Forecasting of Electricity Prices Accounting for Wind Power Predictions

Tryggvi Jónsson

Kongens Lyngby 2008
IMM-THESIS-2008-

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253351, Fax +45 45882673
reception@imm.dtu.dk
www.imm.dtu.dk

Abstract

For players in deregulated energy markets such as Nord Pool and EEX, price forecasts are paramount when it comes to designing bidding strategies and are an important aid in production planning. In addition, price forecasts can be of great value for grid operators who are responsible for keeping the grid in balance. It is a known fact that electricity prices on Nord Pool's spot market are, in the long run, mainly influenced by the level of water in the reservoirs of the Norwegian and Swedish hydropower plants. However, changes in the water level happen slowly and are therefore not a matter of great relevance when forecasts are made for the prices at the Nord Pool spot market on a relatively short horizon. In this thesis, the effects of predicted wind power production on the spot prices in Nord Pool's Western Danish price area (DK-1) are investigated. Moreover, ways of including the predicted wind power production in a forecasting model not only for the mean spot price in DK-1, but also the full distribution of the prices, are explored. It turns out that the effects of forecasted wind power production on the spot price is substantial and even more effects can be found with small modifications. The forecasting model constructed consists of three mains parts. The first part accounts for the effects of external factors on the prices while the second one is a dynamic model of the spot prices that accounts for the effects found be the first model. The final layer adds valuable information about the uncertainty or the distribution of the prices. Combined these models give reliable non-parametric description to the full distribution of the spot prices. Given the result of this thesis, it is very likely that the same methodology will give good results when forecasting the prices on other electricity pools. It is expected that the approach will be highly beneficial both for pools where wind power penetration is relatively high, and for markets with other characteristics, such as regulation markets.

Keywords: *electricity spot prices, wind power, forecasting, statistical modeling, non-parametric modeling*

Preface

This thesis was prepared at Informatics Mathematical Modelling, the Technical University of Denmark, under the supervision of Prof. Henrik Madsen and Assoc. Prof. Pierre Pinson, in partial fulfillment of the requirements for acquiring the M.Sc. degree in engineering.

The thesis deals with forecasting of prices at deregulated electricity markets.

The project was carried out in the period from August 1st 2007 until June 1st 2008.

Lyngby, May 2008

Tryggvi Jónsson

Acknowledgements

First and foremost I would like to thank my supervisors, Prof. Henrik Madsen and Assoc. Prof. Pierre Pinson for their guidance and help. I would also like to thank Jan Frederik Foyn at Nord Pool ASA, Nina Detelefsen at Energinet.dk, Torben Skov Nielsen at ENFOR A/S for providing the necessary data. Furthermore, I would like to thank Sigurón Björnsson, Lars Bruun Sørensen and Peter Lyk-Jensen at Dong Energy for fruitful discussions and their interest in the project. Same goes for Jacob Vive-Munk at Nordjysk Elhandel and Nis Kjær, Philip Røpcke and Jacob Skovsby Toft at Markedskraft Danmark A/S. Finally, I would like to thank Fannar Örn Thordarson, Snorri Páll Sigurdsson and Thórhildur Thorkelsdóttir for proof reading parts of the text and assistance regarding the layout of the thesis.

Contents

CHAPTER 1

Introduction

With growing proportion of energy trading being done on international energy exchanges, such as Nord Pool in Scandinavia and EEX in Germany, and with expanding geographical areas which these exchanges cover, the need for more advanced market price forecasting methods has increased. Furthermore, increased focus on renewable energy sources, many of which produce non dispatchable energy, has made prices more volatile and therefore forecasting of them more difficult. This development will without doubt continue with EU's recently presented 20% 2020 target[1] and with increased focus on renewable energy in the USA and P.R. China. These forecasts are however essential for traders bidding on these markets on behalf of both producers and retailers. For them, knowing not only a predicted price, but also knowing the uncertainty of the forecast is paramount in their decision making. Furthermore, price forecasts can be helpful for Transmission System Operators (TSOs) during planning. Existing models do not properly account for dependence of market prices on variable generation sources (such as wind power), and pay no attention to related forecast uncertainty.

The objective of this thesis is to derive a model suitable for forecasting electricity spot prices on deregulated energy exchanges. The focus is primarily on Nord Pool's Western Danish price area (DK-1), nevertheless the desired possibility of applying the model on other areas as well is kept in mind during the

[1] The target states that at least 20% of every EU member country's energy consumption has to be generated by renewable energy sources by 2020. The target is binding for all member countries.

development of the model. However, due to difficulties with acquiring data, no analysis is done on other areas. Still, DK-1 is a very interesting area to conduct this kind of research on. A large share of the generation capacity in DK-1 is installed in the form of conventional thermal plants, however the system is also heavily penetrated by wind power, a non dispatchable energy source. In fact, DK-1 is currently the grid area in the world which has the largest share, of its total energy production, generated by wind turbines. Furthermore, Denmark's geographical location makes it a sort of connection hub between the hydro power region in the North and the rest of Europe. DK-1 is therefore a good representative for what the future in European energy markets will look like, a grid heavily penetrated by energy generated by a free source with very variable availability and with large connections to surrounding areas.

A consequence of the large proportion of wind energy in the system is that considerable efforts are spent on estimating the effects of wind power, or forecasted wind power, on the prices. Recent studies have shown that the measured wind power affects the spot prices [11, 30] and also in theory this effect should exist, owing it the virtually nil marginal cost of wind power. However, wind power production is very volatile and it can not be planned in great detail before delivery. Consequently, measured wind power production can not be used as an input for a model that is used for forecasting. However, commercial forecasting models for the power generated by wind turbines exist and are widely used for production planning and bidding. With the goal of finding a model suitable for forecasting the spot price several hours ahead in mind, the effects of such forecasts are analyzed in this thesis. An analysis that is not known to have been conducted before. As it turns out the forecasts also affect the spot prices in DK-1 and are therefore used for improved forecasting of the spot prices. Moreover, the impact of other variables, such as time factors is examined as well.

