,* vol. 30(4), no. doi: 10.1016/j.ijforecast.2014.08.008, pp. 1-46, 2014.

[30] P. F. Pai and C. S. Lin, “A hybrid ARIMA and support vector machines model in stock price forecasting,” *Omega,* vol. 33, no. 6, p. 497–505, 2005.

[31] R. J. Hyndman and G. Athanasopoulos, Forecasting : principles and practice, Melbourne, Australia: 2nd edition, OTexts, 2014.

[32] J. Fattah, L. Ezzine, Z. Aman, H. E. Moussami and A. Lachhab, “Forecasting of demand using ARIMA model,” *International Journal of Engineering,* vol. 10, no. doi: 10.1177/1847979018808673, pp. 1-9, 2018.

[33] A. Buteikis, “Lecture notes: Stationary time series,” 02 2020. [Online]. Available: http://web.vu.lt/mif/a.buteikis/wp-content/uploads/2020/02/Lecture_01.pdf. [Accessed 08 12 2020].

[34] S. E. Sim, K. G. Tay, A. Huong and W. K. Tiong, “Forecasting Electricity Consumption Using SARIMA Method in IBM SPSS Software,” *Universal Journal of Electrical and Electronic Engineering,* vol. 6, no. 5B, pp. 103-114, 2019.

[35] D. Kwiatkowski, P. C. Phillips, P. Schmidt and Y. Shin, “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,* vol. 54, no. 1-3, pp. 159-178, 1992.

[36] R. Alwee, S. M. H. Shamsuddin and R. Sallehuddin, “Hybrid Support Vector Regression and Autoregressive Integrated Moving Average Models Improved by Particle Swarm Optimization for Property Crime Rates Forecasting with Economic Indicators,” *The Scientific World Journal* , vol. (4):951475, no. doi: 10.1155/2013/951475, 2013.

[37] K. Himakireeti and T. Vishnu, “Air Passengers Occupancy Prediction Using Arima Model,” *International Journal of Applied Engineering Research,* vol. 14, no. 3, pp. 646-650, 2019.

[38] N. Singh and . S. Mohanty, “A review of price forecasting problem and techniques in deregulated electricity markets,” *Journal of Power and Energy Engineering,* vol. 3, no. 9, p. 1, 2015.

[39] P. Burman and R. H. Shumway, “Generalized Exponential Predictors for Time Series,” *Journal of the American Statistical Association,,* vol. 101, no. 476, p. 1598–1606, 2006.

[40] L. Zhou, P. Zhao, D. Wu, C. Cheng and H. Huang, “Time series model for forecasting the number of new admission inpatients,” *BMC Medical Informatics and Decision Making,* vol. 18, no. 39, p. 1, 2018.

[41] ENTSO-E, “Transparency Platform,” [Online]. Available: https://transparency.entsoe.eu/dashboard/show. [Accessed 03 10 2020].

[42] “Intercontinental Exchange,” [Online]. Available: https://www.theice.com/. [Accessed 03 10 2020].

[43] “Lithuanian Hydrometeorological service,” [Online]. Available: http://www.meteo.lt/en/. [Accessed 14 11 2020].

[44] R. Tawn, J. Browell and I. Dinwoodie, “Missing Data in Wind Farm Time Series: Properties and Effect on Forecasts,” in *21st Power Systems Computation Conference*, Porto, Portugal, 2020.

[45] K. Zor, Ö. Çelik, O. Timur and B. Yıldırım, “Simple Approaches to Missing Data for Energy Forecasting Applications,” in *Conference: 16th International Conference on Clean Energy*, Gazimağusa, Turkish Republic of Northern Cyprus, 2018.

[46] T. Kim, W. Ko and J. Kim, “Analysis and Impact Evaluation of Missing Data Imputation in Day-ahead PV Generation Forecasting,” *Applied Sciences,* vol. 9(1), no. doi:10.3390/app9010204, p. 204, 2019.

