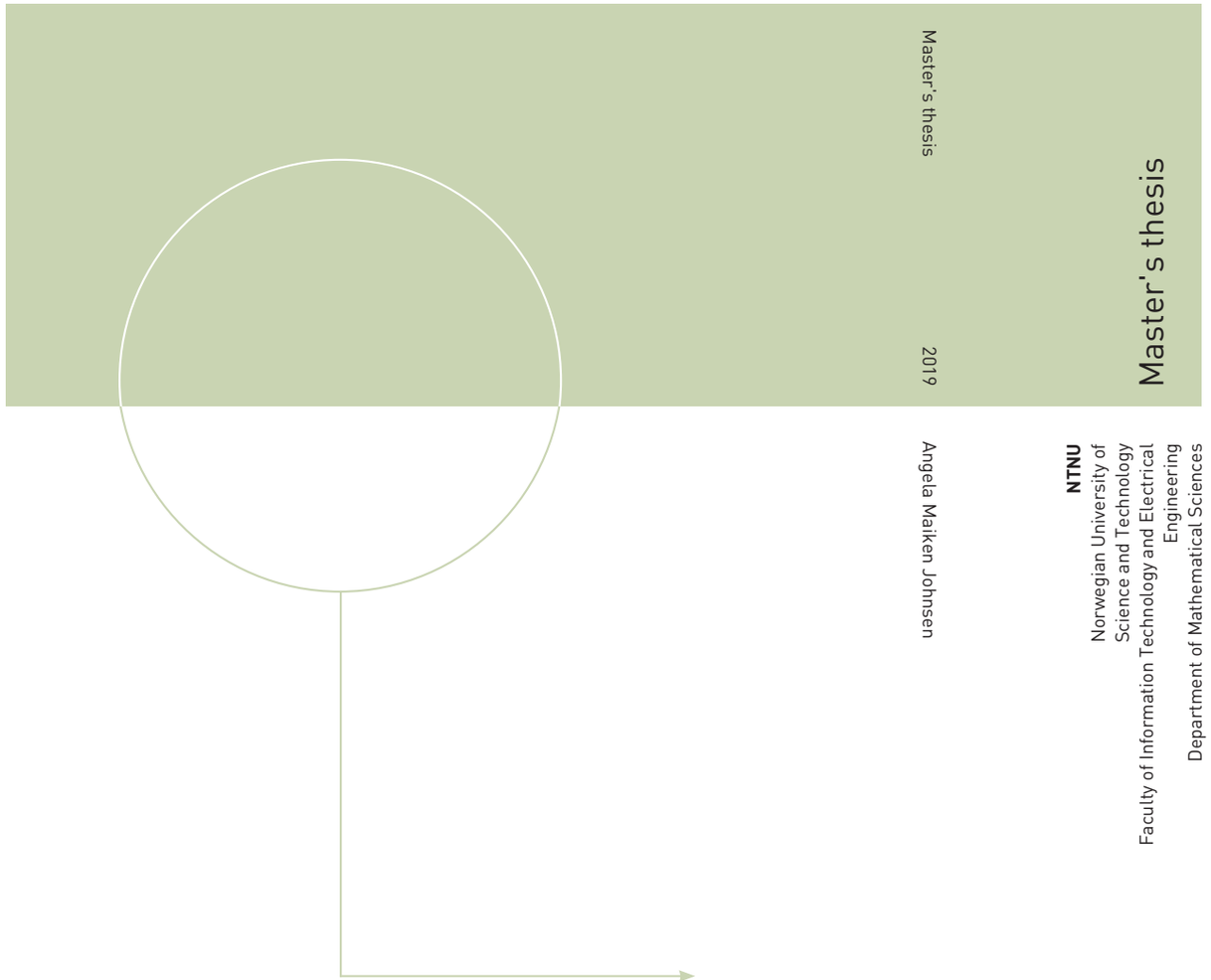


Angela Maiken Johnsen

Forecasting Day-Ahead Electricity Spot Prices, With Applications to the German Electricity Market

July 2019

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

Forecasting Day-Ahead Electricity Spot Prices, With Applications to the German Electricity Market

Angela Maiken Johnsen

Master of Science in Applied Physics and Mathematics
Submission date: July 2019
Supervisor: Jo Eidsvik

Norwegian University of Science and Technology
Department of Mathematical Sciences

Abstract

In this thesis the German electricity market is studied, with the aim of predicting the day-ahead electricity spot prices. Three forecast models are presented; the first model being a persistence model, which serves as a baseline model, only incorporating previous values of the electricity spot prices. The second model is a rolling window regression (RWR) model, while the third model is a state-space model. The two latter models incorporate previous electricity spot prices in the model, as well as values of other fundamental variables that are assumed to influence the the day-ahead electricity spot price. These include historical oil, coal and gas prices, spot price volatility and electricity demand, to name a few. The RWR model uses information from a window of previous days, whereas the state-space model only considers the preceding day.

The parameter estimation of the RWR model is performed using the least squares equivalents of this model. As for the state space model, a Kalman filter is implemented to perform parameter estimation. The estimates from these models are then used to forecast the spot prices, in addition to perform inference on the estimates.

When comparing the models it is found that both the RWR and state-space model outperform the persistence model, with the RWR model having the most accurate predictions. Evidence indicates that this is due to the electricity prices having larger memory than one day. Unlike the persistence model, both the RWR model and the state-space model manage to capture some of the stochastic nature of the spot price.

Sammendrag

Denne masteroppgaven har studert det tyske elektrisitetsmarkedet, i den hensikt å predikere neste dags elektrisitetsspotpriser. Tre modeller er presentert; den første modellen er en ”persistence”-modell, som fungerer som en grunnmodell som kun inkluderer tidligere elektrisitetsspotpriser. Den andre modellen er en ”rolling window regression”-modell (RWR), og den tredje er en tilstandsmodell. De to sistnevnte modellene inkluderer, i tillegg til tidligere priser for elektrisitetsspotkontrakten, også verdier til andre variabler som antas å påvirke neste dags spotpris. Disse inkluderer historiske olje-, kull- og gasspriser, spotprisvolatilitet og etterspørsel etter elektrisitet, for å nevne noen. RWR-modellen utnytter informasjon fra et vindu av tidligere dager, mens tilstandsmodellen kun tar i bruk informasjon fra dagen før.

Parameterestimering for RWR modellen foregår gjennom den tilsvarende minste kvadraters metode. For å estimere parametrene i tilstandsmodellen er et Kalmanfilter blitt implementert. Estimatene fra begge modellene brukes så til å predikere spotprisene i tillegg til å utføre inferens.

Ved sammenligning av modellene ble det funnet at både RWR- og tilstandsmodellen ga bedre resultater en persistence-modelle, hvorav RWR-modellen hadde de mest presise prediksjonene. Dette tyder på at elektrisitetsspotprisene har lengre minne en en dag. I motsetning til persistence-modellen klarer både RWR- og tilstandsmodellen å fange opp noe av den stokastiske oppførselen til spotprisene.

Preface

This thesis concludes my studies at the Norwegian University of Science and Technology (NTNU) at the study programme Applied Physics and Mathematics. The work was carried out during the spring of 2019 at the Department of Mathematical Sciences.

I would like to thank my supervisor, Jo Eidsvik, for his invaluable support during this semester. Our weekly meetings have provided me with both knowledge and motivation, and to that I am very grateful. I would also express my gratitude towards Florentina Paraschiv, who has provided me with data and domain knowledge of the German electricity market. Lastly, to my friends, family and boyfriend: Thank you for your unwavering support to this day.

Trondheim, July 2019
Angela Maiken Johnsen

Table of Contents

Chapter 1

Introduction

Ever since the first light bulb was lit in the 19th century, electricity has been crucial in the development of the modern society. Today it is nearly impossible for any civilised society to imagine a life without electricity, but our extensive consumption of it has led to irreversible global challenges we now suffer from. At the same time, electricity is essential to alleviate developing countries from poverty and to stimulate economic growth. Thus, to keep global warming at bay, without the expense of drastically decreasing the electricity consumption, countries around the world are forced to focus more on the renewable energy sources, which are more dependent on the weather than the traditional ones. The power generation from these sources are therefore subject to short-time changes and has a higher volatility compared to traditional sources, making it an important price driver for the electricity markets [1].

Whereas the early electricity markets were characterised by governmental monopolies and regulations, today's markets are to a greater extent privatised and deregulated. With this new market structure, competition is present, where the electricity prices are governed by the law of supply and demand, among other things. Now, utility companies have to take other factors into account than before, such as the market situation as well as financial risk [2]. Considering the above-mentioned factors, models that capture the dynamics of the electricity prices are of interest for the market participants in order to make decisions and plan ahead [3].

Numerous modelling approaches have been developed in order to predict the electricity prices in the different markets in the world, and the preferred models depend, among other things, on the time horizon of the prediction. An attempt to classify the numerous models were carried out in [4], in which six different groups

of electricity price forecast models are proposed: game theoretic, fundamental, reduced-form, statistical and artificial intelligence-based models. However, many of the models considered in the literature are hybrid, combining techniques from two or more of the groups mentioned.

In this thesis the aim is to predict the day-ahead electricity spot prices in Germany. Data from the German electricity market were provided by Florentina Paraschiv, and include the prices of the electricity spot contract, coal, gas and oil, as well as the demand, to name a few. These data are from the year 2010 to 2016. Three models, which can be categorised as statistical and fundamental models, are then formulated with the aim of predicting the day-ahead electricity spot prices in Germany. These models are the persistence model, rolling window regression model and the state-space model. A description of the electricity market, as well as a more detailed presentation of the data, are given in chapter 2. In addition, a more detailed literature review is given here, focusing on the fundamental, statistical and artificial intelligence-based models. Further, in chapter 3, relevant theory for the development of the models will be presented, including means of evaluating their predictive performance. In chapter 4 the models are presented, discussed and compared, while the final chapter, chapter 5, concludes the thesis as well as suggesting further work.

Chapter 2

The Electricity Market

In this chapter the relevant background of this thesis is presented, starting in 2.1 by presenting the market from which the data is collected, that is the European Energy Exchange. Then the price formation on the electricity spot market will be explained in further detail, followed by a presentation of the data used in the thesis, including some data analysis. Lastly, a literature review focusing on the fundamental, statistical and artificial intelligence-based models, is conducted.

With the liberalisation of the power sector the need for organised markets at the wholesale level emerged. In these markets electricity is treated as a commodity, and different electricity contracts may be traded much similar to other financial assets and commodities. These contracts can be sold over-the-counter or in organised markets, the latter of which will be the focus in this thesis. The contracts may be physical contracts for delivery of electricity or financial contracts for hedging or speculations [5]. Further, the various physical contracts have different maturities, and they are classified as either long-term or short-term, that is spot contracts. The former includes futures, forwards and bilateral agreements which may have maturities exceeding one year. It is the latter that will be the focus for this thesis, so the long-term contracts will not be explained any further. As for the short-term contracts there exist both day-ahead and intraday markets, but this thesis will only look at the day-ahead market.

2.1 The European Energy Exchange

In the early days of electricity markets their only purpose was to deliver electricity to customers. Today, however, their role has expanded to also address the new challenges in the political and environmental climate. Now the electricity should be produced in a sustainable manner to ensure environmental protection and prevent climate change mitigation [6]. In addition they serve as regulated and neutral marketplaces, so that all producers and consumers have non-discriminatory access to the market [7]. This thesis will focus on a particular market, that is the German electricity market, operated by the European Energy Exchange (EEX). EEX is the leading energy exchange in central Europe and was founded in 2002 in Leipzig, Germany [8]. EEX offers various energy and commodity products; power, natural gas, emission allowances and oil, to name a few [9].

As mentioned, this thesis will focus on the electricity spot market, more specifically the European Power Exchange (EPEX), which is part of the EEX group. EPEX consists of Germany, as well as France, United Kingdom, the Netherlands, Belgium, Austria, Switzerland and Luxembourg, which represent $50\ \%$ of the electricity consumption in Europe [10]. The spot market is a day-ahead market where the spot is an hourly contract between producers and consumers with physical delivery of electricity the following day. The 24 spot prices are determined in a daily auction, which takes place at 12:00 (noon) every day. Here market participants bid on the spot contract, in €/MWh, for hourly blocks the following day, with information from the previous days in mind. The bidding process is schematically illustrated in figure 2.1. The final spot price for the individual hours, the market clearing price, is then determined after all the bids have been collected by the principles of supply and demand [11].

To understand how the electricity spot prices are determined, it is important to understand the concept of the merit order curve (the supply curve) and the demand curve. The merit order curve is constructed based on the merit order principle, meaning that the suppliers' electricity offers are ranked based on the marginal costs of the power plants used to generate it. One example of this (stepwise) curve is illustrated in figure 2.2, with the producers' supply bid in MWh on the horizontal axis and the marginal cost in €/MWh on the vertical axis. Electricity production from renewable energy sources, such as wind and solar, has the lowest marginal cost, and may even be negative when factoring in the renewable support schemes [6]. These sources are then followed by, with increasing marginal cost, the power plants covering the base load (lignite and nuclear), coal, gas and oil fired power plants. As can be seen from figure 2.2, the marginal cost of a power plant depend

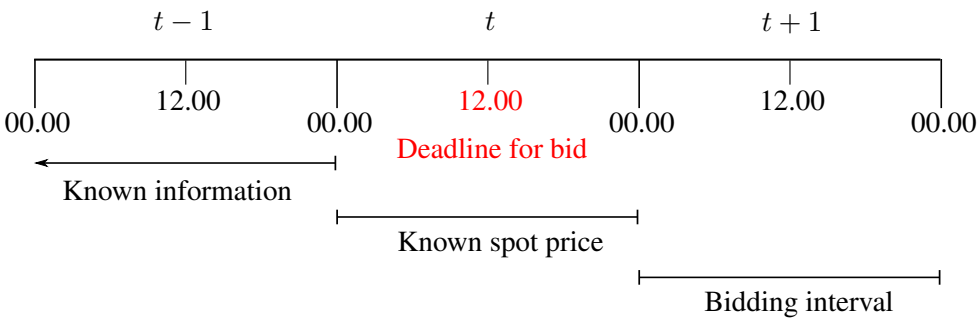


Figure 2.1: Overview of the bidding process. At day t the bids for the electricity prices for day $t+1$ are determined in hourly intervals, the first interval being from 00:00-01:00, the second 01:00-02:00 etc. At the time of the bidding the spot prices for all intervals of the current day are known, as these were settled upon the day before.

on the fuel cost and the emission cost, as shown in grey and blue, respectively. The merit order curve also shows the preferred sources at any time, favouring the power production from the power plants having the lowest marginal cost. Note that this curve is not constant in time, and is affected by the weather, changes in fuel prices and power plant outages, among other things.

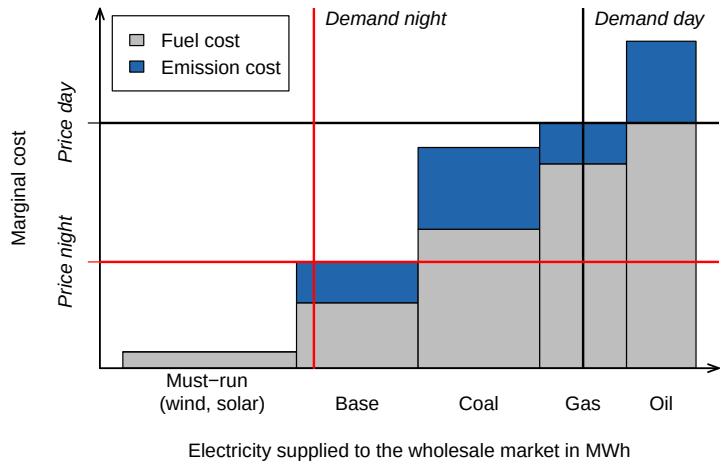
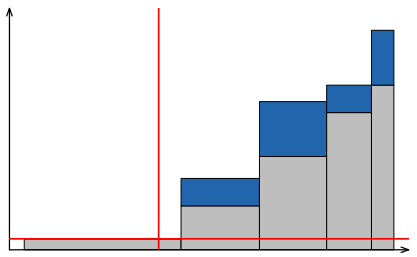


Figure 2.2: The merit order curve, inspired by [12]. The vertical lines represent the demand during night (red) and day (black). The intersections between the demand and supply are illustrated by the horizontal lines, determining the prices during the night and day

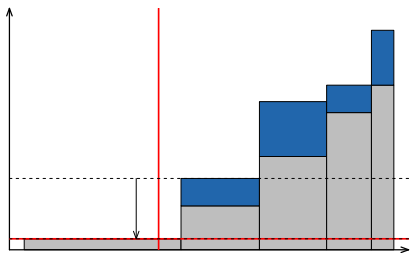
With the merit order curve the *market clearing price*, that is the spot price, is determined by the intersection of the supply and demand curve. As the demand is nearly inelastic in the short term [7], the demand curve is illustrated by a vertical line. This is due to the fact that electricity is considered a necessity, and an increase in the price is not likely to effect consumers' behaviour to a great extent as there exist no other alternative to electricity. During the day the demand is naturally higher than during the night hours, meaning that during the hours with high demand more CO_2 intensive power plants will often be used to meet the demand. In turn this leads to higher electricity prices during these hours. During the night, on the other hand, it typically suffices with the electricity generated by wind and the power plants covering the base load to cover the demand.