Knowledge of these effects is then used to construct a model that estimates not only the mean price but also the full distribution of the prices in a non-parametric fashion. Since a model of the spot prices is only one bit in a larger puzzle of modeling an energy market completely, the last part of the study are devoted to a preliminary analysis of how the next steps towards this goal could look like.

1.1 Previous Work

During the past decade, forecasting of electricity prices has been of increased interest. As the proportion of electricity traded on deregulated electricity markets has increased, price forecasts have become a more important tool in strategic planning. Several papers have been written on constructing point forecast-

ing models for electricity prices on these competitive markets, taking many different approaches to the modeling - Both statistical and economical. In [5], the spot prices on the Leipzig Power Exchange, LPX (later merged with the European Energy Exchange, EEX), are modeled with linear univariate time series models. Despite some interesting results for 1 hour ahead forecasts, the approach taken in [5] does not, as implied by the name, account for any non-linear behavior in the prices and external effects are completely ignored. Same applies for [4] where other techniques of univariate modeling are compared. The lack of attention given to non-linearities and external factors limits the models' potential of being extended to a longer horizon. Furthermore, the assumption of static market conditions throughout the year leads to very volatile performance of the models. In [1], many different tasks in modeling energy prices are addressed with some interesting results. Yet the majority of the criticism presented above, applies for most of the models proposed there.

Many papers have addressed the issue of including external factors. Both [25] and [32], consider the effects of electricity demand in the models presented. In [48] and [52] different versions of oligopoly equilibrium models are presented as a way of modeling prices with external factors accounted for. External factors such as weather data and demand. As an alternative to pre-defining which external effects are to be accounted for, [15] presents Input/Output Hidden Markov Models as a way of switching between market states. The different methods used in all these papers all have their distinct pros and cons. However they all have in common, that point forecasts are provided, by the model, along with a prediction interval that is normally distributed around the mean. However, this assumption is known to be wrong as stated in [6] as the distribution of energy prices is skewed in the positive direction and more heavy tailed than a normal distribution. Therefore, these models provide useful information about the mean price only, since information about the forecasting uncertainty is obtained under the wrong assumptions.

In [6] a quantile based forecasting method is proposed for obtaining better uncertainty estimates on electricity prices in California and North Eastern USA. This method is more flexible and does not assume normal distribution and does therefore serve better for estimating uncertainty of the electricity prices.

1.2 The Data

The majority of the historical market data is obtained from Energinet.dk's website, `www.energinet.dk`, while some parts of the data were made available by Nord Pool ASA through their database. Wind power production forecasts are made by the Wind Power Prediction Tool (WPPT)[31] and were provided by ENFOR A/S with permission from Energinet.dk.

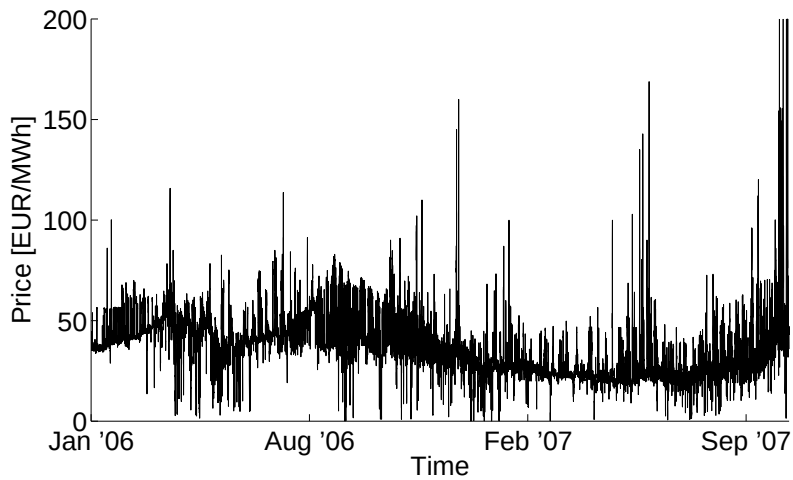

Figure 1.1: Time series plot of the spot prices at Nord Pool's Elspot

All analysis and modeling made were carried out on a data set that covers the period from 4th of January 2006 until 31st of October 2007. It consists of hourly observations of the electricity spot prices in the Western Danish price area (DK-1) along with the hourly measurements of electricity consumed in the same area. Furthermore, quarterly forecasts of produced wind power in DK-1 measured in MW were available. The forecasts account for 15 out of the 16 regions DK-1 is divided into, but forecasts for the offshore wind farm at Horns Reef were not available due to confidentiality issues.

For coordinating the time and measurement units on observations, the wind power forecasts are converted into hourly production forecasts measured in MWh. This is done by linearly interpolating between each two adjacent forecasts in every hour and taking the result as the production in MWh for that quarter. Then these interpolations are summed up for each hour. So in mathematical terms, each hourly forecast is obtained by

$$\widehat{WG}_{hh}^{hourly} = \sum_{i=2}^{5} 0.25 \cdot \left(\frac{\widehat{WG}_{t(i-1)}^{quarterly} + \widehat{WG}_{t(i)}^{quarterly}}{2} \right) \tag{1.1}$$

During the work on the thesis, forecasts of the electricity consumption were desired, but were not available. Since it was desired to focus on forecasting prices, consumption forecasts are simulated by taking the measured consumption and adding to it a white noise with standard deviation of 2% of the mean consumption. The percentage chosen as standard deviation is deemed reason-

Figure 1.2: A demonstration of when prices become available on Nord Pool's spot market

able since most state of the art methods mentioned in [50] perform that well for all prediction horizons relevant here. Moreover, standard deviation on the interval $0 - 5\%$ only has very little effect on the results presented.