[47] S. Jung, J. Moon, S. Park, S. Rho, S. W. Baik and E. Hwang, “Bagging Ensemble of Multilayer Perceptrons for Missing Electricity Consumption Data Imputation,” *Sensors,* vol. 20, no. doi:10.3390/s20061772, p. 1772, 2020.

[48] S. v. Buuren, “Multiple imputation of discrete and continuous data by fully conditional specification,” *Stat Methods Med Res,* vol. 16(3), no. doi: 10.1177/0962280206074463, pp. 219-42, 2007.

[49] D. A. Dickey and W. A. Fuller, “Distribution of the Estimators for Autoregressive Time Series With a Unit Root,” *Journal of the American Statistical Association,* vol. 74, no. 366, pp. 427-431, 1979.

[50] J. Korpihalkola, “Forecasting Nord Pool’s Day-ahead hourly spot prices in Finland,” Bachelor’s thesis, Finland, 2019.

[51] T. Jakaša, I. Andročec and I. Sprčić, “Electricity price forecasting – ARIMA,” in *2011 8th International Conference on the European Energy Market (EEM),* Zagreb, Croatia, 2011.

[52] A. Skrivarhaug, “Energy Markets & Price Analysis in Practice,” Presentation during Lecture of the Course ENE421 at 07.10.2010 at NHH, Bergen, Norway, 2010.

[53] A. H. Studenmund, Using econometrics: A practical guide (4th ed.), Pearson, USA, 2016.

[54] D. O. Afanasyeva and E. A. Fedorovab, “On the impact of outlier filtering on the electricity price forecasting,” in *International Conference on Applied Research in Economics 2018*, Perm, Russia, 2018.

[55] T. S. Christensen and F. E. Benth, “Modeling the exact convergence of electricity prices in interconnected markets,” *SSRN Electronic Journal,* no. doi: 10.2139/ssrn.3312602, p. 19, 2019.

[56] R. J. Hyndman, “Forecasting with daily data,” 17 09 2013. [Online]. Available: https://robjhyndman.com/hyndsight/dailydata/. [Accessed 29 12 2020].