Another peculiar aspect that can be deduced from figure 2.2 is the *merit order effect*. This is a term for describing the decrease in the electricity prices due to an increase in electricity produced by renewable energy sources. This effect is illustrated in figure 2.3; in both 2.3a and 2.3b the demand is the same, but in the latter the supply from renewable energy sources is larger, resulting in a lower electricity price than in the former case, illustrated by the arrow.

Figure 2.3: Illustration of the merit order effect. The axes and the bars are the same as in figure 2.2, and the vertical and horizontal line represent the demand and intersection with the merit order curve, respectively. As can be seen, this particular intersection decreases from the left panel to the right panel, as indicated by the arrow. This results in a lower spot price.



(a) The merit order curve of a certain day.



(b) The merit order curve of another day with higher supply of electricity generated from renewable energy sources, but with same demand as in figure 2.3a.

2.2 Data

In this section, the data provided for this thesis will be presented. These include the German electricity spot prices, coal, gas and oil prices, price for CO_2 emission allowances (EUA[1]), expected infeed from photovoltaics and wind, expected power plant availability and expected demand were provided for the analysis. These were, as mentioned, supplied by Florentina Paraschiv, one of the authors behind the article "The Impact of Renewable Energies on EEX Day-Ahead Electricity Prices" [11], which this thesis is influenced by. The period the data come from is from the 1st of January 2010 to the 31st of August 2016, except from the spot price, with the first entry from the 25th of December 2009. The data are summarised in table 2.1, and includes a description of these. Summary statistics may be found in table 2.2, while the plots of the development of the variables with daily granularity are shown in the figures 2.4.

Before describing the data any further, some of the terms found in table 2.1 will be explained. As seen, the value for the coal price is the latest available price of the front-month Amsterdam-Rotterdam-Antwerp (ARA) futures contract, that is the price of the futures contract having the closest expiration date. ARA denotes the port in this specific area, and is one of the most important gateways for coal in Europe [7]. As for the gas price, the value in the data is the NCG (NetConnect Germany) day-ahead natural gas spot price, which constitutes the relevant gas spot price for the German area [13]. Lastly, the value for the oil price is the active ICE (Intercontinental Exchange) Brent Crude futures contract, which is the futures contract for a specific type of oil originating from the North Sea [7].

Furthermore, it is important to understand why the data in table 2.1 are included in the analysis. With the merit order curve in figure 2.2 in mind, it is logical that the prices of coal, gas, oil and CO_2 emission allowances are included in a formulation of a fundamental model. These prices directly influence the electricity generation from the fossil-fired power plants, hence having an impact on the merit order curve. The expected infeed of electricity from the renewable energy sources is included because of the merit order effect; as seen in figure 2.3, an increase in the infeed from electricity generated by photovoltaic and wind tends to decrease the electricity price. The demand also influences the spot price, as it determines the intersection with the merit order curve. Lastly, the power plant availability is also included, because it simply gives as a cap for the potential electricity production at a given time. As for the spot price volatility, this will be explained in chapter 4.

[1] One EUA permits emitting one tonne of CO_2 equivalent.

Table 2.1: A list of the data provided for this thesis including a description of these. The units and granularity of the data are also given, the latter being either hourly (h), that is 24 values per day, or daily (d), that is one value per day. The variable marked with * is a derived variable.

Variable (units) [granularity]	Description
Spot price (€/MWh) [h]	Market clearing price for all hours.
Spot price volatility* (€/MWh) [h]	Standard deviation of the market clearing prices for the same hour of the five last delivery days.
Coal price (€/12,000 t) [d]	Latest available price of the front-month ARA futures contract before the electricity price auction occurs. This is auctioned daily.
Gas price (€/MWh) [d]	Last available price of the NCG day- ahead natural gas spot price on the day before the electricity price auction takes place.
Oil price (€/barrel) [d]	Last available price of the active ICE Brent Crude futures contract on the day before the electricity price auction takes place.
Price for EUA (€0.01/EUA) [d]	Latest available price for the EEX Carbon Index. Auctioned daily at 10:30 am.
Expected PV and wind infeed (MWh) [h]	Sum of expected wind and photovoltaic electricity infeed into the grid. Published by German transmission system operators after the electricity price auction.
Expected power plant availability (MWh) [d]	Ex-ante expected power plant availability for electricity production on the delivery day. Published daily at 10:00 am.
Expected demand (MWh) [h]	Demand forecast data for the relevant hour on the delivery day.

As can be seen in table 2.2, as well as in figure 2.4a, the spot prices may also be negative, which is one of the aspects that make electricity spot prices different from that of other financial assets or commodities. The occurrences of these prices are marked with red dashed lines in all the plots in figure 2.4, in total 10 occurrences.[2] These counterintuitive prices occur in times when low demand meets a high inflexible power generation, often accompanied with a higher power generation from fluctuating sources like wind. This fact is supported by the plots in figure 2.4; as seen in figure 2.4c the negative prices coincide with high infeed from wind and solar power, while the demand is low. The occurrences of negative prices mostly happen during nighttime, when these events occur. The fact that wind constitute a large part of the electricity generation during nighttime also make the spot prices during these hours more volatile than during the day.

Table 2.2: Descriptive statistics of the variables in table 2.1, with corresponding units.

Variable	Mean	Std. dev.	Minimum	Maximum
Spot price	38.55	16.55	−222.99	210
Coal price	7.88	1.73	4.61	12.16
Gas price	21.44	4.69	11	39.50
Oil price	40.37	10.12	15.02	56.66
Price for EUA	8.46	3.80	2.72	16.84
Expected PV and wind infeed	9157	7077	263.4	44607
Expected power plant availability	55323	4862	40016	64169
Expected demand	41557	8548	18233	63715

The reason why the spot prices may be negative can be explained by the fact that electricity is economically non-storable [4], meaning the power system requires a constant balance between production and consumption. Thus, prices may fall in times of low demand to signal generators to reduce the power generation to avoid overloading the grid, and may even be negative because the cost of shutting down and restarting a power plant may exceed the cost of accepting negative prices [14]. With 24 hourly blocks, there are 58608 spot prices in the spot price data provided for this thesis, and 401 occurrences of negative prices. With only 0.68 % of the prices in the data set being negative, this phenomenon is rather rare.

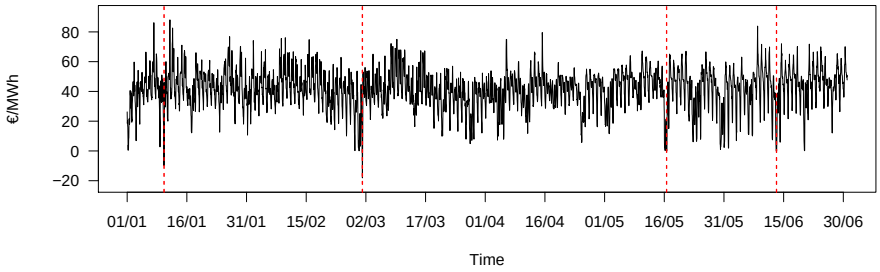
[2] It appears to only be four occurrences, however the negative prices often occur in consecutive hours, causing the dashed lines to (seemingly) coincide.

Another peculiar aspect of the electricity spot is its *mean reverting* behaviour [5]. Mean reversion is the assumption that the price of a commodity, or a financial asset, will tend to move to its long-term average price. Whereas stock prices are governed by the principles of supply and demand only, and thus can become arbitrarily large or small (but never negative), electricity spot prices are in addition to being influenced by supply and demand, also associated with the cost of generating the electricity itself. Although there exist price spikes in both the negative and positive direction in the short run, electricity spot prices express a mean reverting behaviour in the long run reflecting the cost of production of this particular commodity [15]. Even when sudden price spikes occur in the short run, the prices rapidly return to the previous price level.

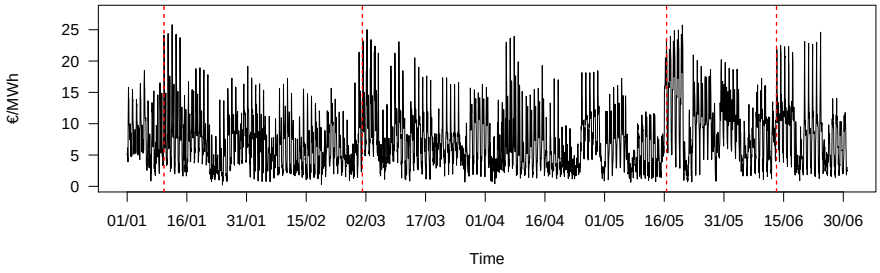
To inspect the correlation between the different variables, a correlation matrix is found in figure 2.5. First, as can be seen, the spot prices and the demand are highly correlated, with a positive correlation of 0.59. This is due to the fact that supply and demand of electricity must balance at every moment. In addition, as seen in the previous section, the demand determines the intersection with the merit order curve, which sets the spot price. To inspect the correlation further, plot of the standardised spot price, shown in black, and expected demand, shown in red, of the first week of 2010 is presented in figure 2.6. The two curves seem to follow each other, however the spot price development is much more ”extreme” in the sense that its curve is spikier than that of the demand. Both the spot price and the demand expresses a seasonal behaviour in the daily pattern, as seen in figure 2.6, but also as in the weekly and yearly patterns. The spiky nature of the spot price development may be explained by the fact that the underlying factors that affect the spot price are unpredictable, like the weather, power plant outages and transmission constraints [1], whereas the demand is more easily predicted. According to said article, an appropriate forecasting model for electricity spot prices should consider the deterministic patterns as well as these stochastic components.

Furthermore, the fluctuant nature of the renewable energy sources can be seen from figure 2.4c. In addition, as seen from the correlation matrix in figure 2.5, both wind and photovoltaic electricity are negatively correlated to the spot price. This supports the merit order effect discussed earlier, that is that an increase in the electricity generated by renewable sources tend to decrease the electricity spot prices. From figure 2.5 it is seen that the wind infeed has a stronger correlation to the spot price than do photovoltaic, more specifically -0.39 against -0.08. With the elecricity prices being correlated to these volatile variables, especially wind, partly explains the stochastic nature of these prices.

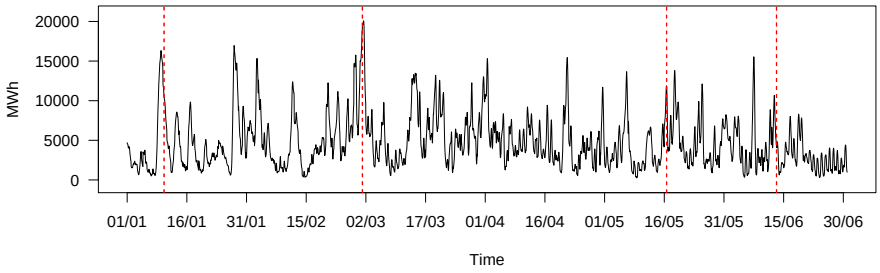
Figure 2.4: Plots of the development of the variables with hourly granularity, as presented in table 2.1. The data are taken from the first half of 2010. The dates on the first axis are presented in the day/month format, while the units are given on the second axis. The red dashed lines indicate the occurrences of negative spot prices.



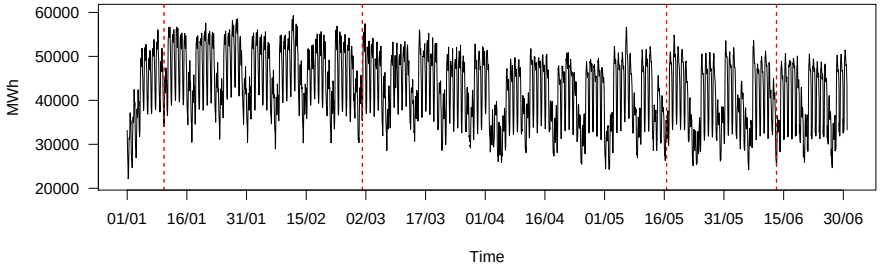
(a) Plot of the price development of the spot in the period considered.



(b) Plot of the volatility development of the spot in the period considered.



(c) Plot of the expected photovoltaic and wind infeed in the period considered.



(d) Plot of the expected demand in the period considered.

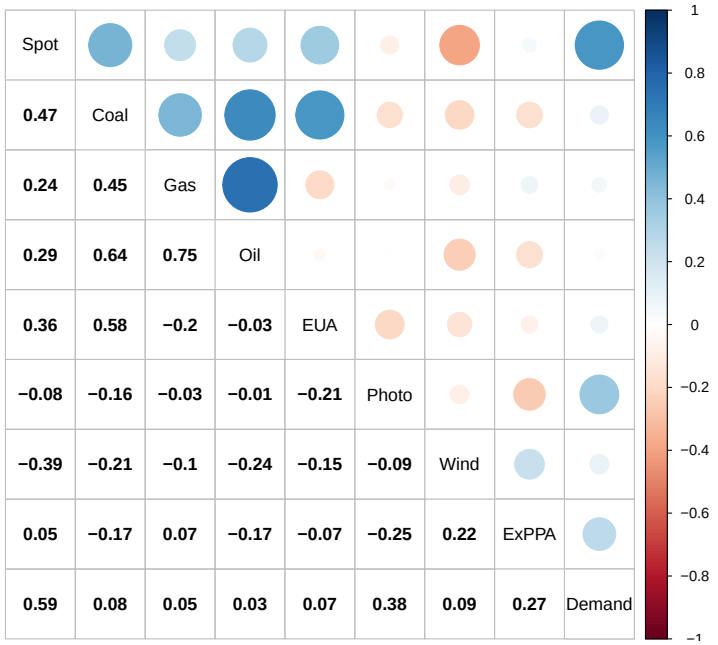


Figure 2.5: Correlation between the data in table 2.1, presented as numbers (lower triangle) and as colours (upper triangle). The correlation is colour coded as shown by the bar to the right. Here ”ExPPA” is an abbreviation for expected power plant availability.

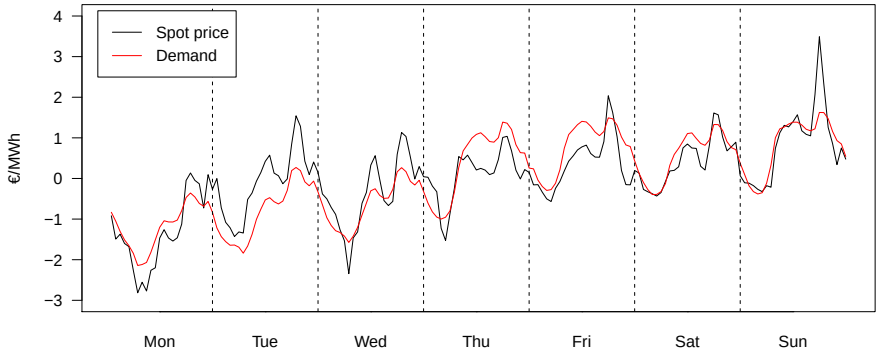


Figure 2.6: Plot of the spot price and expected demand of the first week in 2010. The vertical dotted lines are included to separate the days.