1.3 Time indexes in Modeling Notation - Mathematical Perspective vs. Real Life

To avoid confusion about the special time indexing of forecasting models for the spot price, a brief explanation will be given here.

A k-step ahead prediction for the value of some output variable Y is commonly written as $\widehat{Y}_{t+k|t}$. This notation is read as "the predicted value of Y at time $t+k$, given observations until time t". This will lead to that the 1 step ahead prediction of a process $\{Y\}$, calculated as regression model of only the latest available observation of $\{Y\}$, will be given as

$$\widehat{Y}_{t+1|t} = \phi Y_t \tag{1.2}$$

where ϕ is an estimated regression coefficient.

Now consider the time-line, representing when observations become available on the spot market, given in Figure 1.2.

Imagine the clock is little before 12:00 (time t) on some given day. As will be explained in detail in Chapter 3, the electricity spot prices are now known until the following midnight (time $t+12$), and are about to be set for the following 24 hours after that. This translates into that in real life, at all forecasting times relevant in the analysis made for this thesis, prices are known for at least the next 12 hours. This means that a forecast made for 13 hours ahead in real life is a one hour forecast from a mathematical point of view.

In the notation of this thesis k refers to for how many real life hours ahead the forecast is made for and t refers to the time, in real life, that the forecast is made. Hence, when forecasts are made for $k = 13$, the latest available observation is

from $t+12$ and (1.2) for this setting becomes

$$\widehat{Y}_{t+13|t+12} = \phi Y_{t+12} \tag{1.3}$$

On the real time market (see Chapter 3 for description) the prices are set continuously. Therefore does $k = 12$ refer both to 12 hour prediction in real life as well as from mathematical point of view.

1.4 Thesis Overview

The thesis is structured as follows:

Chapter 2 provides an overview of the power production in the Nordic countries and how the energy market has transformed from a oligopoly structure to a market based structure.

Chapter 3 covers the functionals of Nord Pool's markets, Elspot, Elbas, the regulation market and the Financial market.

Chapter 4 lists the mathematical theory used in the study.

Chapter 5 shows the construction of static seasonal models for the spot price.

Chapter 6 presents analysis of the external factors affecting the spot prices.

Chapter 7 contains a description of the more elaborate point forecasting model derived for the spot price.

Chapter 8 covers the construction of prediction intervals using quantile regression.

Chapter 9 introduces preliminary analysis of how the first steps towards a model for the regulation market could be done.

Chapter 10 contains conclusions drawn from the study.

CHAPTER 2

Energy Production in Northern Europe and the Nordic Power System

The electricity generation methods and the power market's structure in northern Europe, have been in constant development during the past decades. This chapter provides an overview of the power production in the Nordic countries and how the energy market has transformed from a oligopoly structure to a market based structure.

2.1 Electricity as a Commodity

As for other free commodities, prices of electricity are expected to reflect production costs of the last unit sold. If not, new producers will enter the market and the old ones fall out. However, due to the nature of electricity and its importance to the western society, there are three significant things that make electricity different from other commodities. [8]

- Electricity is difficult to store[1] and has to be available on demand. Con-

[1] Numerous methods exists but are not cost efficient, have technical limitations and rely on specific natural conditions.

sequently, it is not possible to stock it or have customers queue for it. Therefore, the generation of electric power must match the demand at all times to prevent frequency fluctuations and ensure system stability.

- Transporting electricity from producers to consumers requires special and expensive infrastructure called a transmission system. The system can only be used for transportation of electricity and can transport it long distances in a split second. There are however limitations on how much energy can be transported simultaneously, and due to high building cost, the transmission system is often fully or close to fully utilized.
- The demand for electric power is inelastic, i.e. consumers do not respond to price changes. For this there are many reasons. Two obvious ones are that no other commodity can easily replace electricity and that small consumers are not affected by the market price instantly, since their contracts with retailers are only revised annually or more infrequently.

2.2 Production

The differences in the landscape of Scandinavia are reflected in the electricity generation methods of the respective countries. The flat landscape of Denmark has been one of the driving forces behind a large wind power industry, while the deep fjords in Norway and the mountains in Lapland give great opportunities for production of hydropower. Nuclear power is utilized in Finland and Sweden and along with thermal power which is also the main source of power in Denmark. Here, the intention is to give a brief overview of the production methods used in Scandinavia and how generation capacity is spread across the region. In Table 2.1 the annual production of each country by each method is summarized.

2.2.1 Hydroelectric Power

Hydroelectric power is produced from the potential energy of the elevation of water. In most cases large dams are used to create a reservoir, from which the water flow is controlled and sent down a penstock and drives a turbine and a generator. So called free flow hydropower plants also exist, where no reservoir is made and they are therefore not controllable. Hydropower now supplies about 19% of the world's electricity and large dams are still being designed and built. However the potential of harnessing more hydropower in Scandinavia is considered exhausted. Using water in this manner to generate power has

Table 2.1: Summary of electricity production in Scandinavia and Germany in 2006 [12, 41]

Country	Hydro		Nuclear		Thermal		Wind		Biomass		Other		Total
	TWh	%	TWh	%	TWh	%	TWh	%	TWh	%	TWh	%	TWh
Denmark	-	-	-	-	34.4	79.6	6.1	14.1	0.2	0.5	1.9	5.8	43.2
Finland	11.3	14.4	22	28	33.9	43	0.2	0.3	10.1	12.9	1.1	1.4	78.6
Norway	119.9	98.5	-	-	0.4	0.3	0.7	0.6	0.4	0.3	0.3	0.3	121.7
Sweden	61.2	43.7	65	46.3	3.8	2.7	1	0.7	8.2	5.8	1.1	0.8	140.3
Total in Scandinavia	192.4	50.1	87	22.7	72.5	18.9	8	2.1	19.5	5.1	4.4	1.2	383.3
Germany	21.5	3.6	159	26.6	335	56.2	30	5.1	18.6	3	31.9	5.5	596
Total	213.9	21.8	246	25.1	407.5	41.6	38	3.9	38.1	3.9	36.3	3.7	979.8

the economic advantage that the cost of fuel is eliminated making the electricity produced cheap, despite high building costs. Furthermore, hydroelectric power plants do not emit any carbon dioxide and are therefore classified as a renewable source of energy. Nevertheless, power plants of this type have met increasing opposition, since they require the sinking of enormous land space and can therefore jeopardize ecosystems. [53]