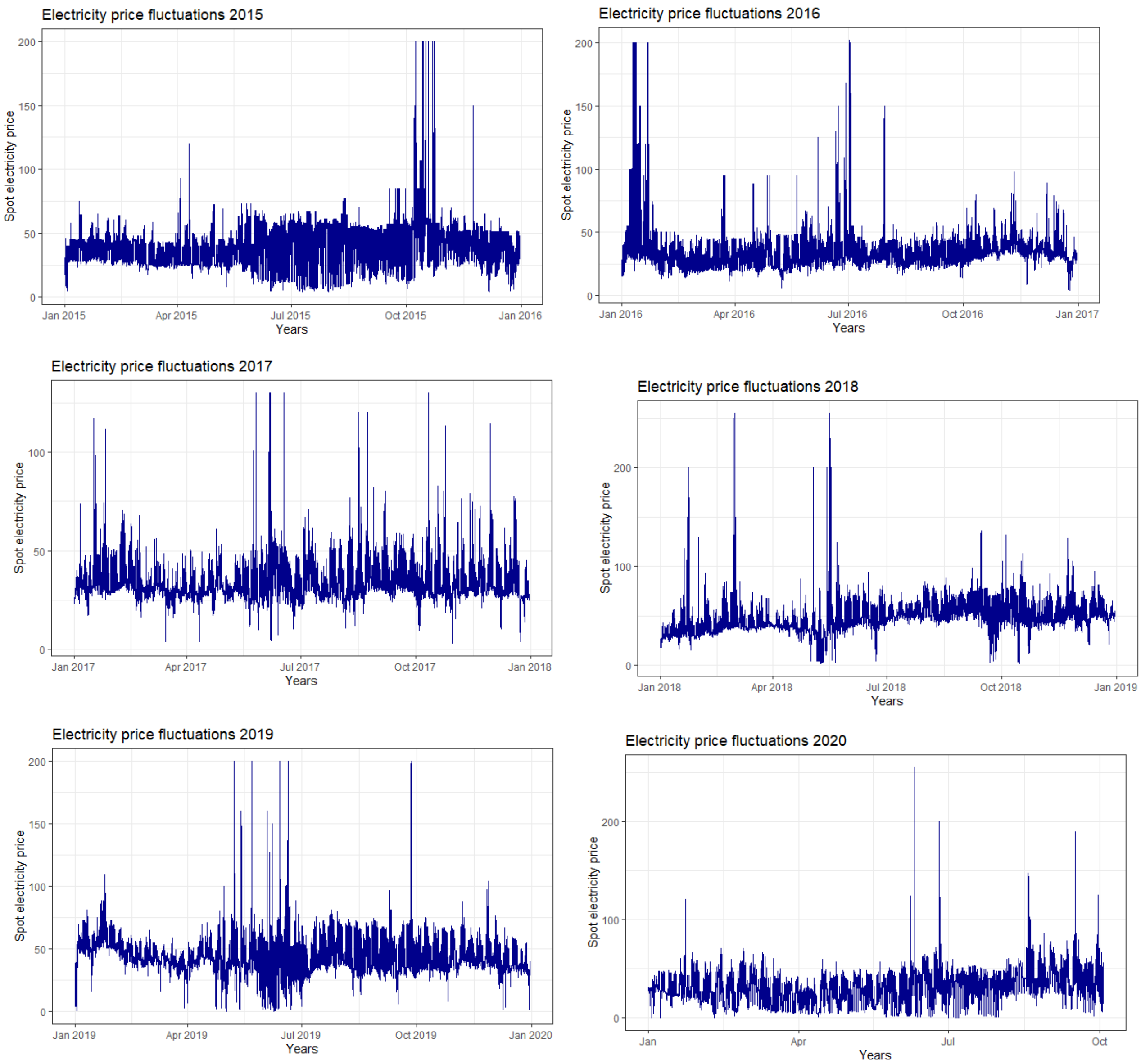
[57] R. J. Hyndman, “Prediction intervals for NNETAR models,” 25 05 2017. [Online]. Available: https://robjhyndman.com/hyndsight/nnetar-prediction-intervals/. [Accessed 28 12 2020].

[58] R. J. Hyndman and Y. Khandakar, “Automatic Time Series Forecasting: The forecast Package for R,” *Journal of Statistical Software,* vol. 27, no. doi: 10.18637/jss.v027.i03, p. 1–22, 2008.

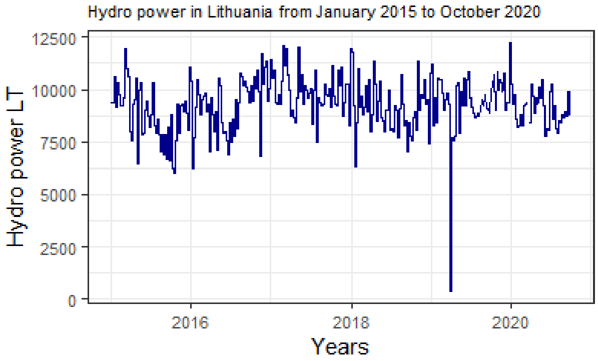
[59] J. Nowotarski and R. Weron, “Computing electricity spot price prediction intervals using quantile regression and forecast averaging,” *Computational Statistics,* vol. 30, no. DOI 10.1007/s00180-014-0523-0, p. 791–803, 2015.