Lastly, to inspect the correlation between the prices of oil, gas, emission allowances and coal, a plot showing the development of these prices is shown in figure 2.7. Prior to plotting them, the prices were standardised in order to compare them, as these are of different magnitude. From the plot it seems that the price developments follow each other, and this is also supported by the correlation matrix in figure 2.5. As seen, the correlation between gas and oil is the highest, with a correlation of 0.75. This is also the strongest correlation found in the dataset. The second strongest correlation is found between the coal and oil prices, with a correlation coefficient of 0.64. As for the price of emission allowances, this is positively correlated to the coal price, with a correlation coefficient of 0.58, while it is negatively correlated to the gas and oil price. Due to the high correlations, a change in one of the variables influences the others. As for oil and gas the relationship is nearly linear. In comparison, the correlation between the demand and spot prices is 0.59, which is also considered a strong correlation.

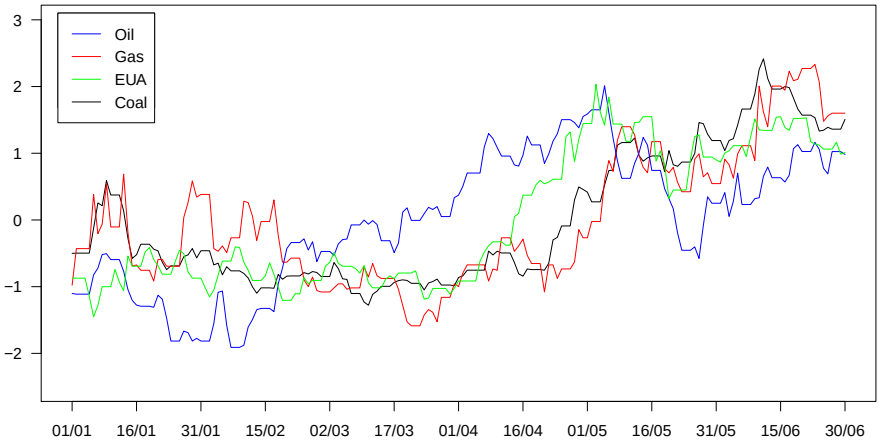


Figure 2.7: Plot of the development of the standardised prices of oil (blue curve), gas (red curve), EUA (green curve) and coal (black curve) in the first half of 2010.

2.3 Literature Review

In this section a literature review will be conducted, with an emphasis on the fundamental, statistical and artificial intelligence-based models. Models in the literature also include game theoretic and the reduced-form models, however, these will not be detailed any further. This is due to the fact that the former groups of models do not focus on predicting the electricity prices per se, but rather intent to analyse the strategic behaviour of the agents in the market. As for the latter groups of models, the reduced form models, these are not useful for short-term electricity pricing, as these focus on the long-term development of electricity pricing. Thus, these models are more important in derivatives valuation and risk analysis. The research involving electricity price forecasting is vast, especially after the deregulation of these markets became a fact. Much of the literature review in this section is based on [4], an excellent paper treating the field of electricity price forecasting, which also gives a more extensive literature review than presented here.

The first type of models that will be presented are the fundamental models. These are models which attempt to capture the physical and economic aspects that are present in the electricity market. As argued by [4], this group of models can be divided into parameter rich models, utilising variables that are assumed to influence the electricity prices, or parsimonious structural models of supply and demand. In [16] an approach is made to model the electricity spot prices in the Nordic market, using stochastic processes of fundamental variables. In this article the fundamental variables are climatic, hydropower, demand and base load supply data. However, as concluded in this particular paper, this method is better at predictions for longer time periods than the day-ahead market, making this method better suited for analysing company risk. Another paper dealing with fundamental models is [17], which developed both regression-based and time series models using fundamental variables such as demand, demand volatility and fuel prices, to name a few. Using these models, the aim was to predict the electricity prices in the day-ahead market as well as the intra-day market in Britain. In the article which this thesis is influenced by, [11], a state-space model incorporating fundamental variables from the German electricity market was formulated and solved by the Kalman filter. In this, the fundamental variables coincide with those of table 2.1. The aim here, however, was to observe the impact the fundamental variables had on the electricity spot price, and did not reach an overall conclusion about the predictive performance of the model compared to other models.

Further, the statistical models use statistical techniques to forecast the electricity prices, using data on the previous electricity prices or previous or current values of

exogenous variables. The statistical models mainly comprise regression and time series models, as well as the popular similar-day methods. The latter methods searches historical data and finds a day having the same characteristics as the day of interest. One of the articles mentioned in the previous article, [17], developed regression models using time-varying parameters, thus including a time series approach. The simplest cases comprise models of univariate time series, where the spot price is dependent on the price from previous days or hours. In [18] various univariate autoregressive (AR) and autoregressive moving average (ARMA) models were proposed and tested on the German market. The spot prices were tested as a single series for the whole time period considered, or as 24 different series modelling each hour of the day. The aforementioned article found that the models performed better on the hourly series. Pure AR type models do not take into account that the electricity spot price may be influenced by other time series, but by combining time series models with fundamental models, this is issue is resolved. This is done in for example [19], which utilises variables based on technologies of the power plants, market concentration, congestions and volumes. In this article, a GARCH structure of the residuals was also included, which is popular for modelling the volatility in the electricity markets. In fact, some of the characteristics of financial time series, including electricity spot prices, are volatility clusters. This phenomenon occurs because the variability of the time series is dependent on its own past.

Lastly, the artificial intelligence based methods consist mainly of various neural networks (NNs) and support vector machines (SVMs). These methods are ”intelligent” in the sense that they learn from the data that is fed into them. These are flexible methods that can handle the complexity of the electricity markets, as well as the non-linearities that may be present in determining the electricity spot prices [5]. In [1], [20] and [21], artificial neural networks were tested on forecasting short-term electricity prices. In the first article, an ANN was implemented and trained using fundamental data to model the day-ahead electricity spot prices in Germany. The results showed that the forecasts errors are competitive to those of the other models tested, and even better in some cases. The other models included a time series approach, modelling the electricity spot price as a time series, accounting for daily seasonality, and two other naïve models. These forecasted the electricity spot prices using the price of the same hour from the previous day or week as forecasts. In the second article, an ANN model based on a similar day method was implemented and tested on the North-American electricity prices, compared with a direct similar approach. In this article, data on load and prices were used to characterise the days. This article found that the ANN based model performed better than the similar day approach alone. In the third article the ANN

was tested using nothing but the historic electricity prices in Spain an California, comparing the results with time series models found in the literature, with success. Recurrent neural networks have also been implemented with the aim of predicting the electricity spot prices, and tested on the markets in Spain and New York [22], with greater accuracy than other, more traditional models. Further. using SVM in the electricity price forecast literature is different than predictions per se. As SVMs are classification tools, these models are first trained as a classifier using part of the data, then exploit this to classify (or predict) the other data. Examples of SVM in the electricity price forecast literature include, among other things, [23] and [24]. In the first article it is shown that the SVM implemented performs better than a specific type of NN, that is radial basis function neural network. The second article trains the SVM with fundamental data that is assumed to influence the electricity price, and shows a better predictive performance than that of an NN.

Chapter 3

Theory

This chapter introduces theory relevant for this thesis. First some relevant details and theory behind time series will be explained, starting with the basic building blocks of these series. Further, the theory behind state-space models is introduced, which are models that are capable of handling a wide range of time series models. The Kalman filter is also presented, which utilises the state-space representation of a system to solve it. Next, linear regression and the connection with state-space models will be presented. Lastly, the evaluation methods which are used to evaluate the predictive performance of the models are presented.

3.1 Time Series

Time series are series of observations y_t that are measured over a set of times $t = 1, \ldots, N$ and may have either continuous or discrete sample spaces. These models assume that the observations depend on each other in some manner related to time, be it daily, weekly, monthly or yearly. Time series are widely used; in finance and economics they are, among other things, used for modelling daily closing stock prices. In demography one may want to study a particular population, and the population size may be modelled by a time series. In environmental studies time series may model the concentration of a certain particle in the air. The range of applications is huge. Based on previous observations, one goal is to fit a statistical model for the time series, and then do forecasting with these models.

3.1.1 Autoregressive Processes

An autoregressive (AR) process is a representation of a time series model where the current state is defined via its former values multiplied with some weights plus error [25]. More formally, an AR process of order p, AR(p), is a representation of a zero-mean process, $y_t, t = 1, \ldots, N$, where

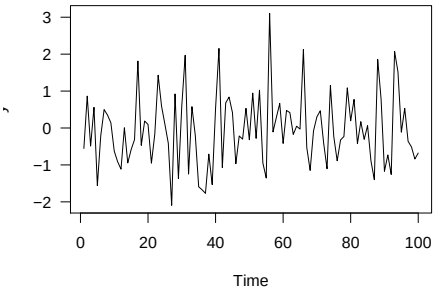
$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \tag{3.1}$$

where p is a non-negative integer. As can be seen from the above equation the process depends on the p previous values of the series, where the ϕ_is determine the dependency of the previous values.

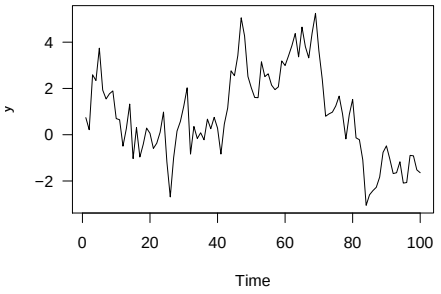
A simple example is the $\mathrm{AR}(1)$ process, which takes the form

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \tag{3.2}$$

where $\phi_1 \neq 0$ and ε_t is as before. This process has been simulated in R over 100 time units for $\phi_1 = 0.1$ and $\phi_1 = 0.9$, which are presented in figure 3.1a and 3.1b, respectively. In both cases $\sigma^2 = 1$.



(a) $\mathrm{AR}(1)$ process with $\phi_1 = 0.1$



(b) $\mathrm{AR}(1)$ process with $\phi_1 = 0.9$

Figure 3.1: Simulated $\mathrm{AR}(1)$ processes over 100 time units with two values of ϕ_1. Here $\varepsilon_t \sim \mathcal{N}(0, 1)$.

As can be seen, the graph of the process in figure 3.1a is rather jagged compared to that of figure 3.1b. This can be explained by the fact that when $\phi_1 = 0.1$ the process depends less on the previous value than when $\phi_1 = 0.9$. As a consequence, the error term ε_t has a larger impact on the process, resulting in a jagged graph more similar to white noise. As for figure 3.1b the values of the process seem to

be more dependent on the previous value, resulting in a smoother graph.
The properties of time series are well understood because of the linear combination of Gaussian terms. For $\mathrm{AR}(1)$ processes, see (3.2), the variance becomes

$$\mathrm{Var}(y_t) = \mathrm{Var}(\phi_1 y_{t-1}) + \mathrm{Var}(\varepsilon_t) = \phi_1^2 \mathrm{Var}(y_{t-1}) + \sigma^2.$$

When the mean and variance of a process are independent of time, the process is said to be stationary [26]. Assuming that the $\mathrm{AR}(1)$ process is stationary, meaning $\mathrm{Var}(y_t) = \mathrm{Var}(y_{t-1}) = \sigma_y^2$, the mean and variance become

$$\begin{aligned}
&\mathrm{E}[y_t] = \phi_1^2 \mathrm{E}[y_{t-1}] + \mathrm{E}[\varepsilon_t] = 0, \\
&\mathrm{Var}(y_t) = \sigma_y^2 = \phi_1^2 \mathrm{Var}(y_{t-1}) + \mathrm{Var}(\varepsilon_t) = \phi_1^2 \sigma_y^2 + \sigma^2,
\end{aligned}$$

where the latter can be written as

$$\sigma_y^2 = \frac{\sigma^2}{1 - \phi_1^2}. \tag{3.3}$$

As can be seen, both the mean and variance of the process are time independent. Because $\sigma_y^2 > 0$, one must require that the denominator be greater than 0, meaning $\phi_1^2 < 1$ or $|\phi_1| < 1$ in order for the process to be stationary.

Returning to the plots in figure 3.1, where $\sigma^2 = 1$, it can be seen that when $\phi_1 = 0.1$ the variance is, according to (3.3), $\sigma_y^2 = 1/(1 - 0.1^2) \approx 1.01$. When $\phi_1 = 0.9$ the variance becomes $\sigma_y^2 = 1/(1 - 0.9^2) \approx 5.26$. As can be seen from figure 3.1, $y_t \in (-2, 3)$ in figure 3.1a, while $y_t \in (-2, 5)$ in figure 3.1b, which correspond well with the calculated variances. In both plots the graphs are symmetric around 0, which corresponds well with the theoretical mean of 0. Alternatively one can enforce $\sigma_y = 1$, so that $\sigma^2 = 1 - \phi_1^2$. With a $\phi_1 = 0.1$, the residual variance becomes $1 - 0.1^2 = 0.99$, whereas when $\phi_1 = 0.9$, it is equal to $1 - 0.9^2 = 0.19$.

A non-stationary $\mathrm{AR}(1)$ process occurs when $|\phi_1| \geq 1$. Then $\mathrm{Var}(y_t) \neq \mathrm{Var}(y_{t-1})$, so that $\mathrm{Var}(y_t) = \phi_1^2 \mathrm{Var}(y_{t-1}) + \sigma^2$. As $|\phi_1| > 1$ it can be seen that the variance increases with time. A special case of a non-stationary process, common in the field of finance, is the one-dimensional *random walk*. This is an $\mathrm{AR}(1)$ process with $\phi_1 = 1$,

$$y_t = y_{t-1} + \varepsilon_t, \tag{3.4}$$

with initial conditions $y_0 = 0$ [26]. Then it can easily be seen that (3.4) may be written as a sum of white noises, that is $y_t = \sum_{i=1}^{t} \varepsilon_i$. It follows that $\mathrm{E}(y_t) =$

$\sum_{i=1}^{t} \mathrm{E}[\varepsilon_i] = 0$, which trivially is independent of time. The variance, on the other hand, becomes

$$\begin{aligned}
\mathrm{Var}(y_1) &= \mathrm{Var}(y_0) + \sigma^2 = 0 + \sigma^2 = \sigma^2, \\
\mathrm{Var}(y_2) &= \mathrm{Var}(y_1) + \sigma^2 = 2\sigma^2, \\
&\vdots \\
\mathrm{Var}(y_t) &= t\sigma^2,
\end{aligned}$$

meaning the variance increases linearly with time. A plot of 50 simulated random walks is shown in figure 3.2. Here it can be seen that the variance increases with time, and that the different walks are symmetric around 0, which corresponds well with the theory.

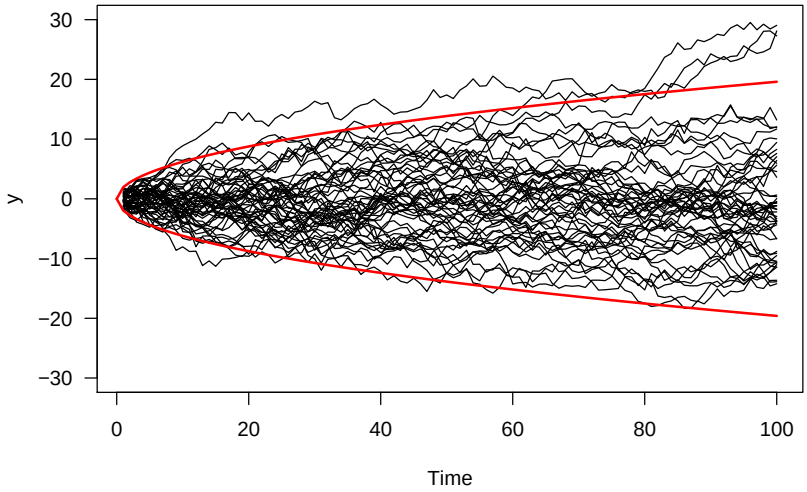


Figure 3.2: 50 simulations of random walks over 100 time units, with red lines showing the 95 % confidence band. Here $\sigma^2 = 1$.

3.1.2 Persistence Model

One of the simplest forecasting models for time series is the persistence model. This model assumes that future values are equal to some values from the past. Say the value of the time series at time t and $t + h$ are related, the persistence model for a stationary time series is [27]

$$y_t = y_{t-h} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2).$$

The simplicity of the model makes it popular as a baseline model, that is a model which other models are compared to in order to evaluate their predictive performance. Another version of the persistence model is obtained when taking the average of a window of previous observations, that is

$$y_t = \frac{1}{p}\sum_{i=1}^{p} y_{t-i} + \varepsilon_t. \tag{3.5}$$

This may be thought of as an $\mathrm{AR}(p)$ process, with $\phi_i = 1/p, i = 1, \ldots, p$, see equation (3.1). The estimate of the mean is then

$$\hat{y}_t = \frac{1}{p}\sum_{i=1}^{p} y_t \tag{3.6}$$

while the variance is estimated from

$$\hat{\sigma}^2 = \frac{1}{p-1}\sum_{i=1}^{p}(y_t - \hat{y}_t)^2. \tag{3.7}$$

The window size p may be determined based on some evaluation method, some of which will be presented in a section 3.2.