Hydropower makes up for virtually all electricity production in Norway, and is also a considerable proportion of the power generation in Sweden (43%) and Finland (14%). The hydroelectric power plants are mainly situated in the northern parts of Sweden and Finland, while they are spread all over Norway. [35, 41]

2.2.2 Nuclear Power

In nuclear power plants, electricity is produced by converting the nuclear energy of fissable uranium into thermal energy by fission. Then the thermal energy is converted into kinetic energy by a steam turbine which drives an electricity generator. Nuclear power plants provide steady energy at a consistent price. Fuel costs are lower than in thermal power plants, but building costs are higher and maintenance and security is more expensive. Since energy production can only be altered slowly, nuclear power plants are typically used to handle the base load in the system, while the peak loads are left to more flexible power plants. Although nuclear generation of electricity produces no greenhouse gasses and virtually no airborne pollution, nuclear power has been contested. The long term storage of the highly radioactive waste nuclear power plants produce, which needs to be handled with great care and forethought due to long half-live, have been an issue and security risks has been connected to accidents at Three Mile Island and Chernobyl.

The only existing nuclear power plants in Scandinavia are located in Sweden and Finland and make up for 46% and 28% of the countries' electricity production, respectively. However, focus has been directed from nuclear generation of electricity in Sweden and moved towards renewable energy sources. Meanwhile, the building of a new nuclear power plant was accepted in the Finnish parliament in 2002 and since 2005 there has been growing vocal support from the industry and government alike, for building a 6th nuclear power reactor in order to lower electricity costs and meet Finland's obligations under the Kyoto protocol. Moreover, Germany is one of the largest producer of nuclear electricity in the world with annual production of around 160 TWh or about 20% of the country's electricity production. In Germany, the situation is the same as in Sweden, due to public protest more attention has been given to renewable energy sources. [20, 51, 53]

2.2.3 Thermal Power

A thermal power plant converts the energy stored in fossil fuels such as coal, gas and oil successively into thermal energy, mechanical energy and finally electric energy. Every plant is a unique custom designed system and its production capacity can vary from KW to GW. Plants are either built to produce electricity only or both electricity and hot water for commercial use. Starting a new thermal power plant is normally quite expensive and operation is only cost efficient when the output is within a certain range. Fuel costs are high for any type of thermal plant and for especially gas plants. Apart from large emissions of greenhouse gases, airborne pollution from thermal plants include heavy metals such as mercury and radioactive waste. Therefore, the introduction of carbon dioxide emission quotas has also led to increased operation costs. Due to the above and international environmental agreements such as the Kyoto protocol, the use of thermal power plant has increasingly been the source of environmental concern. The flexibility of the thermal power plants varies with the fuel used for generation. Plants running on natural gas are very flexible and can therefore be used to handle peak load, while coal plants are very inflexible and are mainly used to supply the base load.[53]

Thermal power plants produce by far the largest share of the electricity production in Denmark and a considerable share in Finland and Germany as well. In Sweden, thermal power is used to a very limited extend and in Norway thermal plants hardly exist. Despite the opposition against thermal power plants, the task of eliminating them seems impossible due to the large share of electricity production they provide. [35, 41]

2.2.4 Wind Power

A wind turbine converts the kinetic energy in wind into mechanical energy, which is then converted into electricity. Modern windmills have production capacity up to 5 MW at optimal conditions. A number of wind turbines are often collected together into a so called wind farm, which can be found both off-shore and on shore. The production of a wind turbine can not be controlled to the same extent as many other power plants. Electricity is only generated when wind speed ranges from 0 m/s to 25 m/s, since the turbines have to be shut down for machine protection if the wind speed exceeds 25 m/s. This results in an annual production quantity that is around 15% of installed capacity on average. Due to the direct relation between wind speed and production quantity, production quantity can be extremely variable throughout the day. Therefore, good and reliable forecasts for wind forecasts and production forecasts are essential when planning the production in a system that contains wind turbines. Starting costs of a wind farm can be rather high, but once they

are up and running, the production cost is extremely low since maintenance and production costs are low and fuel cost is non existent.[7, 53]

Denmark is a leading nation in design, production and harnessing wind power, with about 15% wind power penetration in its power grid and installed capacity of little over 3 GW. Germany is also among the leading nations in the wind energy sector with the largest installed wind power capacity of around 18 GW, which provides Germany with 5% of their annual electricity production. Other countries in Scandinavia all have few wind turbines installed, and the production is so small that it is almost negligible. [35, 41]

2.2.5 Other Power Generation Methods

Various alternative methods are also used to a little extend in the countries that are of interest here. All countries have some electricity production from bio mass and waste. Germany gets a small share of its electricity production from geothermal power plants and in some of the countries experiments have been made with solar cells.

2.3 Transmission

Once the electricity has been generated, it has to be transmitted from the production plant to the users. The transmission systems in the Scandinavian countries are all driven by a non profit organizations called Transmission System Operators (TSOs), which are responsible for operating the country's transmission grid and connections to other countries. A further discussion of the TSOs can be found in Section 2.4.2.1.