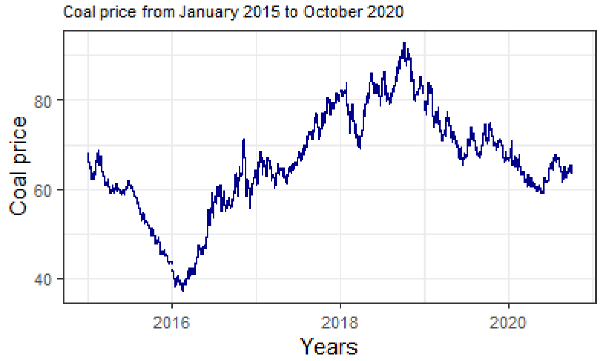
A APPENDIX

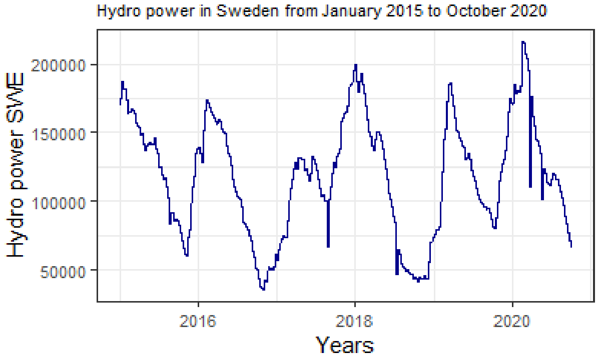
Appendix 1. Electricity price fluctuations in 2015-2020.

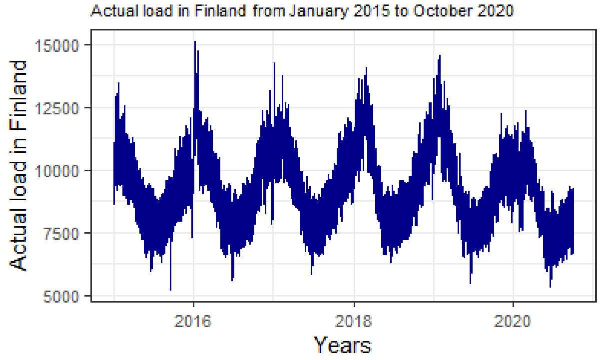


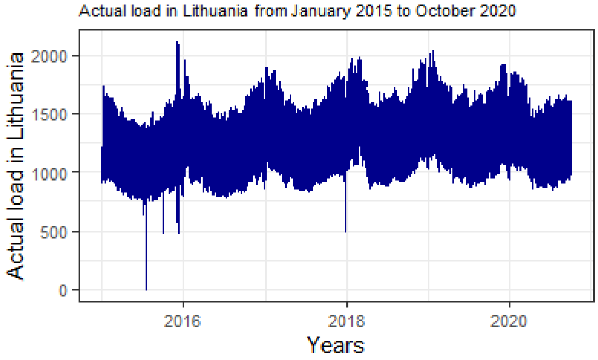
Appendix 2. External variables graphical representation.

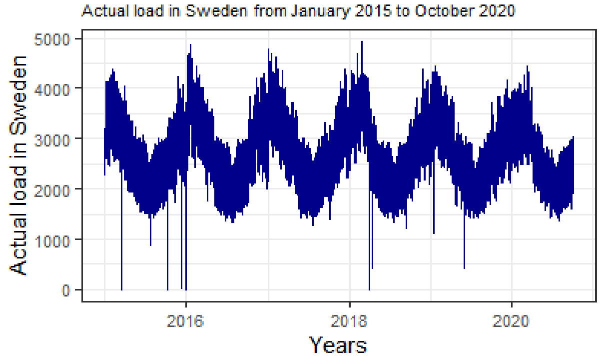

Hydro power in Lithuania from January 2015 to October 2020
Hydro power LT
12500
10000
7500
5000
2500
0
2016
2018
2020
Years


Coal price from January 2015 to October 2020
Coal price
80
60
40
2016
2018
2020
Years

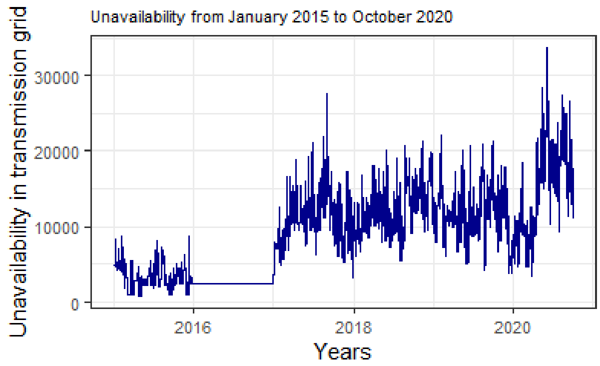

Hydro power in Sweden from January 2015 to October 2020
Hydro power SWE
200000
150000
100000
50000
2016
2018
2020
Years