3.1.3 Multiple Linear Regression in the Time Series Context

One of the basic prediction models in statistics is the multiple linear regression (MLR) model, which assumes a linear relationship between the variable of interest, the response, and the explanatory variables, or predictors. In the time series context, the response and the explanatory variables are themselves time series. Rather than modelling the response as a single time series, that is relating its current value to its own past, the MLR model also take into account that the variable of interest may be influenced not only by its past values, but also by current and past values of other exogenous time series [26].

Now, denote by $\boldsymbol{y}_t, t = 1, \ldots, N$ the response at time t and $\boldsymbol{c}_t$ the corresponding vector of predictors at time t. With k predictors, $\boldsymbol{c}_t$ is a $k \times 1$ vector. The response may also be a vector, but throughout the remainder of this thesis, assume it it a scalar. The relationship between y_t and $\boldsymbol{c}_t$ may then be written [28]

$$y_t = \boldsymbol{c}_t^\top \boldsymbol{x}_t + v_t \quad v_t \sim \mathcal{N}(0, R_t),$$

where $\boldsymbol{x}_t$ is the $k \times 1$ vector of the coefficients, or weights, of the explanatory variables, and v_t white noise with variance R_t. Now let $\boldsymbol{y}$ and $\boldsymbol{v}$ be the $N \times 1$

vectors with elements y_t and v_t, respectively, and C the matrix with $c_t^\top$ as rows, that is

$$C = \begin{bmatrix} \boldsymbol{c}_1^\top \\ \boldsymbol{c}_2^\top \\ \vdots \\ \boldsymbol{c}_N^\top \end{bmatrix} = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,k} \\ c_{2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ c_{N,1} & \cdots & \cdots & c_{N,k} \end{bmatrix}.$$

With this notation the MLR model may be written more compactly as

$$\boldsymbol{y} = C\boldsymbol{x} + \boldsymbol{v}.$$

The goal is then to find the "best" estimate $\hat{\boldsymbol{x}}$. This amounts to minimising some objective function $f(\boldsymbol{x})$, so that

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}).$$

With the estimate $\hat{\boldsymbol{x}}$ the predicted value of the MLR model is given by $\hat{\boldsymbol{y}} = C\hat{\boldsymbol{x}}$. A standard approach when solving for $\boldsymbol{x}$ is the method of least squares. Then the objective function is

$$f(\boldsymbol{x}) = (\boldsymbol{y} - C\boldsymbol{x})^\top (\boldsymbol{y} - C\boldsymbol{x}) = \boldsymbol{r}^\top \boldsymbol{r},$$

where $\boldsymbol{r}$ is the vector of residuals. In other words, the optimal estimate of $\boldsymbol{x}$ is the one that minimises the square of the residuals. With this method, one obtains the least square estimate

$$\hat{\boldsymbol{x}} = (C^\top C)^{-1} C^\top \boldsymbol{y},$$

while the variance of v_t is estimated by

$$\hat{R} = \frac{1}{N-k}(\boldsymbol{y} - C\hat{\boldsymbol{x}})^\top (\boldsymbol{y} - C\hat{\boldsymbol{x}}) = \frac{1}{N-k}\hat{\boldsymbol{r}}^\top \hat{\boldsymbol{r}}.$$

Both $\hat{\boldsymbol{x}}$ and $\hat{R}$ are unbiased estimators for $\boldsymbol{x}$ and R, respectively, meaning $\mathrm{E}[\hat{\boldsymbol{x}}] = \boldsymbol{x}$ and $\mathrm{E}[\hat{R}] = R$. In addition, the covariance matrix of $\hat{\boldsymbol{x}}$, used for inference, is $\mathrm{Cov}(\hat{\boldsymbol{x}}) = (C^\top C)^{-1}\hat{R}$.

3.1.4 Rolling Window Regression

Parameter instability is considered a crucial issue when forecasting in various fields. To deal with the instability it is common to use only the most recent observations, and not all the available data as in the regular MLR model, to estimate parameters. The observations used makes up the so-called window, which are then used to perform parameter estimation through least squares. Contrary to the MLR

model, which assigns the same weight to the explanatory variables, the rolling window regression (RWR) model allows for changing weights over time. With a window size of $p \leq N$, an illustration of the method is found in figure 3.3. When being at day $t-1$, trying to predict the value for y_t,the weights $\boldsymbol{x}_t$ are fitted from the data in the window, which can be written

$$\boldsymbol{y}_t(p) = \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{bmatrix}, \quad C_t(p) = \begin{bmatrix} \boldsymbol{c}_{t-1}^\top \\ \boldsymbol{c}_{t-2}^\top \\ \vdots \\ \boldsymbol{c}_{t-p}^\top \end{bmatrix} = \begin{bmatrix} c_{t-1,1} & c_{t-1,2} & \cdots & c_{t-1,k} \\ c_{t-2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ c_{t-p,1} & \cdots & \cdots & c_{t-p,k} \end{bmatrix}.$$

With the notation above, the RWR model may be written

$$\boldsymbol{y}_t(p) = C_t(p)\boldsymbol{x}_t(p) + \boldsymbol{v}_t(p), \tag{3.8}$$

so that the least squares estimates of this model are [29]

$$\hat{\boldsymbol{x}}_t(p) = \left[C_t(p)^\top C_t(p)\right]^{-1} C_t(p)^\top \boldsymbol{y}_t(p), \tag{3.9}$$

$$\hat{R}_t(p) = \frac{1}{p-k}\left[\boldsymbol{y}_t(p) - C_t(p)\hat{\boldsymbol{x}}_t(p)\right]^\top \left[\boldsymbol{y}_t(p) - C_t(p)\hat{\boldsymbol{x}}_t(p)\right], \tag{3.10}$$

$$\mathrm{Cov}(\hat{\boldsymbol{x}}_t(p)) = \left[C_t(p)^\top C_t(p)\right]^{-1} \hat{R}_t(p).$$

With the estimated weights $\hat{\boldsymbol{x}}_t(p)$, the electricity price at day t may be calculated as

$$\hat{y}_t = \boldsymbol{c}_t \hat{\boldsymbol{x}}_t. \tag{3.11}$$

The size p of the window may be determined based on some evaluation measure, which will be presented in section 4.2, as mentioned in the previous section.

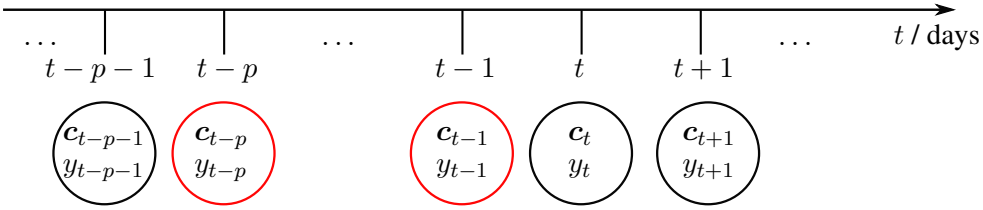


Figure 3.3: Illustration of the rolling window. When estimating $\hat{\boldsymbol{x}}_t(p)$ only the data in the window are used, here emphasised by the red circles.

3.1.5 State-Space Models

Another model that, in the regression sense, allows for changing regression coefficients over time are state-space models, or dynamic linear models. These are general models that are able to represent various systems in only two equations; the *state equation* and *observation equation*. The model was originally introduced as a method for use in aerospace-related research, but has proved to be applicable in various other fields [26]. The techniques used for these models are flexible and capable of handling a much wider range of problems than other time series models [30].

In the state-space representation of a model it is the state of the model that is the variable of interest. The state equation may be written

$$\boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + \boldsymbol{w}_t, \quad t = 1, \ldots, N, \quad \boldsymbol{w}_t \sim \mathcal{N}(\boldsymbol{0}, Q_t), \tag{3.12}$$

where the $k \times 1$ vector $\boldsymbol{x}_t$ defines the current state. This depends linearly on the past state $\boldsymbol{x}_{t-1}$, where the dependency is determined by the $k \times k$ transition matrix A, plus some white noise $\boldsymbol{w}_t$. Here Q_t is the time dependent $k \times k$ covariance matrix of the error. In addition it is assumed that the first state is equal to $\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0)$, with known $\boldsymbol{\mu}_0$ and Σ_0. For $k = 1$, this is simply the $\mathrm{AR}(1)$ process, defined in equation (3.1), where $A = \phi_1$.

However, in state-space models it is assumed that the state $\boldsymbol{x}_t$ is not observed directly, but through a linearly transformed version of it, which is the observation y_t. The observation equation determines this relation, and is

$$y_t = \boldsymbol{c}_t^\top \boldsymbol{x}_t + v_t, \quad v_t \sim \mathcal{N}(0, R_t), \tag{3.13}$$

where $\boldsymbol{c}_t$ is, as before, a $k \times 1$ vector. Recall that y_t was defined to be a scalar for the purpose of this thesis. As can be seen, $\boldsymbol{c}_t$ is the time dependent vector determining the linear relationship between the state $\boldsymbol{x}_t$ and the observation y_t. In addition, v_t is white noise with time dependent variance R_t. It is assumed that $\boldsymbol{w}_t$ and v_t are independent. As $\boldsymbol{x}_t$ is of interest, the goal is to estimate $\boldsymbol{x}_t$ by $y_1, y_2, \ldots, y_t$. Note that when $\boldsymbol{x}_t = \boldsymbol{x}_{t-1} = \boldsymbol{x}\ \forall t$, the state-space model reduces to an MLR model, where the regular least squares estimates may be used. When this is not the case, however, more intricate techniques are required which will be presented in the subsequent section.

3.1.6 Kalman Filter

Once the model is formulated in a state-space form, the Kalman filter may be used to estimate the states $\boldsymbol{x}_t$. The object of filtering is to update our knowledge of the

system as new observations y_t are observed [30]. Let $Y_t = \{y_1, \ldots, y_t\}$, that is the set of all the observations up to time t, and denote $\mathrm{E}[\boldsymbol{x}_t|Y_{t-1}] = \hat{\boldsymbol{x}}_{t|t-1}$ and $\mathrm{Var}(\boldsymbol{x}_t|Y_{t-1}) = P_{t|t-1}$. Then, using (3.12),

$$\begin{aligned}\hat{\boldsymbol{x}}_{t|t-1} &= \mathrm{E}[A\boldsymbol{x}_{t-1} + \boldsymbol{w}_t|Y_{t-1}] = A\hat{\boldsymbol{x}}_{t-1|t-1},\\ P_{t|t-1} &= \mathrm{Var}(A\boldsymbol{x}_{t-1} + \boldsymbol{w}_t|Y_{t-1}) = AP_{t-1|t-1}A^\top + Q_t.\end{aligned}$$

Further, the error in the estimation, called the innovation residual, can be found by

$$e_t = y_t - \mathrm{E}[y_t|Y_{t-1}] = y_t - \mathrm{E}[\boldsymbol{c}_t^\top \boldsymbol{x}_t + v_t|Y_{t-1}] = y_t - \boldsymbol{c}_t^\top \hat{\boldsymbol{x}}_{t|t-1},$$

with associated innovation covariance

$$S_{t|t-1} = R_t + \boldsymbol{c}_t^\top P_{t|t-1}\boldsymbol{c}_t.$$

Now, having observed y_t, the updated state estimate can be calculated by

$$\hat{\boldsymbol{x}}_{t|t} = \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{K}_t e_t,$$

where $\boldsymbol{K}_t = P_{t|t-1}\boldsymbol{c}_t S_{t|t-1}^{-1}$ is the Kalman gain. The updated covariance is

$$P_{t|t} = (1 - \boldsymbol{K}_t\boldsymbol{c}_t^\top)P_{t|t-1}.$$

To summarise, the Kalman filter is an algorithm which iteratively estimates the system's state and updates it as new observations are available. These steps are done through the *prediction step* and the *updating* or *filtering* step. Before these steps the system has to be initialised with some initial conditions,

$$\begin{aligned}\hat{\boldsymbol{x}}_{0|0} &= \boldsymbol{\mu}_0,\\ P_{0|0} &= \Sigma_0.\end{aligned} \tag{3.14}$$

The prediction step is then

$$\begin{aligned}\hat{\boldsymbol{x}}_{t|t-1} &= A\hat{\boldsymbol{x}}_{t-1|t-1},\\ P_{t|t-1} &= AP_{t-1|t-1}A^\top + Q_t,\\ S_{t|t-1} &= R_t + \boldsymbol{c}_t^\top P_{t|t-1}\boldsymbol{c}_t,\end{aligned} \tag{3.15}$$

while the updating step consists of

$$\begin{aligned}\boldsymbol{K}_t &= P_{t|t-1}\boldsymbol{c}_t S_{t|t-1}^{-1},\\ \hat{\boldsymbol{x}}_{t|t} &= \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{K}_t e_t,\\ P_{t|t} &= (1 - \boldsymbol{K}_t\boldsymbol{c}_t^\top)P_{t|t-1}.\end{aligned}$$

3.2 Evaluation Measures

In this section, the evaluation measures used to evaluate the predictive performance of the models in the thesis will be presented. These include the mean square error, the continuous ranked probability score and the mean absolute error, which will be presented in the following.

3.2.1 Mean Square Error

The mean square error (MSE) is a common measure to evaluate the prediction accuracy of a certain model. Denote by y_t the observation at time t, $t = 1, \ldots, N$, and $\hat{y}_t$ its point prediction. Then the MSE is defined as

$$\text{MSE} = \frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2 \tag{3.16}$$

As can be seen, a model yielding perfect predictions correspond to an MSE of 0, while a higher MSE indicates a poorer prediction [31].

3.2.2 Continuous Ranked Probability Score

The continuous ranked probability score (CRPS) is a measure for evaluating probabilistic forecasts. Contrary to point forecasts, where a specific value is given, probabilistic forecasting assigns a probability distribution to the forecast. Let F_t be the probabilistic cumulative distribution function (cdf) at time t of the forecast with the corresponding observation y_t. Then its CRPS is defined as [32]

$$\text{CRPS}_t = \int_{-\infty}^{\infty}(F_t(x) - \mathbb{1}\{x \geq y_t\})^2 dx,$$

where $\mathbb{1}\{x \geq y_t\}$ is the *indicator function*, defined as

$$\mathbb{1}\{x \geq y_t\} = \begin{cases} 0 & x < y_t, \\ 1 & x \geq y_t. \end{cases}$$

For several probabilistic forecasts, each having a cdf $F_t(y_t), t = 1, \ldots, N$, the CRPS becomes

$$\text{CRPS} = \frac{\sum_{t=1}^{N}\text{CRPS}_t}{N}. \tag{3.17}$$

The CRPS measures the squared difference between the cdf of the forecast and that of the observation y_t, which takes the form of the indicator function defined above. The difference of the cdf of a forecast and an observation is illustrated in figure 3.4

by the grey area. In this particular case the forecasted distribution is standard normal, shown in red, while the cdf of the observation is shown in black. Here the observation is at 1.2. In the same figure the cdf of a $\mathcal{N}(1.2, 0.2^2)$ distribution is presented in blue. As can be seen, the CRPS depend on the forecasted distribution, and is lower for the cdf of $\mathcal{N}(1.2, 0.2^2)$. However, this requires a high precision and low variance. As shown in [32], in the particular case of a normal distribution with mean μ and variance σ^2, the CRPS at time t takes the form

$$\mathrm{CRPS}_{\mathcal{N},t} = -\sigma \left[\frac{1}{\sqrt{\pi}} - 2\varphi\left(\frac{y_t - \mu}{\sigma}\right) - \frac{y_t - \mu}{\sigma}\left(2\Phi\left(\frac{y_t - \mu}{\sigma}\right) - 1\right)\right], \tag{3.18}$$

where $\varphi(\cdot)$ denote the probability density function and $\Phi(\cdot)$ the cumulative distribution function of the normal distribution.