The national grids along with the interconnections between countries allow the power to flow from areas where production is high to areas where production does not meet the demand. For instance in periods of sufficient reservoir levels in the hydropower plants in the North, the transmission system is used to transport the hydropower to the more dense populated areas in the South, and the other way around when reservoir levels are low or wind power production is extremely high. Power is not only transmitted within the Nordic countries, but also sold further South. Energy is, therefore often transported through Denmark due to its geographical location between the hydropower area in the North and the rest of Europe. Finally, cheap nuclear power is imported to Finland through interconnections to Russia.

2.3.1 Market areas

The Nordic Power Exchange region is divided into areas, which have internal transmission capacity that can be considered unlimited. Finland and Sweden make up for one such area each. Denmark is divided by the Great Belt into the Eastern and Western areas, and until October 2007, Norway was divided in three areas. Then, the fourth Nord Pool area was created in southern Norway, due to a new connection between Southern Norway and the Netherlands. In Figure 2.1, the boundaries of the market areas in the Nord Pool region are shown as they were in September 2007, along with the transmission capacities between areas and to other countries.

2.3.2 Varying capacities

The transmission system is not a static system. Maintenance, breakdowns and limitations often lead to reduced transmission capacity between areas, so that little or no power can be transported through individual connections. Such reduced transmission capacity is bound to effect the power price, although the extent of the effects depend on between which areas the interconnection is down and other aspects such as the climate as well as availability of the different fuel types in the down period.

2.3.3 Point tariff system

Like all other service companies, the TSOs receive tariffs for their services of transmitting power from producer to consumer. The idea behind the system of point tariff, used by the Nordic TSOs, is that the producers pay a fee to the grid operator for every *KWh* they pour into the grid. Correspondingly, the end users pay a fee for every *KWh* they draw from the grid. Moreover, the kilowatt hour can be traded freely within the whole area. This means for instance that a consumer in southern Sweden can purchase power from a producer in the northern part of the country without any limitations. In such a trade, the producer ensures that the quantity of electricity agreed on is poured into the grid and the consumer draws the same quantity from the grid. However, the energy does not flow directly from producer to consumer and therefore does the consumer not receive power specifically generated by the producer. [35, 42]

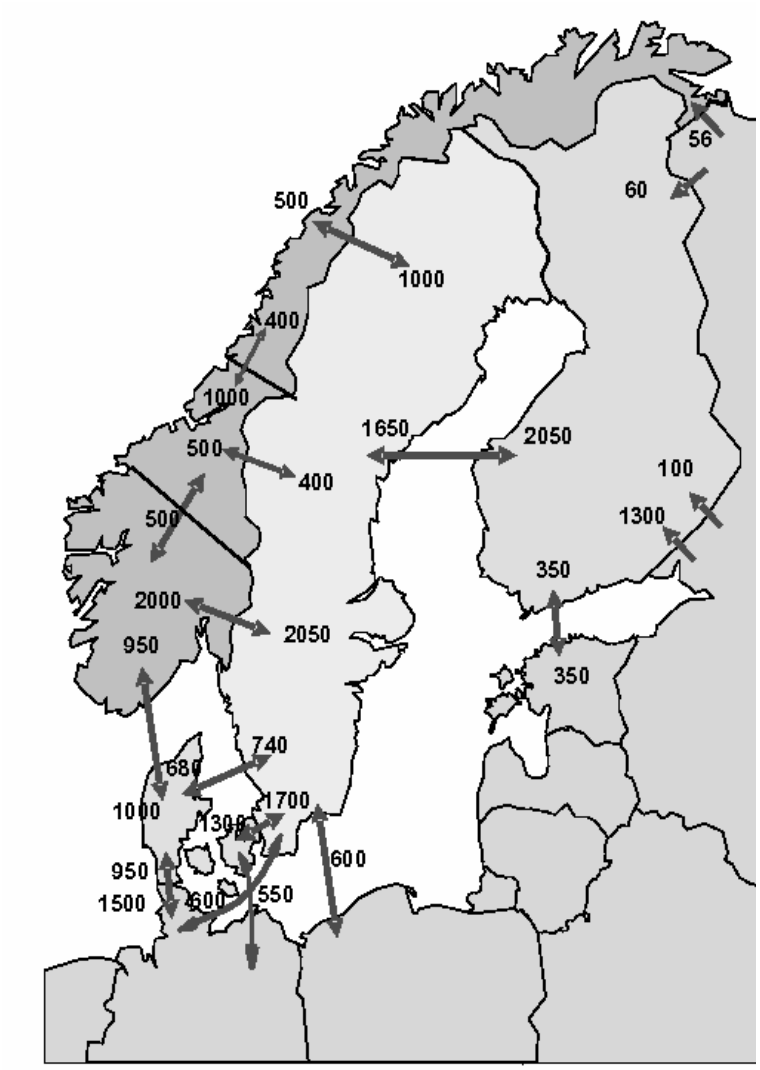

Figure 2.1: Geographical layout of the areas in the Nord Pool region and transmission capacities between them[34]

2.4 The Market Structure - Reforms and Current structure

The history of electricity trading in Scandinavia spans almost 50 years and in this section, a brief overview of this history is given along with a discussion of the markets functions today.

2.4.1 The market structure before deregulation

Before the deregulation of the electricity markets in Scandinavia began in 1991, the markets in Norway, Sweden and Finland all had an oligopoly structure where state owned companies held a dominant position and also controlled the transmission grid. Despite the markets shared similar characteristics, the situation were not entirely the same in the countries.

In Norway, the state owned Statkraft dominated the power sector and also operated the transmission grid. Local and regional utilities gained access to the national grid in 1969 and could exchange energy on a spot market. Many small local and regional utilities, commonly owned by local authorities, were involved in transmission of energy at the regional level. Distribution was on the hand of about 200 local companies, mainly owned by municipalities.

In Sweden, about 50% of the generation was carried out by the government owned Vattenfall, which also operated the transmission grid. The remaining generation was mainly on the hands of about ten other utilities of various sizes, though all rather big. High transmission fees made difficult for smaller utilities to operate. Similar to Norway, numerous distribution companies existed, many owned by municipalities.