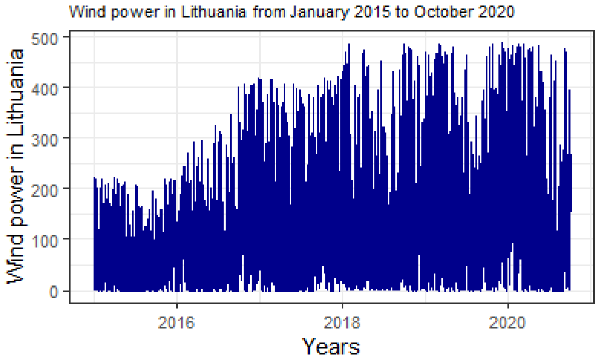

Actual load in Finland from January 2015 to October 2020
Actual load in Finland
15000
12500
10000
7500
5000
2016
2018
2020
Years

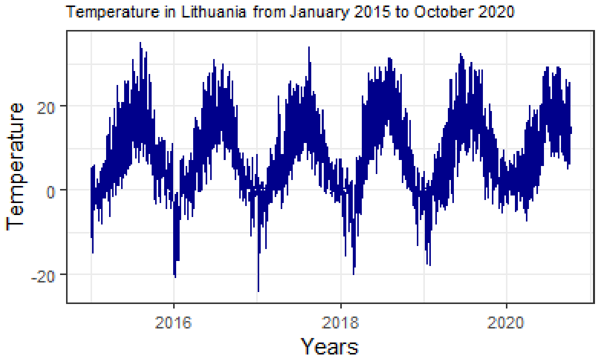

Actual load in Lithuania from January 2015 to October 2020
Actual load in Lithuania
2000
1500
1000
500
0
2016
2018
2020
Years


Actual load in Sweden from January 2015 to October 2020
Actual load in Sweden
5000
4000
3000
2000
1000
0
2016
2018
2020
Years

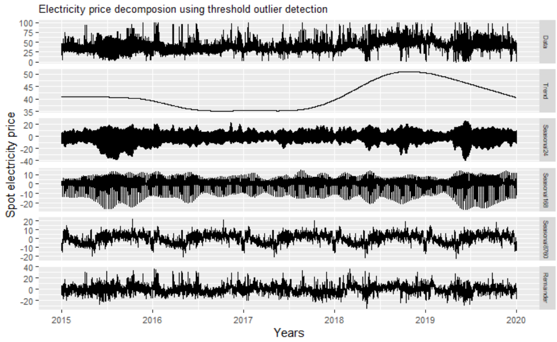

Natural gas price from January 2015 to October 2020
Natural gas price
21
18
15
2016
2018
2020
Years

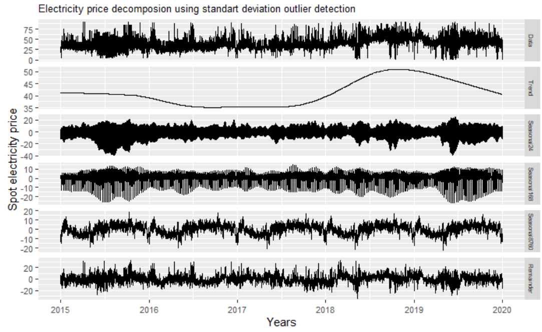

Unavailability from January 2015 to October 2020
Unavailability in transmission grid
30000
20000
10000
0
2016
2018
2020
Years