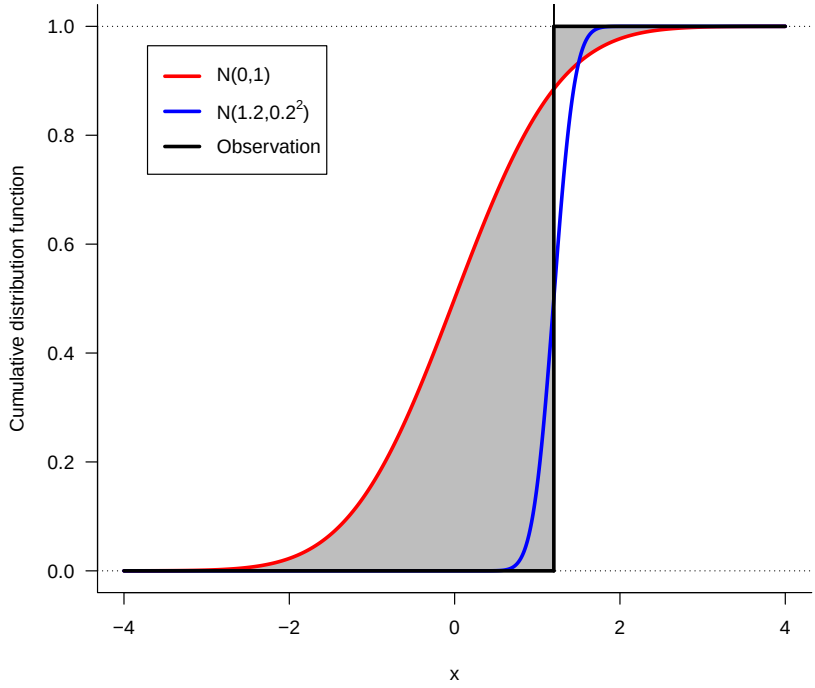


Figure 3.4: Illustration of the CRPS. The plot shows the cdf of a standard normal distribution (red curve), the cdf of an observation (black curve), and the cdf of a $\mathcal{N}(1.2, 0.2^2)$ distribution (blue curve). The observation is at at 1.2. Assuming the forecasted distribution is a standard normal, the CRPS is the square of the grey area. As can be seen, for this particular observation, the $\mathcal{N}(1.2, 2^2)$ distribution yields lower CRPS.

3.2.3 Mean Absolute Error

The last evaluation measure considered is the mean absolute error (MAE). This is defined as

$$\mathrm{MAE} = \frac{\sum_{t=1}^{N} |y_t - \hat{y}_t|}{N}, \tag{3.19}$$

where $t = 1, \ldots, N$, as before. Note that equation (3.17) reduces to the mean absolute error for point forecasts [32]. As seen, this measure is rather similar to the MSE, however, as pointed out in in [33] the MAE is "a more natural measure of average error". This is due to the fact that the MSE penalises large errors, as this measure square the errors. Thus, with outliers present, yielding large prediction errors, the MSE may underestimate a model's performance.

Chapter 4

Forecasting the Electricity Spot Prices at the European Energy Exchange

In this chapter three models for forecasting the day-ahead electricity spot price on the European Energy Exchange are studied. The electricity market, as well as the data provided for this project, have already been described in chapter 2. The models that have been developed and studied in this thesis are the persistence, RWR and the state-space model. The former model is a pure statistical model, more specifically a time series model, as it relates the electricity prices to its own past. The other models also take into account that the electricity prices may be influenced by other time series as well as its own historic values, and are thus classified as a hybrid between the statistical and fundamental models. Further, the persistence model serves as a baseline model which the other models are compared with to inspect their predictive performance. All the models are tested with data from 2010, and the hourly intervals 03:00-04:00, 12:00-13:00, 18:00-19:00. Henceforth, these intervals are referred to as hour 03:00, 12:00 and 18:00, respectively. A plot of the spot price development for these hours are shown in figure 4.1. Note that negative prices only occur in hour 03.00, when the demand is at its lowest. This supports the statement that these prices mostly occur during the night, as mentioned in chapter 2. On the other hand, price spikes in the positive direction mostly occur during the evening, at 18:00, where the demand reaches its evening peak. Despite the fact that the demand is highest at noon, there are less price spikes than for the other hours.

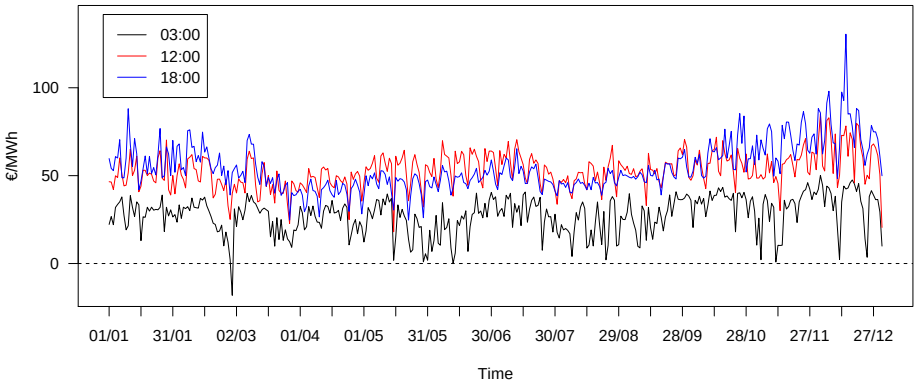


Figure 4.1: The spot price development for hour 03:00 (shown in black), 12:00 (shown in red) and 18:00 (shown in blue) for various dates in 2010. The dates are presented in the day/month format. A horizontal dotted line is included to indicate a spot price of 0.

In the following the spot price at day t and hour i is denoted $y_{i,t}$, while $\hat{y}_{i,t}$ denote its prediction. As mentioned, the RWR and the state-space model use past values of other time series, in addition to the spot price, to forecast the day-ahead electricity price. These explanatory variables are based on the data in 2.1, and are contained in the vector defined as

$$\boldsymbol{c}_{i,t} = \begin{bmatrix} \text{Spot price}_{i,t-1} \\ \text{Volatility}_{i,t-1} \\ \text{Coal price}_{t-1} \\ \text{Gas price}_{t-1} \\ \text{Oil price}_{t-1} \\ \text{EUA price}_{t-1} \\ \text{Expected PV and wind}_{i,t} \\ \text{ExPPA}_{i,t} \\ \text{Demand}_{i,t} \\ \text{Demand}_{i,t-1} \end{bmatrix}.$$

In the following, $y_{i,t}$ and $\boldsymbol{c}_{i,t}$ replace y_t and $\boldsymbol{c}_t$ as presented in chapter 3. As seen, the number of explanatory variables is $k = 10$. Note that for the vector $\boldsymbol{c}_{i,t}$ the subscript i is omitted for some of the entries. This is due to the fact that these are given on an hourly basis, as seen in table 2.1.

To understand why the entries of $\boldsymbol{c}_{i,t}$ are chosen, recall the price formation of the spot contract discussed in chapter 2. In this chapter, the inclusion of the prices

of coal, gas, oil and CO_2 emission allowances are already argued for, as well as the expected infeed from photovoltaic and wind and the demand. For the lag spot price, it is argued in [11] that this is included to reduce autocorrelation in the data and incorporate historic price and risk signals, which usually influence agents' price expectations and risk aversion. In addition, as noted in [17] and [18], the $\mathrm{AR}(1)$ process is a benchmark model for mean reversion, which supports having the lag spot price incorporated in the model. Lastly, as for the volatility of the spot price, this is included to serve as another indicator for historic price instability and risk.

In this chapter the models will be presented in separate sections, starting with the persistence model, then the RWR model and lastly the state-space model. The different sections will also assess the models, both in terms of their predictive performance as well as assessing the model assumptions of the residuals. The quality of the predictions are measured through the MSE, CRPS and MAE, as presented in section 3.2. As for the CRPS, this requires a predictive distribution, which for all the models is assumed to be normal, meaning equation (3.18) may be used to calculate it.

4.1 Persistence Model

The simplest model that is studied for predicting the day-ahead electricity spot prices is the persistence based model given in equation (3.5). In this model the electricity price at i, t is modelled as an average of the prices of the same hour from the p previous days, thus generating 24 time series for each hour i. The window size p is chosen to minimise the MSE, CRPS and MAE, which takes the form of (3.16), (3.18) and (3.19), respectively. These are calculated using the estimates of the mean and the variance, which are estimated as in the equations (3.6) and (3.7).

Plots of the MSE and CRPS for different window sizes p and the hours 03:00, 12:00 and 18:00, may be found in figure 4.2. The optimal window sizes are either 2, 7 or 8, depending on the hour and which error measure considered, as summarised in the tables 4.1a and 4.1b. The window size is chosen to be 7 even though this is not the optimal window for all hours for the CRPS. However, as mentioned earlier, the spot prices express, among other things, weekly seasonality, which makes a window size of 7 plausible. With this window size, the MSE, CRPS and MAE for the hour 12:00, is 77.64, 4.09 and 6.79, respectively, while these values for the other hours are found in table 4.1a. Corresponding tables as in 4.1, as well as similar plots to the ones found in figure 4.2, using MAE as the error

measure may be found in table A.1 and figure A.1 in the appendix.

Further, it is seen that the persistence model predicts best for the hour 18:00 when $p = 7$, as the MSE, CRPS and MAE are lowest for this hour. The predictions for the hour 12:00 is almost as good as for the hour 18:00, with an MSE, CRPS and MAE of only 1.97 %, 0.39 % and 8.30 % higher than the best prediction, using a window size of $p = 7$. As for the hour 03:00, on the other hand, the MSE, CRPS and MSE are 20.50 %, 15.12 % and 14.83 % higher than the predictions for hour 18:00. This may be due to the fact that the prices during nighttime are more volatile than during the day, as mentioned in chapter 2. As seen from the figure 4.1, it can also be seen that the variability in the spot prices at 03:00 is greater than for the other hours, causing the average of the p last days to be a bad prediction for this particular hour.

In the following, predictions for the hours 12:00 are considered. Even though the persistence model performs better on the hour 12:00 with respect to the error measures considered, the other models which are presented later in this chapter predict the price at noon better. In addition, the values of both MSE, CRPS and MAE for the predictions at 12:00 are rather similar to those of hour 18:00. Now, the spot price predictions for $p = 7$, as well as the actual spot price, can be found in figure 4.3 in red and black, respectively. A 95 % confidence interval for the prediction is also shown in said figure. As can be seen, the curve of the predicted electricity prices seem to smooth the curve of the true prices, and fail to capture the jaggedness of the true spot price development. This is as expected, as this model takes the average of the 7 last electricity prices. Despite this, the true spot price is within the 95 % confidence interval for all the days considered.

Table 4.1: The optimal p for the persistence model with respect to MSE and CRPS, for all the hours. The MAE is also given.

(a) The p yielding lowest MSE.

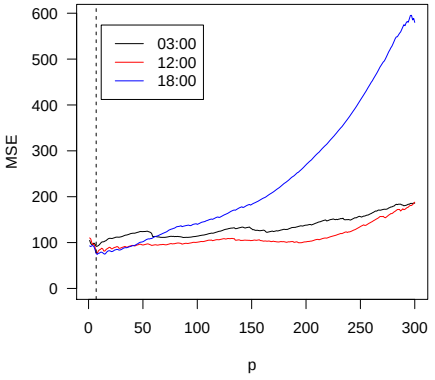
Hour	p	MSE	CRPS	MAE
03:00	7	91.75	4.34	7.20
12:00	8	76.44	4.09	6.79
18:00	7	76.14	3.77	6.27

(b) The p yielding lowest CRPS.

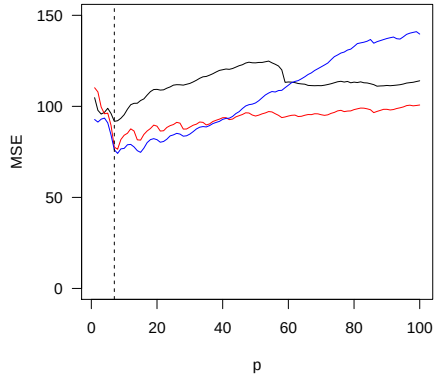
Hour	p	MSE	CRPS	MAE
03:00	2	97.82	4.25	7.06
12:00	8	76.44	4.10	6.79
18:00	7	76.14	3.77	6.27

To assess the persistence model further, various diagnostic plots are plotted and

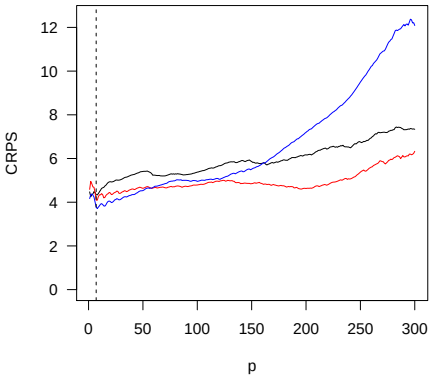
Figure 4.2: The MSE and CRPS for the persistence model for various window sizes p and the hours 03:00 (black curve), 12:00 (red curve) and 18:00 (blue curve). A vertical dotted line at $p = 7$ is also shown in all the plots.



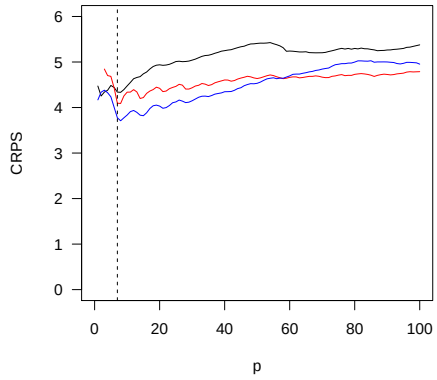
(a) The MSE for window sizes $p \leq 300$.



(b) The MSE for window sizes $p \leq 100$



(c) The CRPS for window sizes $p \leq 300$.



(d) The CRPS for window sizes $p \leq 100$.

presented in figure 4.4. As seen in figure 4.4a, the histogram of the model is shown, with a normal density curve in red. The histogram shows that there are more negative residuals than positive, so the histogram is not perfectly symmetrical. As for the plot of the predicted versus the actual spot price, $\hat{y}_{12,t}$ versus $y_{12,t}$, shown in figure 4.4b, the predictions are not very off in the sense that it seems to be a linear relationship between them. In fact, the regression slope is 0.997, with 1 being the ideal slope. The standard error of the slope, indicating the spread of the points, is $8.75 \cdot 10^{-3}$. Moving on to figure 4.4c, it is seen that the residuals express a slight linear relationship to the fitted values, having a regression slope of $-3.07 \cdot 10^{-4}$, as illustrated by the red line. However, when conducting a significance test for the linear relationship, the null hypothesis of a regression slope of 0 is not rejected

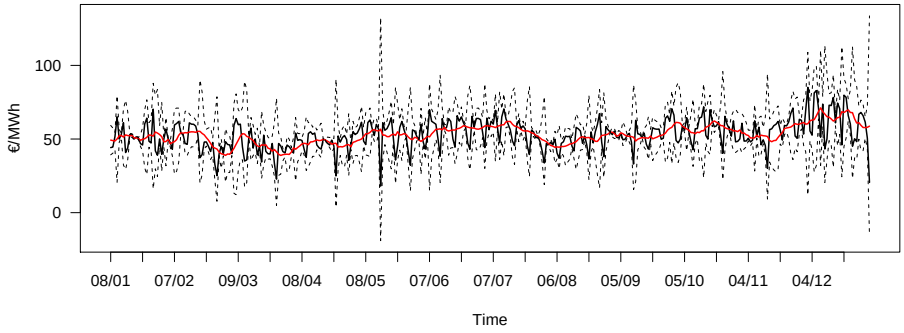
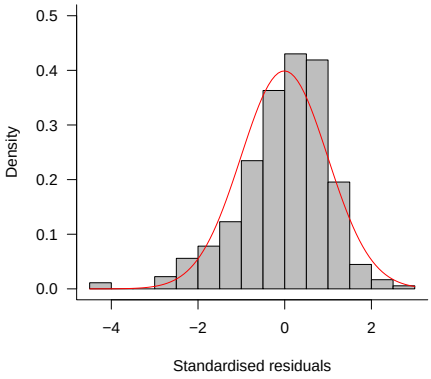


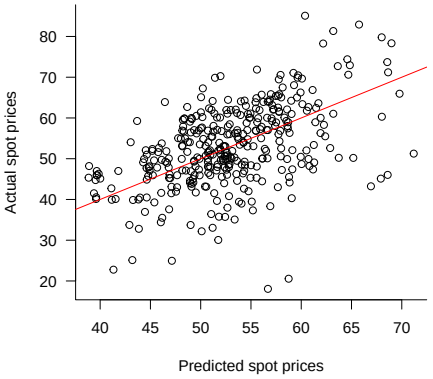
Figure 4.3: The spot price predictions, $\hat{y}_{i,t}$, from the persistence model with $p = 7$ (red curve) and the actual spot prices, $y_{i,t}$, for the year 2010 (black curve). Here $i = 12$. A $95\ \%$ confidence interval is included and shown as dotted lines. The axes correspond to those of figure 4.1. Note that the start date is the 8th of January as the first $p = 7$ days are used to calculate the first prediction.

on neither a $5\ \%$ nor $10\ \%$ significance level. Thus it may be concluded that the linear relationship is not significant. Moreover, figure 4.4d shows a QQ plot with a red reference line, as well as a 95% confidence interval for the normal distribution. The slight concave relationship indicates a heavy left tail and a light right tale, even though the right tail is within the $95\ \%$ confidence interval, as seen from the QQ plot. This gives evidence that the data are left-skewed. This is also supported by figure 4.4a. Thus, other distributions meeting this criteria should be analysed further. In fact, when conducting a Shapiro-Wilks test [34] for testing for normality, there is adequate evidence that the null hypothesis that the residuals come from a normally distributed population be rejected on a $5\ \%$ significance level.