The Finnish power sector was also dominated by a state owned company, Imatran Voima Oy (IVO), which operated the national grid. However a large share of the generation was owned by Finnish industries, which formed their own transmission company, TVS, to interconnect their generation to supply areas. Hence, in Finland there were two overlapping grids.

Due to geographical reasons, the grid in Denmark was divided in two main parts. One consisting of Jutland, Funen and the islands west of the Great Belt (DK-West, DK-1), while Zealand and the islands east of the Great Belt make up the other one (DK-East, DK-2), apart from the island of Bornholm. In both of these areas, the generation and distribution was on the hands of companies mostly owned by municipalities, which formed organizations specially for the purpose of manage the extra-high voltage grids and coordinated operations.[2]

2.4.1.1 The Nordel cooperation

In 1963, the Nordel organization was founded by the largest electricity producers in Norway, Sweden and Finland. The objective was to enable cooperation between the producers. Nordel was based on the principle that each country would be self-sufficient in terms of generating capacities and trading was meant to achieve optimal dispatch of a larger system. Investments in interconnections were primarily made based on expected savings, but not on net exports. The countries exchanged information about marginal costs and if there was a difference, an exchange was made, resulting in a price that was the average of the two prices.

The cost-plus structure of the Nordic power sector led to over investments and poor return on equities. However the competition had positive effects on the utilities, where no significant efficiency problems were experienced.[2]

2.4.2 Shifting to a market based structure

Worldwide, the shift to a market based structure was triggered by the deregulation of the electricity system in England and Wales in 1990. In Scandinavia, Norway led the way by the Energy act in 1990, which took effect in 1991. The intention was to reduce regional differences in power costs, enhance operational efficiency in generation and distribution and develop the power sector towards more efficiency. A new state owned company, Statnett SF, was founded around Statkraft´s transmission activities and all transmission networks were opened for third-party access. Furthermore vertically integrated companies were forced to adopt separate accounting for generation, distribution and supply activities.

Similar reforms took place in Sweden, though on more steps. In 1991, Vattenfall's generation and distribution activities were corporatized and a special government owned institution, Svenska Kraftnät, was founded to operate the national grid. Vattenfall remained government owned. The networks then were gradually opened to new players and finally in 1996, a competitive market took effect with a new Energy act.

The Energy Market Act was introduced in Finland in 1995, opening market for competition and setting the ground for further modifications in 1998, that allowed customers to choose their supplier of electricity freely and at no additional cost. The state owned IVO had already separated its transmission activities into a separate company, IVS, when the Energy Market Act was introduced. Finland had two overlapping grids, owned and operated by IVS and the privately owned TVS until 1997, when the companies merged and formed

Fingrid. Fingrid is a fully privatized company, jointly owned by institutional investors, power producers, and the state.

Due to the different structure of the power sector in Denmark, reforms moved more slowly. New legislation opening the grids to negotiated third-party access and allowing competition for large consumers, distributors, and generators was introduced in 1996. In 1998, a market competition was introduced for large producers and consumers. The threshold for participating in that market was 100 GWh, but was reduced gradually until all customers could trade freely from January 1st 2003. In August 2005 the two system operators in Denmark merged in the state owned Energinet.dk, effective from January 1st 2005.

After the reforms of the electricity market in the four Scandinavian countries their power sectors all shared the same characteristics. The electricity system is split into four main parts: Generation, transmission, distribution and retail. Competition is allowed in generation and retail, while transmission and distribution is considered to be a natural monopoly. The non-profit grid companies operate the monopoly parts and are called the transmission system operator (TSO).[2, 10, 34]

2.4.2.1 The role of the TSOs and the new Nordel

The unbundling of power generation and transmission results in decentralized decision making. Therefore, if not controlled, imbalances in the grid are bound to happen. Maintaining the instantaneous balance between supply and demand, along with ensuring operational security of the power grid in it´s area, is therefore the main role of a TSO on daily basis. Furthermore, ensuring and maintaining the short-term and long-term adequacy of the transmission system and enhancing the efficient functioning of the electricity market, also fall under the responsibility of the TSOs.

With the new structure of the power sector in the Nordic countries, Nordel has transformed into a cooperation organization between the TSOs with the Danish and Icelandic TSOs as new members. Current members are therefore: Energinet.dk (Denmark), Fingrid Oyj (Finland), Landsnet (Iceland), Statnett SF (Norway) and Svenska Kraftnät (Sweden). The objectives of Nordel are listed on their website as:

- Development of an adequate and robust transmission system aiming at few large price areas
- Seamless cooperation in the management of the daily system operations to maintain the security of supply and to use the resources efficiently across the borders

- Efficient functioning of the North-West European electricity market with the aim to create larger and more liquid markets and to improve transparency of the TSO operations
- Establishment of a benchmark for European transparency of the TSO information.

[43]

CHAPTER 3

Nord Pool

As a result of the deregulation of the energy market in Norway a power exchange, Statnett Marked A/S, was established as an independent company in 1993. However since all electricity in Norway was hydroelectric power, the spot prices were very volatile. In Sweden, the establishment of such a market was impossible, due to the large market share of Vattenfall and Sydkraft, which controlled about 75% of the generating capacity in Sweden. Since a combined Norwegian-Swedish market would address problems of both nations, a decision of establishing such market was made. The joint exchange market began operating in January 1995 with a design based on Norwegian experience and was named Nord Pool. Nord Pool was owned by the countries´ two grid companies, Statnett SF (50%) and Svenska Kraftnät (50%). Finland entered into Nord Pool in 1998 after agreement was reached with the Finnish power exchange, EL-EX, to represent Nord Pool in Finland. Western Denmark began trading on Nord Pool in July 1999 and Eastern Denmark joined Nord Pool in 2000.