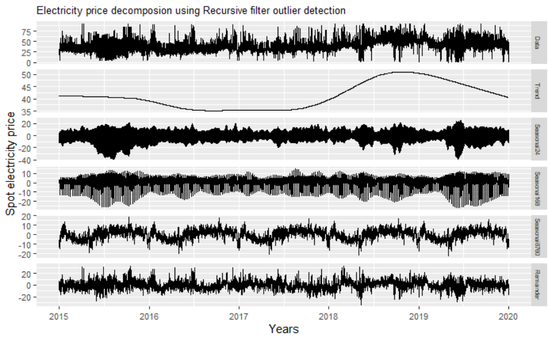


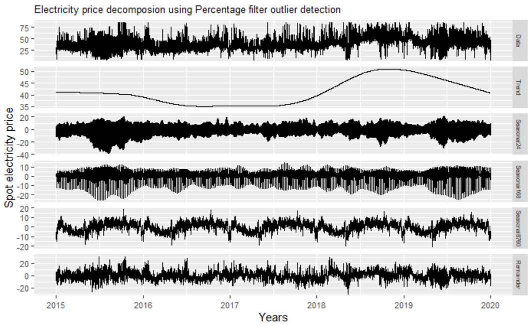


Appendix 3. Decomposition

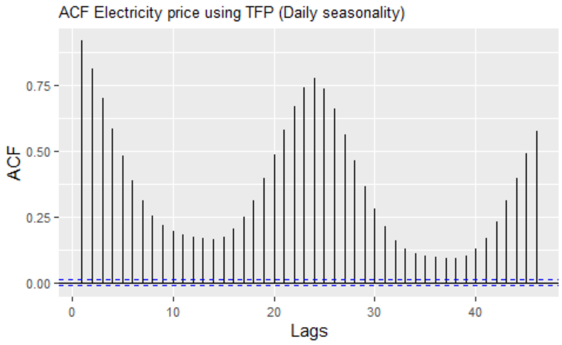


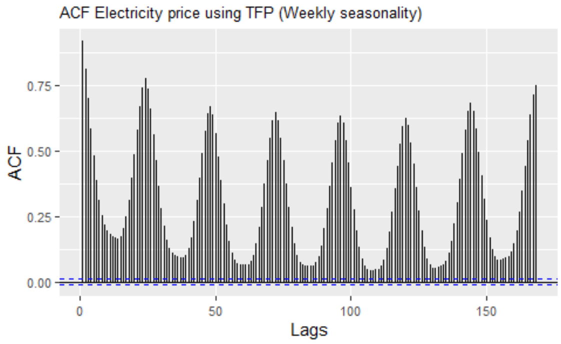


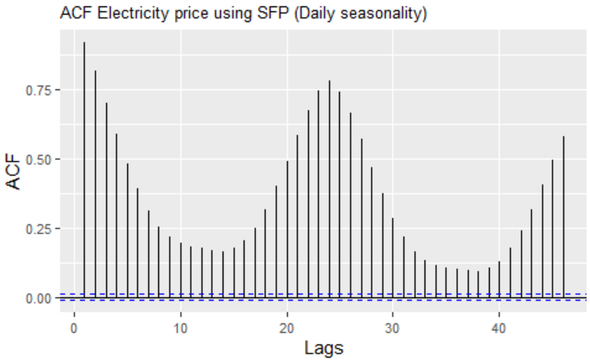


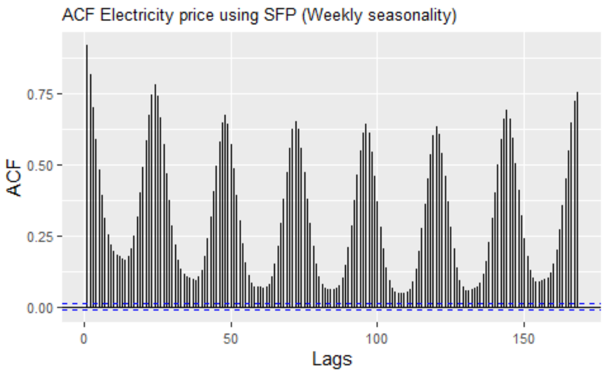


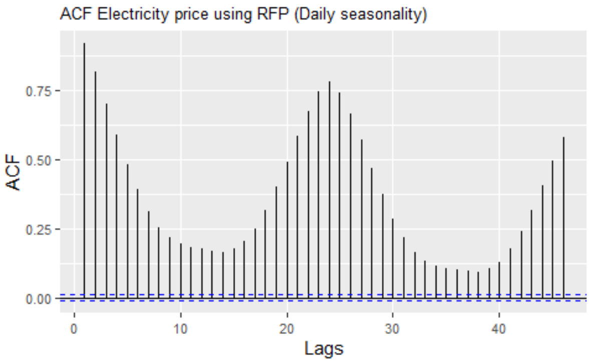
Appendix 4. ACF

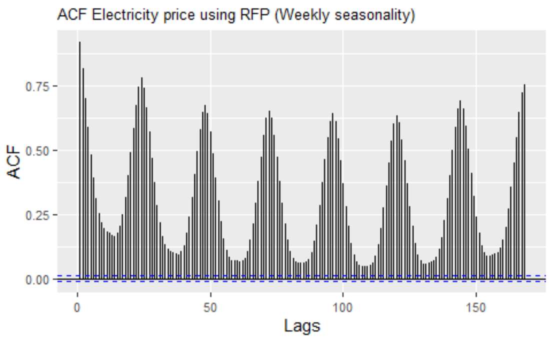


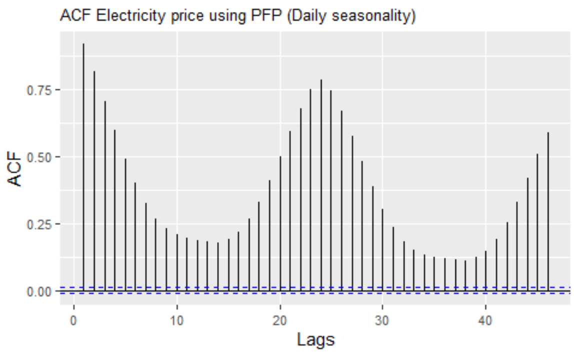


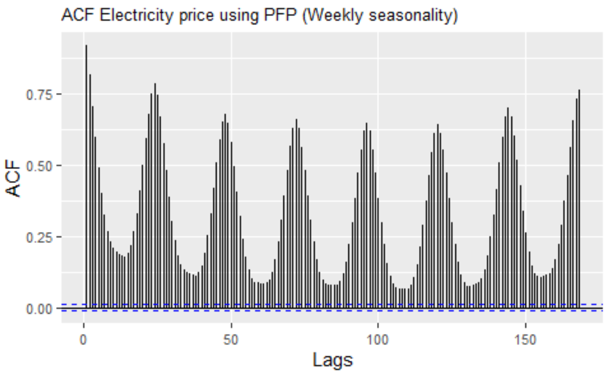












Appendix 5. A summary of models forecasting accuracy measures after hydro volumes elimination

Outlier detection method	**Models**									
	ARIMA (1,0,0)		ARIMA(0,1,0)		ARIMA(1,1,1)		ARIMA (3,1,1)		ARIMA (3,1,2)	
	Uni	Multi	Uni	Multi	Uni	Multi	Uni	Multi	Uni	Multi
Accuracy measure - RMSE										
TFP	5.81	5.28	5.94	5.50	5.89	5.23	5.89	5.22	5.11	5.06
SFP	5.67	5.15	5.79	5.35	5.74	5.09	5.47	**5.08**	4.97	4.93
RFP	5.81	5.28	5.93	5.49	5.89	5.23	5.88	5.22	5.10	5.06
PFP	5.61	5.10	5.73	5.31	5.70	**5.03**	5.70	**5.03**	4.86	**4.82**
Accuracy measure - MAE										
TFP	3.36	3.27	3.13	3.11	3.07	3.26	3.07	3.24	3.12	3.10
SFP	3.32	3.22	3.09	3.07	**3.03**	3.21	3.23	3.20	3.09	3.06
RFP	3.36	3.27	3.13	3.11	3.07	3.25	3.07	3.24	3.12	3.10
PFP	3.30	3.21	3.08	3.05	**3.03**	3.20	**3.03**	3.19	3.07	**3.04**