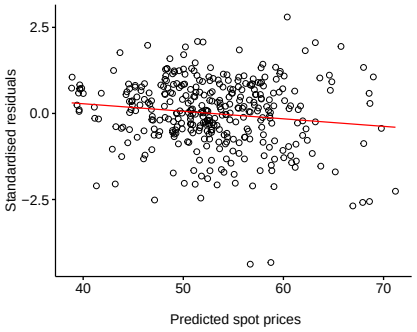
Figure 4.4: Different diagnostic plots to evaluate the persistence model with a window size of $p = 7$.



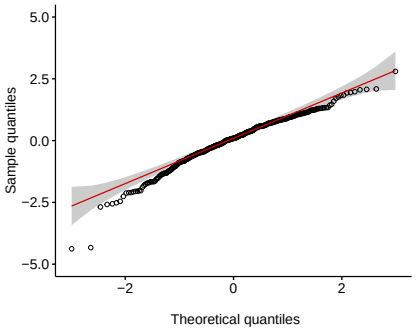
(a) The histogram of the standardised residuals with the normal density curve in red.



(b) Scatter plot of the predicted values of the spot prices, $\hat{y}_{12,t}$, versus the actual spot prices, $y_{12,t}$. The red diagonal line indicates a perfect fit. Here the regression slope between $\hat{y}_{12,t}$ and $y_{12,t}$ is found to be 0.997.



(c) Scatter plot of the predicted values, $\hat{y}$, versus the standardised residuals. The red line shows the regression line, with a slope of $-3.07 \cdot 10^{-4}$.



(d) The normal QQ plot with a 95 % confidence interval indicated by the grey area.

4.2 RWR Model

A drawback with the persistence model is that it only uses historical values of the spot price itself, and do not take into account that it may be influenced by other variables. In the RWR model, on the other hand, the day-ahead electricity price may also be influenced by other underlying time series. This model takes the form of equation (3.8), with the subscript t replaced by i, t. The window size p was, as for the persistence model, chosen to minimise the MSE, CRPS and MAE. These were calculated using the estimated mean and variance, estimated by (3.9) and (3.10), respectively.

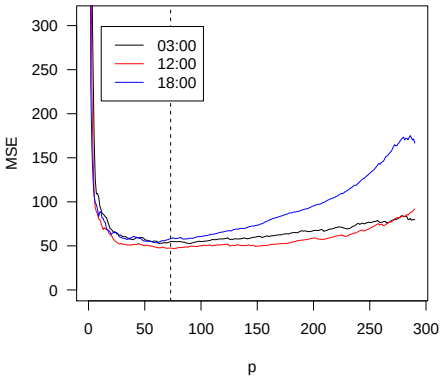
Plots of the MSE and CRPS for different window sizes p and the hours 03:00, 12:00 and 18:00, may be found in figure 4.5. As for the persistence model, there are different optimal window sizes depending on the hour, summarised in table 4.2. However, the MSE and CRPS for hour 18:00 increases for p larger than around 70, while staying approximately the same for the hours 03:00 and 12:00. Therefore, to keep both the MSE and CRPS small for all hours, $p = 73$ was chosen. For this window size, the MSE, CRPS and MAE for the hour 12:00 are 47.70, 3.74 and 5.12, while for 18:00 they are 54.46, 3.83 and 5.22. As in the case for the persistence model, corresponding tables as in 4.2, as well as similar plots to the ones found in figure 4.5, using MAE as the error measure, may be found in table A.1 and figure A.1 in the appendix.

As seen from table 4.2, the RWR model predicts the day-ahead spot prices at noon better, even with the non-optimal window size of $p = 73$. Contrary to the persistence model, the RWR model's predictions are better for the hour 03:00 than 18:00 with respect to MSE. However, with respect to CRPS and MAE, the predictions for hour 18:00 are better. To understand this, recall that the MSE penalises large errors, and favours small errors, as this measure squares the errors. This may lead to either underestimating or overestimating the model's predictive performance. Keeping this in mind, it might be the case that the some of the predictions for the hour 18:00 yield larger errors, causing its MSE to be larger than that of the hour 03:00. Alternatively, some of the predictions for hour 03:00 may be close to the actual value, leading to overestimating the performance of the prediction for this hour. Nevertheless, all the error measures are lower than the corresponding values for the persistence model, meaning the predictive performance of the RWR model is better compared to that of the persistence model, considering these error measures.

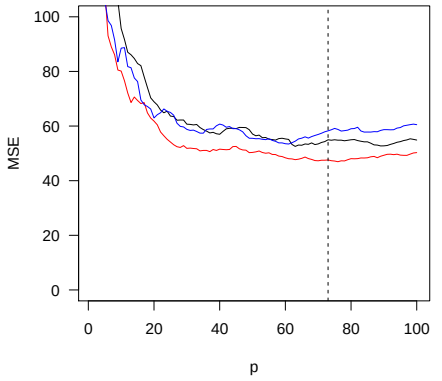
The spot price predictions from the RWR model with $p = 73$, as well as the actual

spot price, for hour 12:00 can be found in figure 4.6 in red and black, respectively. In this figure a 95 % confidence interval is also included and shown as dashed lines. Unlike the persistence model, the RWR model seems to capture the stochastic nature of the spot prices to a greater extent. However, whereas the 95 % confidence interval for the persistence model contain 100 % of the true values, the corresponding percentage is 89 % for the RWR model. Thus, by taking a more aggressive approach, which captures the stochastic behaviour of the electricity spot prices, some of the actual spot prices consequently lie outside the confidence interval.

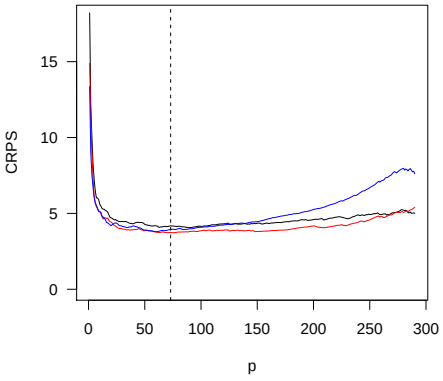
Figure 4.5: The MSE and CRPS for the RWR model for various window sizes p and the hours 03:00 (black curve), 12:00 (red curve) and 18:00 (blue curve). A vertical line at $p = 73$ is also shown in all the plots.



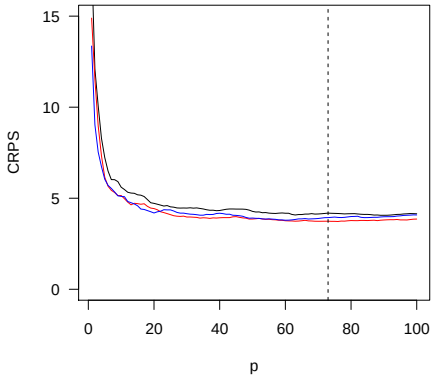
(a) The MSE for window sizes $p \leq 300$.



(b) The MSE for window sizes $p \leq 100$



(c) The CRPS for window sizes $p \leq 300$.



(d) The CRPS for window sizes $p \leq 100$

As for the previous model, various diagnostic plots are plotted to evaluate the

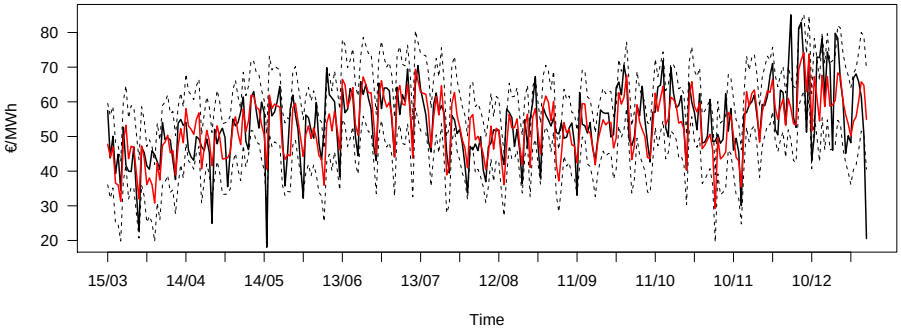


Figure 4.6: The spot price predictions from the RWR model with $p = 73$ (red curve) and the actual spot price for the year 2010 (black curve). A $95\ \%$ confidence interval is included and shown as dotted lines. The axes correspond to that of figure 4.1. Note that the start date is the 15th of March as the first $p = 73$ days are used to calculate the first prediction.

model further. These can be found in figure 4.7. Firstly, as seen in figure 4.7a, the residuals still express a heavier left tail, however in this case there are extreme values in both the left and the right tail. Due to this, a heavy-tailed distribution is worth investigating further. Secondly, as figure 4.7b suggests, the relationship between the predicted and actual spot price seem to be linear, with a regression slope of 1.009 and a standard error of $7.52 \cdot 10^{-3}$. The fact that the points are closer to the red reference line than in figure 4.4b, as the standard error of this particular slope is less than that of the persistence model, this is another indicator that the RWR model performs better in terms of predictions than the baseline model. Thirdly, the residual scatter plot in figure 4.7c indicates a slight linear relationship between the residuals and the predicted spot prices. Nonetheless, as in the case of the persistence model, the null hypothesis of a non-significant regression coefficient cannot be rejected on neither significance level considered. Lastly, the QQ plot in figure 4.7d, as well as figure 4.7a, suggests a heavy right and left tail. Similarly to the persistence model, a Shapiro-Wilks test rejects normality of the residuals.

Furthermore, plots of the development of the weights for the coal price, gas price, photovoltaic and wind and the expected demand may be found in figure 4.8. These are shown for hour 03:00, 12:00 and 18:00. As for the weight of coal, this is negative for the most part for all hours. This would mean that an increase in the coal price tend to decrease the electricity spot price, when keeping all the other variables constant. This is rather unintuitive, as the coal price is positively correlated to the spot price, as seen in the correlation matrix in figure 2.5. However, from

Table 4.2: The optimal p for the RWR model with respect to MSE and CRPS, for all the hours. The MAE is also given.

(a) The p yielding lowest MSE.

Hour	p	MSE	CRPS	MAE
03:00	73	52.55	4.08	5.70
12:00	86	46.95	3.72	5.11
18:00	71	53.43	3.80	5.16

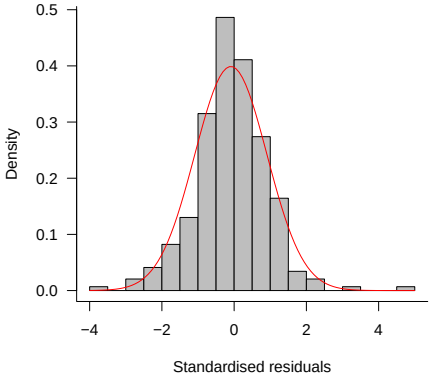
(b) The p yielding lowest CRPS.

Hour	p	MSE	CRPS	MAE
03:00	99	52.72	4.06	5.71
12:00	86	46.95	3.72	5.11
18:00	70	53.53	3.79	5.14

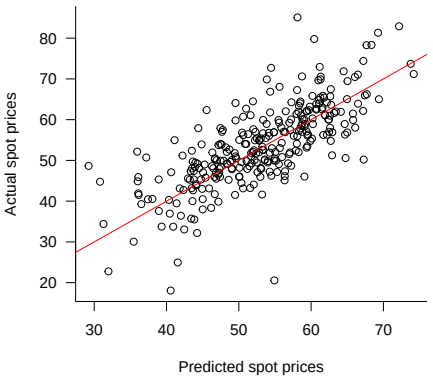
the correlation matrix it is also seen that the coal price is strongly correlated to gas, oil and the emission allowance price, with the correlation coefficients of 0.45, 0.64 and 0.58, respectively. Thus, this may give evidence that the coal price can be linearly predicted by these, and that there exists *multicollinearity* in the data. This is a phenomenon in which there are near-linear dependencies between the explanatory variables [35]. As pointed out in this book, the determinant of the correlation matrix may also be an indicator for multicollinearlity; if the determinant is 1, the explanatory variables are independent, whereas when it is 0 there exists an exact linear relationship between the variables. In this case, the determinant of the correlation matrix is $8.11 \cdot 10^{-3}$. Further, the issue of multicollinearity leads to problems with the least squares estimates, hence also causing problems when doing inference. In fact, in the case of strong multicollinearity, the least square estimates of the weights and their variances tend to be large. Especially the latter gives problems when doing inference on the weights, and is the reason why a 95 % confidence interval is not seen in the plots in figure 4.8, as the intervals are too large. Despite these issues, multicollinearity do not reduce the predictive power of the model itself, hence the MSE, CRPS and MAE are rather low. However, this would explain the unintuitive nature of the plot in figure 4.8a, namely the unstability of the weight and the fact that it even changes sign.

To deal with multicollinearity, [35] proposes three methods, the first being to collect additional data, the second to respecify the model, and the third being to perform Ridge regression. As the model already contains $k = 10$ explanatory variables, the first option does not seem to be necessary, and it may also be the case that obtaining additional data is costly. The two other options seem plausible. As mentioned, however, the presence of multicollinearity does not affect the model fit. Thus, for the purpose of this thesis, the multicollinearity is merely addressed, and makes an interesting topic to resolve for further work.

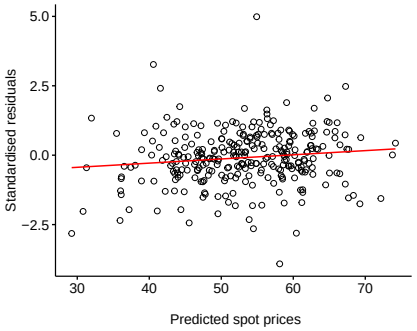
Figure 4.7: Different diagnostic plots to evaluate the RWR model with a window size of $p = 73$



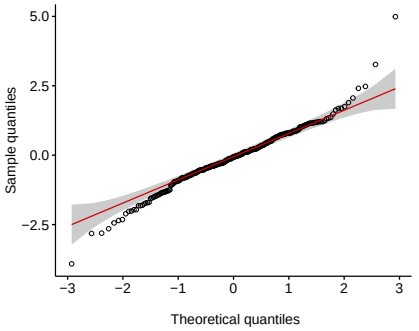
(a) The histogram of the standardised residuals with the normal density curve in red.



(b) Scatter plot of the predicted values of the spot prices, $\hat{y}_{12,t}$, versus the actual spot prices, $y_{12,t}$. The red diagonal line indicates a perfect fit. Here the regression slope between $\hat{y}_{12,t}$ and $y_{12,t}$ is found to be 1.009.



(c) Scatter plot of the predicted values, $\hat{y}$, versus the standardised residuals. The red line shows the regression line, with a slope of $7.52 \cdot 10^{-3}$.

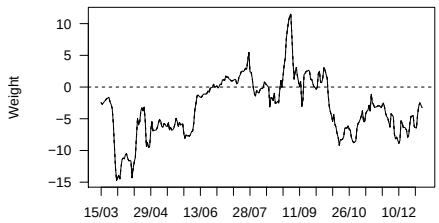


(d) The normal QQ plot with a 95 % confidence interval indicated by the grey area.