Financial derivatives were first introduced on the Nordic Power Exchange's financial market in 1997. At the same time Nord Pool started offering financial market participants expanded clearing services; in addition to clearing all contracts traded on the Nordic Power Exchange. At last trading of emission allowances started on Nord Pool in 2005.

In January 2002 Nord Pool's physical market was, through a demerger, organized into a separate company, Nord Pool Spot ASA. Nord Pool Spot is jointly

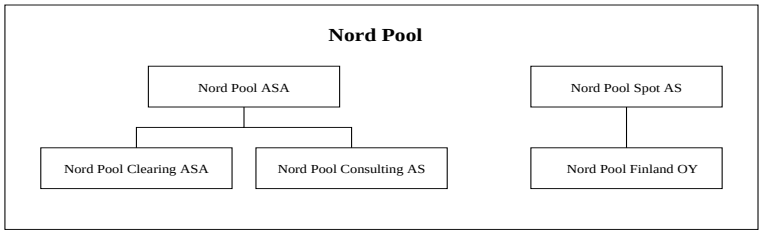

Figure 3.1: Organization chart for Nord Pool ASA

owned by Nord Pool ASA and the TSOs, each with a share of 20%. In March 2002, the clearing services of Nord Pool also demerged into a separate company, Nord Pool Clearing ASA, fully owned by Nord Pool ASA.

Currently the Nord Pool Group comprises four companies, three of which are fully owned by Nord Pool, Nord Pool ASA, Nord Pool Clearing ASA and Nord Pool Consulting AS. The fourth one is Nord Pool Spot ASA.[2, 34, 35]

3.1 The Physical Markets

In the Nord Pool area, there are three electricity markets active, all with different functions.

3.1.1 Elspot

Elspot is a physical power market, organized by Nord Pool, where energy is both sold and bought. Participation is free on the market, meaning that no producers are forced to sell on the market[1].

All participants who meet the requirements set by Nord Pool Spot are given access to the Elspot market. However, Elspot market participants must have a physical grid connection for power delivery or take-off in the area they want to trade in. Trading in Elspot requires signing a balance agreement with the transmission system operator responsible in the Elspot Area or areas with the physical grid connection.

On Elspot, hourly power contracts are traded daily for physical delivery in the next day's 24-hour period. At noon each day, bids for either purchase or sale,

[1] In some electricity markets, producers are forced to sell, in order to ensure enough supply at all times.[20]

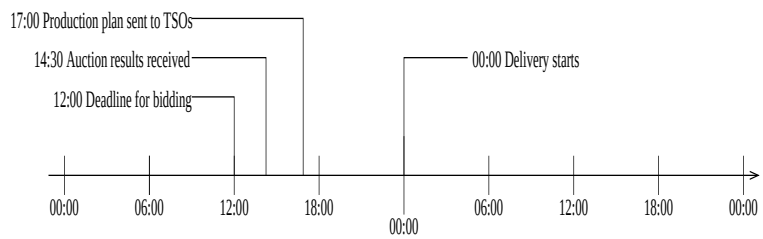

Figure 3.2: A time line showing the process of bidding, accepting and planning at Elspot

for each of these hour long contracts, are submitted (see time line in figure 3.2). Three different types of bids can be submitted.

Hourly bid: is the basic type of Elspot market order. Each bid can have up to 62 price intervals in addition to the current ceiling and floor price limits set by Nord Pool. Both energy price and quantity are specified in the bid.

Bids are said to be either price-independent bids or price dependent bids. In a price-independent bid, no price range are given apart from the ceiling and floor limits, and same amount is bought regardless of the price, see Table 3.1 for an example. In a price dependent bid ranges of prices and the corresponding volume are given. A participant that submits a price dependent bid, accepts that Nord Pool linearly interpolates volumes between each adjacent pair of submitted price steps. See Table 3.2 for an example.

Table 3.1: Example of a price-independent bid: 70 MW is bought each hour, regardless of the price

Hour/Price	0	2000
$1-24$	70	70

Table 3.2: Example of a price dependent bid: In hour 1, if the system price will be between EUR 0 and EUR 20, 50 MW will be bought in that hour. In hour 2, if the system price is EUR 23, $10+(25-10)\cdot(23-22.1)/(25-22.1)=14.7$ MW will be sold in that hour

Hour/Price	0	20	20.1	22	22.1	25	25.1	2000
1	50	50	0	0	−10	−10	−30	−30
2	50	50	0	0	−10	−25	−30	−30

Block bid: gives the participant the opportunity to set an "all or nothing" condition for all the hours within the block. A block bid can be made for any range of hours and it consists of n hourly bids, valid in n consecutive hours and must either be accepted entirely or rejected entirely; thus if accepted, the contract covers all hours and the volume specified in the bid. The Block Bid price is compared with the mean Elspot price for the hours to which the block period applies. An example of a block bid is given in table 3.3

Flexible hourly bid: is a sales bid, where price and volume are specified, but with no hour specification. The bid is accepted for the hour with the highest price which is lower than the bidding price and if no such hour exists, the bid is rejected.

This gives companies with intensive power consumption the opportunity to sell back energy by shutting down the industrial process in the hour when the bid is accepted.

Once all bids have been collected at noon, supply and demand curves for the whole Nordic area and for each hour are created from the bids. First, the supply and demand curves are constructed for the Nord Pool region as a whole with transmission capacities taken to be infinite. The intersection point of those curves then defines the system price for that hour as shown in Figure 3.3. Afterwards, all bidding areas are categorized as either sales surplus areas or sales deficit areas. The trading system checks whether transmission capacities are sufficient between the sales surplus areas and the sales deficit areas. If that turns out to be the case the system price will be valid in that hour for the whole Nordic region. However if that is not the case, the Nordic region is separated in to two different areas, the sales surplus area and the sales deficit area, and new equilibrium points in between the supply and demand curves are found for each area (see Figure 3.4). This will result in a lower area price for the sales surplus area than it is on the sales deficit area. Therefore the possibility

Table 3.3: Example of a block bid: If the calculated mean Elspot area price turns out to be EUR 22 in the hours $1-7$ and EUR 23 in the hours $8-17$, this participant will receive a contract for purchase of 200 MW in the hours 1 to 7 and a contract for sale of 50 MW in the hours 8 to 17.