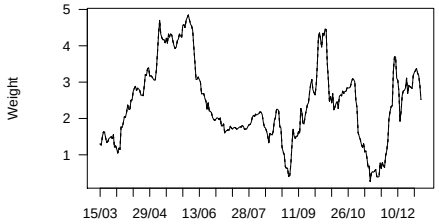
As for the other plots in the figure 4.8, the signs are in accordance with the correlation matrix. From the plot of the weight of gas, seen in figure 4.8b, it is seen that the gas price has less influence on the electricity spot price during the summer, when the value of the weight is around 2. This makes sense as the electricity demand reaches its minimum during the summer months, thus extra electricity generated by the power plants situated to the right of the merit order curve, illus-

trated in figure 2.2, is typically not needed. Further, the weight of photovoltaic and wind in figure 4.8c is negative for the whole period considered, meaning an increase in the electricity generated by these sources decreases the electricity spot price. In addition to correspond to the correlation matrix, this is expected from the merit order effect, discussed in section 2.1. As seen the effect of this variable is less negative around August, meaning an increase in the infeed from these sources do not have such a negative impact on the electricity prices than during the first half of the year. This also makes sense, as the demand is low, and to a greater extent than in high-demand months, met with the electricity generated by the base load. Lastly, as for the weight of the demand, this is highest during the summer months. As the demand is lowest during these months, it is plausible that its weight is higher so as to keep the electricity prices at a reasonable level. Note that although it appears that the the weight of the photovoltaic and wind infeed and expected demand are small compared to those of coal and oil, recall that these variables are of magnitude around 10^3 and 10^4, respectively, as seen in table 2.2. Coal and gas are, in comparison, of magnitude 1 and 10.

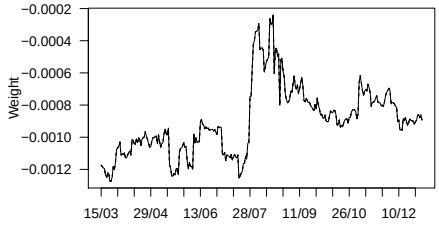
Figure 4.8: Different plots of the development of the weights, estimated by equation (3.9), of the RWR model with a window size of $p = 73$ and the hour 12:00. The dates are displayed on the first axis in the day/month format.



(a) The weight of coal price.



(b) The weight of gas price.



(c) The weight of photovoltaic and wind.

(d) The weight of the expected demand.

4.3 State-Space Model

In the last approach a state space model for the electricity spot prices is formulated as in equation (3.12) and (3.13), based on [11]. As before, y_t and $\boldsymbol{c}_t$ are replaced by $y_{i,t}$ and $\boldsymbol{c}_{i,t}$, but in addition, the transition matrix in the state equation 3.12 is also dependent of the hour i. For the sake of ease of reading, the state-space model is restated here,

$$\begin{aligned} y_{i,t} &= c_{i,t}^{\top}\boldsymbol{x}_{i,t} + v_{i,t}, \quad v_{i,t} \sim \mathcal{N}(0, R_t) \\ \boldsymbol{x}_{i,t} &= A_i\boldsymbol{x}_{i,t-1} + \boldsymbol{w}_{i,t}, \quad \boldsymbol{w}_{i,t} \sim \mathcal{N}(\boldsymbol{0}, Q_t). \end{aligned}$$

As before, $t = 1, \ldots, N$. Two different values of the transition matrix A_i are tested, the first case is when A_i is equal to the identity matrix I for all i, as in said article. In the other case the transition matrix is a diagonal matrix with its entries estimated by the weights from the RWR model for some p. In this case, p was chosen to be 73. The cases are denoted case 1 and case 2, respectively. More specifically, for $j = 1, \ldots, k$ and $i = 1, \ldots, 24$, let $\hat{\boldsymbol{x}}^{i,j}_{[t_0:t_1]}$ denote the vector of the weights of the jth explanatory variable for hour i in the time period $t \in [t_0, t_1]$, as estimated by the RWR model. In addition, let $\hat{\boldsymbol{y}}^{i}_{[t_0,t_1]}$ be the vector containing the predicted electricity spot prices for hour i in the same period, and $\boldsymbol{y}^{i}_{[t_0,t_1]}$ the corresponding vector of the actual spot prices. That is

$$\hat{\boldsymbol{x}}^{i,j}_{[t_0:t_1]} = \left[\hat{x}^{i,j}_{t_0}, \hat{x}^{i,j}_{t_0+1}, \ldots, \hat{x}^{i,j}_{t_1-1}, \hat{x}^{i,j}_{t_1}\right]^{\top},$$

$$\hat{\boldsymbol{y}}^{i}_{[t_0:t_1]} = \left[\hat{y}^{i}_{t_0}, \hat{y}^{i}_{t_0+1}, \ldots, \hat{y}^{i}_{t_1-1}, \hat{y}^{i}_{t_1}\right]^{\top},$$

while $\boldsymbol{y}^{i}_{[t_0,t_1]}$ is similar to the latter, but with the circumflex omitted. Then, the entries of the state-space models are calculated as

$$\hat{A}^{i}_{j,j} = \operatorname{corr}\left(\hat{\boldsymbol{x}}^{i,j}_{[2:(N-p)]}, \hat{\boldsymbol{x}}^{i,j}_{[1:(N-p-1)]}\right), \tag{4.1}$$

$$\hat{Q}_{i,j} = \operatorname{Var}\left(\hat{\boldsymbol{x}}^{i,j}_{[2:(N-p)]} - A^{i}_{j,j}\hat{\boldsymbol{x}}^{i,j}_{[1:(N-p-1)]}\right), \tag{4.2}$$

$$\hat{R}_i = \operatorname{Var}\left(\boldsymbol{y}^{i}_{[1:(N-p)]} - \hat{\boldsymbol{y}}^{i}_{[1:(N-p)]}\right), \tag{4.3}$$

where N is the number of days considered and p is the size of the window used by the RWR. Recall that with a window size of p, $N - p$ estimates are obtained for each hour i of the coefficients $\hat{x}_{i,t}$ and the predicted spot prices, $\hat{y}_{i,t}$. Observe that the above estimates are independent of the day t, and only depend on the hour i and on the variable j considered. This assumption is also made in the article this model is based on. Here $\hat{A}^{i}_{j,j}$ and $\hat{Q}_{i,j}$ are the jth diagonal entry for the transition

matrix and state covariance matrix, respectively, for hour i, and $\hat{R}_i$ the variance of the observation. The Kalman filter, presented in section 3.1.6, is then used to solve the state-space model. In the following, the results from the Kalman filter, using the different transition matrices, will be presented, starting with case 1. In both cases $\hat{Q}_{i,j}$ is estimated as in (4.2), while the initialisation step of the filter, see (3.14), takes the form

$$\mu_0^{i,j} = \frac{1}{N-p} \sum_{t=1}^{N-p} \hat{x}_t^{i,j},$$

$$\Sigma_{0,j,j}^{i} = \frac{\hat{Q}_{i,j}}{1 - \hat{A}_{j,j}^{i2}},$$

Here $\mu_0^{i,j}$ is the jth entry of the $k \times 1$ vector μ_0 for hour i, and is estimated from the weights from the least squares solution of the RWR model. $\Sigma_{0,j,j}^{i}$ is the jth entry of the diagonal matrix, Σ_0 for the same hour, calculated by using equation (3.3). In case 1, with $A = I$, the state equation (3.12) is non-stationary, and in fact a k dimensional random walk, as described in section 3.1.1. As seen here, as well as in figure 3.2, the variances of the states increase linearly with time. In case 2, with A estimated as in (4.1), however, the state equation is stationary, meaning the variances for the states $\boldsymbol{x}_t$ are the same for all t and can be determined by equation (3.3).

To evaluate the predictive performance for the model in both cases, the MSE, CRPS and MAE for these and the hours 03:00, 12:00 and 18:00, can be found in table 4.3. From these tables it is seen that the model with $A = I$ performs better with respect to the evaluation methods and for all the hours considered. Hence, this model is evaluated further, for the hour 12:00. The spot price prediction, as well as the actual spot price, for 2010 are shown in figure 4.9, including a 95 % confidence interval for the prediction. The predictions seem to follow the actual spot prices rather well, and improves the prediction from the persistence model with respect to the error measures considered. However, it fails to outperform the predictions from the RWR model considering the MSE and CRPS for all the hours. The Kalman filter solution yields a smaller MAE for the hour 03:00 than for the RWR model, however this measure is higher for the other hours. Based on these measures, the overall performance of the RWR model is better. This may in turn lead to the conclusion that the electricity spot prices have a larger "memory" than one day, as the state-space model only considers the data in $\boldsymbol{c}_t$, whereas the RWR model considers data from a window of p days, that is $C_t(p)$.

Also here the quality of the prediction for hour 03:00 is the worst, considering the

CRPS and MAE. This is a recurring problem throughout all the models considered, and indicates that the models fail to capture the price jumps that occur during nighttime. With a distribution having heavier tails this issue may be solved. In addition, as all the models include data from previous electricity spot prices, the extreme prices may have too much influence on the following predictions. This is especially the case for the state-space model, which only incorporates data from the previous day in the model, contrary to the other models, which uses a window of the p previous days. A possible solution to overcome this problem include damping the effect of the prices under or above some threshold around its long-term mean. Another solution could be to treat them as outliers, and substitute their values with the mean of the preceding and following hour within the same day t.

Table 4.3: The MSE and CRPS of the Kalman filter solution of the state-space model for both cases considered, in addition to all the hours. The MAE is also given

(a) Case 1.

Hour	MSE	CRPS	MAE
03:00	60.82	6.43	5.63
12:00	60.13	5.43	5.58
18:00	70.64	6.10	5.46

(b) Case 2.

Hour	MSE	CRPS	MAE
03:00	91.11	6.77	6.72
12:00	85.89	5.85	6.64
18:00	94.05	6.39	6.57

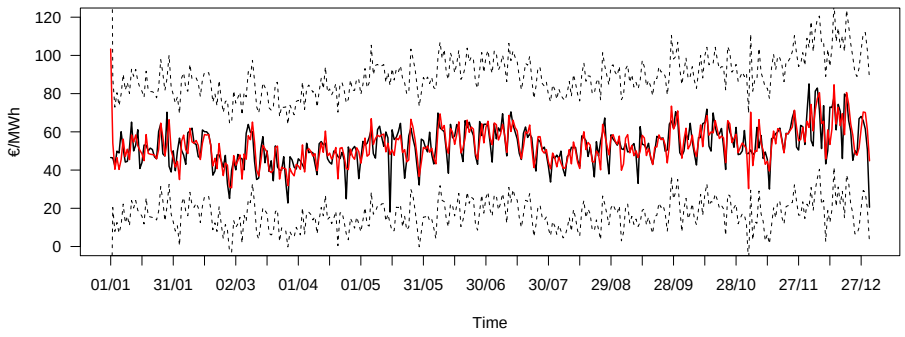


Figure 4.9: The spot price predictions from the Kalman filter (red curve) with the actual spot price for 2010 (black curve). A 95 % confidence interval is included and shown as dotted lines. The units of the second axis is €/MWh, and the dates on the first axis are displayed as day/month.

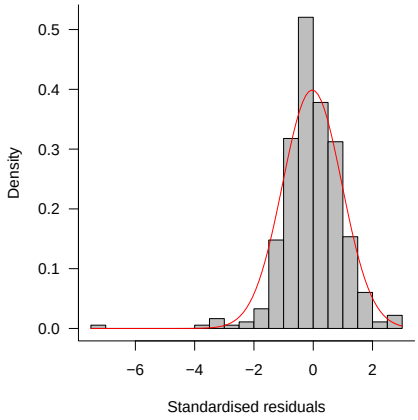
Another interesting aspect that can be seen from figure 4.9 is that the confidence interval is rather large. This is due to the large innovation variances in the prediction step (3.15), that is $S_{t|t-1}$. As pointed out in [36], the Kalman filter also suffers from the presence of multicollinearity, and as discussed in section 4.2, there is evidence of this phenomenon. The same consequences as for the RWR apply to the Kalman filter solution, which explain why the innovation variances are as large as seen here.

Further, in [11] the MAE for the Kalman filter solution for the hours 11:00-14:00 is, in comparison, 3.70. However, in this article the period studied is from the 28th of Janury 2010 to the 28th of February 2013. As the data studied here are from the year 2010, as well as the hourly interval 12:00-13:00, the MAE is not directly comparable. However, this indicates that the model in the article performed better with respect to the predicting the electricity spot prices. This may be due to the method from which the parameters in the state-space model are estimated; in the article mentioned, the maximum likelihood is used to perform parameter estimation. Thus, this gives evidence that parameter estimation through maximum likelihood is preferred rather than the method which is used here.

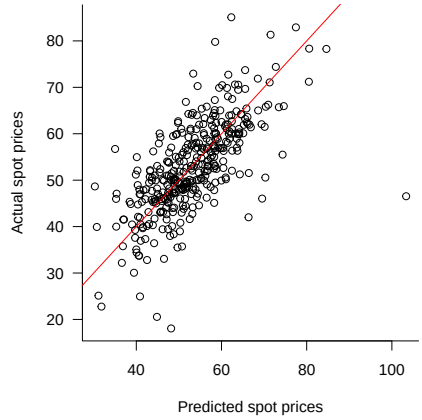
To assess the model further, various diagnostic plots are plotted and shown in figure 4.10. The histogram of the residuals, shown in figure 4.10a, looks rather symmetric, and seems to follow the normal curve in red rather well. However, there is one extreme value in the left tail. In the plot in figure 4.10b it is evidence that there exists a linear relationship between the predictions and the actual values of the spot price. In fact, the regression slope between the predicted and actual spot price is 0.987, with a corresponding standard error of $7.51 \cdot 10^{-3}$. Although this regression slope is furthest away from that of a perfect fit compared to the other models, the standard error is the smallest. Thus, this indicates that the predictions from the state-space model are the most precise. Also here there are some extreme values, which causes the regression slope to be further away from a perfect fit. As for the plot in figure 4.10c, there is a negative relationship between the residuals and the fitted values. This relationship is stronger than found in the corresponding plot of the persistence and RWR model, with a slope of $-1.661 \cdot 10^{-3}$. When testing the significance of the regression slope, the null hypothesis of a slope of 0 is not rejected on a 5 % significance level. On a 10 % significance level, on the other hand, there is sufficient evidence to reject the null hypothesis. This may indicate that the assumption that the variance of $v_{i,t}$, that is R_i, is independent of the day t is wrong. Hence, a GARCH structure of the $v_{i,t}$ may be more appropriate [37]. From the QQ plot in figure 4.10d it is seen that both tails are heavier than for a normal distribution. A Shapiro-Wilks test also rejects normality of the residuals.

Due to these observations, a distribution having heavier tails may also be tested further.

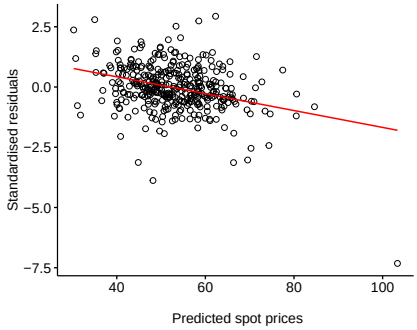
Figure 4.10: Different diagnostic plots to evaluate the persistence model with a window size of $p = 7$.



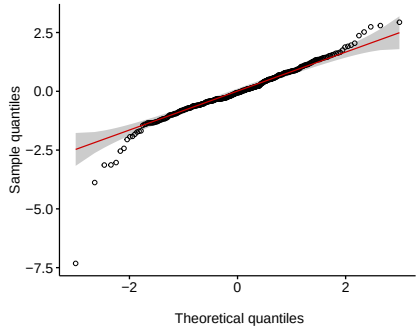
(a) The histogram of the standardised residuals with the normal density curve in red.



(b) Scatter plot of the predicted values of the spot prices, $\hat{y}_{12,t}$, versus the actual spot prices, $y_{12,t}$. The red diagonal line indicates a perfect fit. Here the regression slope between $\hat{y}_{12,t}$ and $y_{12,t}$ is found to be 0.987.



(c) Scatter plot of the predicted values, $\hat{y}$, versus the standardised residuals. The red line shows the regression line.

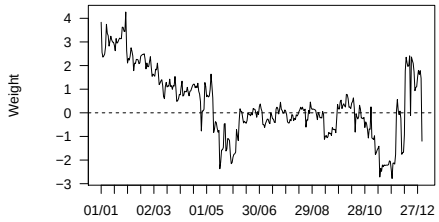


(d) The normal Q-Q plot with a 95 % confidence interval indicated by the grey area.

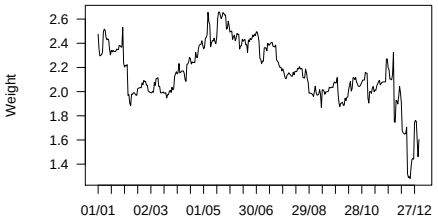
Similarly for the RWR model, the weights for coal price, gas price, photovoltaic and wind and the expected demand are plotted and displayed in figure 4.11. First, the weight of coal in figure 4.11a behaves rather strange, as it does for the RWR

model. Similarly, this may be explained by the multicollinearity. As for the other weights, these correspond better with respect to the correlation matrix in figure 2.5, as their signs coincide. The weight of gas, seen in figure 4.11b, exhibits a different behaviour than seen in figure 4.8b. Whereas in the least squares solution of the RWR model the weight is in the interval $[0.27, 4.85]$, the Kalman filter provides the narrower interval $[1.28, 2.66]$. With this solution, the impact of the gas prices do not vary as much throughout the year as in the RWR model's solution. The weight of photovoltaic and wind in figure 4.11c also behaves differently than in the RWR model, and obtains a maximum around November. However, as the Kalman filter estimates the weights differently, some deviations are expected. Lastly, the weight of the demand in figure 4.11d behaves rather similarly as in the case of the RWR model's solution, as seen in figure 4.8d. Note, however, that in the Kalman filter solution, the demand has less impact in the peak months, with a maximum of $1.56 \cdot 10^{-3}$, as opposed to $2.73 \cdot 10^{-3}$ in the least squares solution of the RWR model.

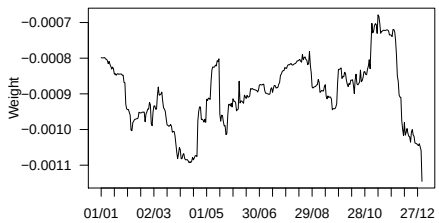
Figure 4.11: Different plots of the development of the weights estimated by the Kalman filter for the hour 12:00. The dates are displayed on the first axis in the day/month format.



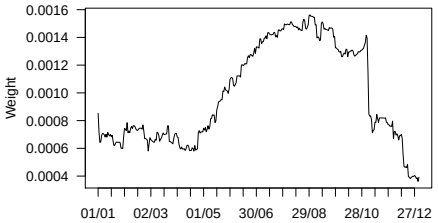
(a) The weight of coal price.



(b) The weight of gas price.



(c) The weight of photovoltaic and wind.



(d) The weight of the expected demand.

Chapter 5

Closing Remarks

In this thesis the aim was to predict the day-ahead electricity spot prices at the German market. Three models were formulated and compared with respect to their predictive performance; the persistence model, belonging to the class of statistical models, the RWR model and the state-space model, the latter ones classified as a hybrid between the statistical and fundamental models. These are fundamental models in that they incorporate variables that are thought to influence the electricity spot prices, thus attempting to explain the physical and economic aspects of the particular market. The fundamental variables included the spot price volatility, fuel prices, demand, to name a few. The persistence model, on the other hand, modelled the electricity spot price as one time series, predicting the day-ahead electricity spot price as the average of a window of previous prices for the same hour. All the models were assumed to be normally distributed. Due to the persistence model's simplicity, the persistence model served as the baseline model, which the other, more sophisticated models, were compared to. All the models were tested with data from 2010 and the hour 12:00. It was found that the RWR model performed best with respect to forecasting the electricity spot prices, with an MSE, CRPS and MAE of 4.70, 3.74 and 5.12, using a window size of $p = 73$. In comparison, the MSE, CRPS and MAE for the persistence model, with a window size of $p = 7$, were 77.64, 4.09 and 6.79, while for the state-space model these were 60.13, 5.43 and 5.58. The fact that the RWR model performed better than the state-space model, as discussed in section 4.3, gives evidence that a model with longer "memory" is preferred.

Despite the satisfying predictions from the RWR and state-space model, performing inference on the weights of the explanatory variables proved difficult. As there was evidence of multicollinearity in the data, this was assumed to cause the prob-

lems. However, this issue did not affect the predictions per se. Further, when assessing the normality assumption of the residuals, the Shapiro-Wilks test rejected the null hypothesis of the residuals. In addition, the diagnostic plots for all the models gave evidence that distribution having heavier tails should be tested further; for the persistence model a distribution having a heavier left tail seem more appropriate, while a distribution having heavier tails in both ends should be tested for the RWR and state-space model. This is thus left for further work. As for the persistence and the RWR model, the residuals did not show any significant linear relationship to the fitted values. This was also the case for the Kalman filter solution on a $5\ \%$ significance level. On a $10\ \%$ significance level, however, this linear relationship was found to be significant. As mentioned, this may indicate that a GARCH model be more appropriate for the errors in this case, which may also be an interesting subject for further work. Lastly, as mentioned in section 3.1.6, there is also evidence that estimating the parameters using maximum likelihood may increase the predictive performance of the model.

All the models considered performed worst when predicting the electricity spot prices at the hour 03:00. As mentioned in section 3.1.6, this may be due to the fact that the price spikes have a too large impact on the following electricity spot prices. As suggested in the same section, the effects of price spikes should be damped, either by limiting the prices below a certain threshold or to treat them as outliers. This is also a suggestion for further work.

Moreover, it would also be interesting to investigate an artificial intelligence based method to forecast the electricity spot prices. As seen in section 2.3, there have been various successful attempts to model the day-ahead electricity spot prices using different NNs. In addition, these are flexible models that are able to handle the complexity and non-linearities that are present in the electricity market. As seen throughout chapter 4, there was evidence that the model assumptions made here were wrong, such as the assumption of normally distributed errors, and in the case of the state-space models, the assumption of homoscedastic errors. Thus, instead of making possibly wrong assumptions of the electricity spot price dynamics, training an NN to learn these dynamics itself is an interesting task for further studies.

Bibliography

[1] Dogan Keles, Jonathan Scelle, Florentina Paraschiv, and Wolf Fichtner. Extended Forecast Methods For Day-Ahead Electricity Spot Prices Applying Artificial Neural Networks. *Applied Energy*, 162:218–230, 2016.

[2] Paula Rocha and Daniel Kuhn. Multistage Stochastic Portfolio Optimisation in Deregulated Electricity Markets Using Linear Decision Rules. *European Journal of Operational Research*, 216(2):397–408, 2012.

[3] Rafał Weron, Ingve Simonsen, and Piotr Wilman. Modeling highly volatile and seasonal markets: evidence from the nord pool electricity market. Econometrics, University Library of Munich, Germany, 2003.

[4] Rafał Weron. Electricity Price Forecasting: A Review of the State-of-the-Art With a Look Into the Future. *International journal of forecasting*, 30(4):1030–1081, 2014.

[5] Rafal Weron. *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, volume 403. John Wiley & Sons, 2007.

[6] Evangelos Kyritsis, Jonas Andersson, and Apostolos Serletis. Electricity Prices, Large-Scale Renewable Integration, and Policy Implications. *Energy Policy*, 101:550–560, 2017.

[7] Andrea Roncoroni, Gianluca Fusai, and Mark Cummins. *Handbook of Multi-Commodity Markets and Products: Structuring, Trading and Risk Management*. John Wiley & Sons, 2015.

[8] Steinar Veka, Gudbrand Lien, Sjur Westgaard, and Helen Higgs. Time-Varying Dependency in European Energy Markets: an Analysis of Nord Pool, European Energy Exchange and Intercontinental Exchange Energy Commodities. *The Journal of Energy Markets*, 5(2):3–32, 2012.

[9] EEX. Products Overview. https://www.eex.com/en/part-of-eex-group/product-overview. [Online: accessed 15/04/2019].

[10] EPEX. About EPEX Spot. https://www.epexspot.com/en/company-info/about_epex_spot. [Online: accessed 01/06/2019].

[11] Florentina Paraschiv, David Erni, and Ralf Pietsch. The Impact of Renewable Energies on EEX Day-Ahead Electricity Prices. *Energy Policy*, 73:196–210, 2014.

[12] Georg Zachmann. A Stochastic Fuel Switching Model for Electricity Prices. *Energy Economics*, 35:5–13, 2013.

[13] NCG. Profile of NetConnect Germany. https://www.net-connect-germany.de/en-gb/About-us/Company-Profile. [Online: accessed 27/06/2019].

[14] EPEX. Negative Prices - How They Occur, What They Mean. https://www.epexspot.com/en/company-info/basics_of_the_power_market/negative_prices. [Online: accessed 01/06/2019].

[15] Alvaro Cartea and Marcelo G. Figueroa. Pricing in Electricity Markets: A Mean Reverting Jump Diffusion Model With Seasonality. *Applied Mathematical Finance*, 12(4):313–335, 2005.

[16] Iivo Vehviläinen and Tuomas Pyykkönen. Stochastic Factor Model for Electricity Spot Price—the Case of the Nordic Market. *Energy Economics*, 27(2):351–367, 2005.

[17] Nektaria V Karakatsani and Derek W Bunn. Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. *International Journal of Forecasting*, 24(4):764–785, 2008.

[18] Jesús Crespo Cuaresma, Jaroslava Hlouskova, Stephan Kossmeier, and Michael Obersteiner. Forecasting Electricity Spot-Prices Using Linear Univariate Time-Series Models. *Applied Energy*, 77(1):87–106, 2004.

[19] Angelica Gianfreda and Luigi Grossi. Forecasting Italian Electricity Zonal Prices With Exogenous Variables. *Energy Economics*, 34(6):2228–2239, 2012.

[20] Paras Mandal, Tomonobu Senjyu, Naomitsu Urasaki, Toshihisa Funabashi, and Anurag K Srivastava. A Novel Approach to Forecast Electricity Price for PJM Using Neural Network and Similar Days Method. *IEEE Transactions on Power Systems*, 22(4):2058–2065, 2007.

[21] J.P.S. Catalão, S.J.P.S. Mariano, V.M.F. Mendes, and L.A.F.M. Ferreira. Short-Term Electricity Prices Forecasting in a Competitive Market: A Neural Network Approach. *Electric Power Systems Research*, 77(10):1297 – 1304, 2007.

[22] S Anbazhagan and Narayanan Kumarappan. Day-Ahead Deregulated Electricity Market Price Forecasting Using Recurrent Neural Network. *IEEE Systems Journal*, 7(4):866–872, 2012.

[23] Chen Yan-Gao and Ma Guangwen. Electricity Price Forecasting Based on Support Vector Machine Trained by Genetic Algorithm. In *2009 Third International Symposium on Intelligent Information Technology Application*, volume 2, pages 292–295. IEEE, 2009.

[24] Wei Sun, Jian-Chang Lu, and Ming Meng. Application of Time Series Based SVM Model on Next-Day Electricity Price Forecasting Under Deregulated Power Market. In *2006 International Conference on Machine Learning and Cybernetics*, pages 2373–2378. IEEE, 2006.

[25] William W.S Wei. *Time Series Analysis: Univariate and Multivariate Methods*. Pearson Addison Wesley, Boston, 2nd edition, 2006.

[26] Robert H Shumway and David S Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer Texts in Statistics. Cham, 4th ed. 2017. edition, 2017.

[27] Carlos F.M. Coimbra and Hugo T.C. Pedro. Chapter 15 - Stochastic-Learning Methods. In Jan Kleissl, editor, *Solar Energy Forecasting and Resource Assessment*, pages 383 – 406. Academic Press, Boston, 2013.

[28] Richard Arnold Johnson. *Applied Multivariate Statistical Analysis*. Prentice Hall, Upper Saddle River, N.J, 4th edition, 1998.

[29] Eric Zivot and Jiahui Wang. *Modeling Financial Time Series With S-Plus®*, volume 191. Springer Science & Business Media, 2007.

[30] James Durbin and Siem Jan Koopman. *Time Series Analysis by State Space Methods*, volume 24 of *Oxford statistical science series*. Oxford University Press, Oxford, 2001.

[31] Rob J Hyndman and Anne B Koehler. Another Look at Measures of Forecast Accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006.

[32] Tilmann Gneiting and Adrian E Raftery. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

[33] CJ Willmott and K Matsuura. Advantages of the Mean Absolute Error (MAE) Over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. *Climate Research*, 30(1):79–82, 2005.

[34] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.

[35] Douglas C Montgomery. *Introduction to Linear Regression Analysis*. Wiley Series in Probability and Statistics. Wiley, Hoboken, 5th ed. edition, 2013.

[36] PK Watson. Kalman Filtering as an Alternative to Ordinary Least Squares - Some Theoretical Considerations and Empirical Results. *Empirical Economics*, 8(2):71–85, 1983.

[37] Robert Engle. Garch 101: The use of arch/garch models in applied econometrics. *Journal of economic perspectives*, 15(4):157–168, 2001.

Appendix

A Evaluation using MAE

Table A.1: The optimal p for the persistence (left table) and the RWR model (right table) with respect to MAE, for all the hours.

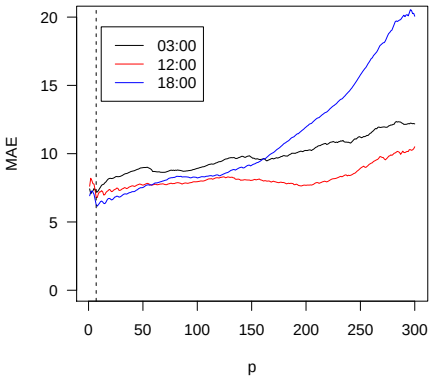
(a) The p yielding lowest MAE for the persistence model.

Hour	p	MSE	CRPS	MAE
03:00	2	97.82	4.25	7.06
12:00	8	76.44	4.10	6.79
18:00	8	74.15	3.71	6.16

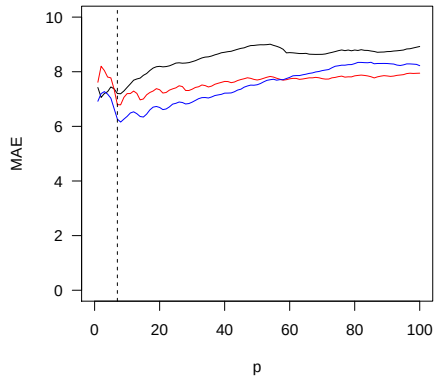
(b) The p yielding lowest MAE for the RWR model.

Hour	p	MSE	CRPS	MAE
03:00	73	52.55	4.08	5.70
12:00	86	46.95	3.72	5.11
18:00	70	53.53	3.79	5.14

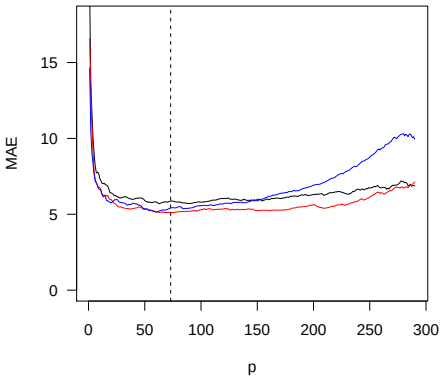
Figure A.1: The MAE for the persistence (upper panel) and RWR model (lower panel) for various window sizes p and the hours 03:00 (black curve), 12:00 (red curve) and 18:00 (blue curve). vertical dotted line at $p = 7$ is also shown in the upper panels, whereas in the lower panels the dotted line is at $p = 93$.



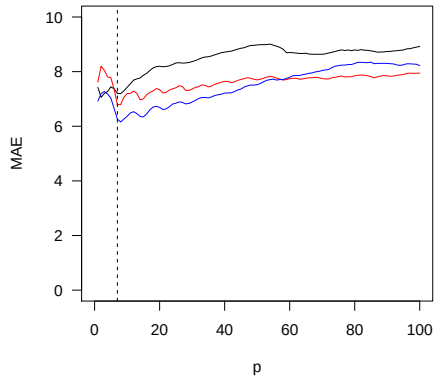
(a) The MAE for for the persistence model. Here $p \leq 300$.



(b) The MAE for the persistence model. Here $p \leq 100$



(c) The MAE for for the RWR model. Here $p \leq 300$.



(d) The MAE for for the RWR model. Here $p \leq 100$.