Block hours	Price	Volume
Hour $1-7$	24	200
Hour $1-7$	20	50
Hour $8-17$	19	-50
Hour $8-17$	24	-100

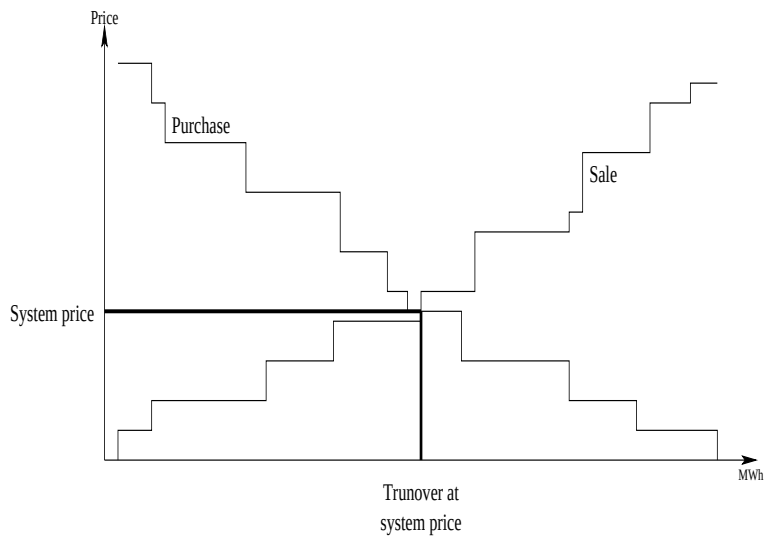

Figure 3.3: Calculation of the system price

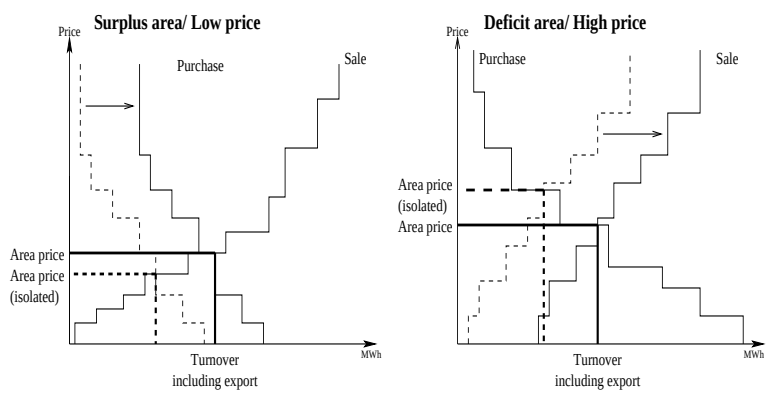

Figure 3.4: Calculation of area prices

of transmitting the transmission capacity between the areas can be considered, from the sales surplus area's point of view, as a price independent wish, from the sales deficit area, to purchase a volume equal to the transmission capacity. Correspondingly, the possibility of transmitting the transmission capacity between the areas can be considered, from the sales deficit area's point of view, as a price independent wish, from the sales surplus area, to sell a volume equal to the transmission capacity. This results in different area prices, lower on the sales surplus area and higher on the sales deficit area, giving balance between

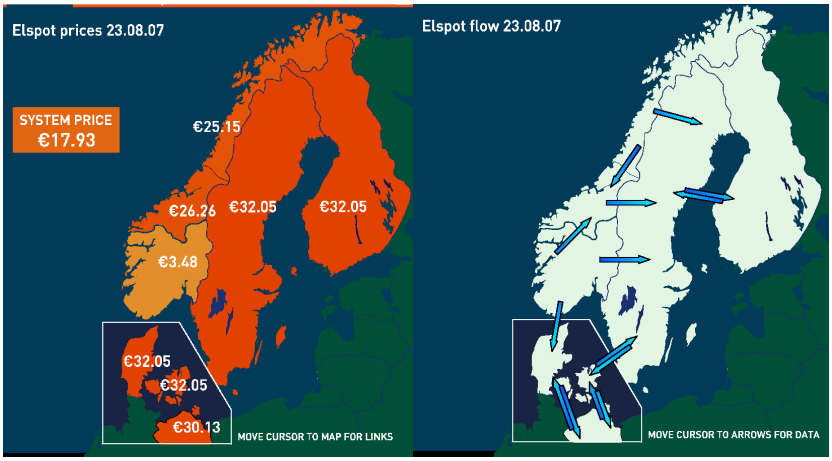

Figure 3.5: Area prices (left) and flow direction (right) of energy on 23.08.2007 between 10:00 and 11:00 in the Nord Pool region

total purchase and total sales in this hour. Furthermore, this results in total utilization of the trading capacity between the surplus and the deficit area. [20, 39, 40]

To get a better understanding of the price mechanism on the Nord Pool spot market, the following example can be viewed. Figure 3.5 shows the spot prices and whether energy is being transmitted into or out of each area between 10:00 and 11:00 on 23.08.2007.

The figure shows how power is only transmitted from areas where the price is lower than the price is in the area where the price is being transmitted to. When the system price was calculated for this particular hour, once equilibrium point between the supply and demand curve had been found, the transmission capacity from Norway 1 (the Southern most part of Norway) had not been sufficient. Therefore two prices were calculated, one for Norway 1 and one for the remaining areas. Then, there is still not sufficient transmission capacity between the other two areas in Norway and the other countries, so the same procedure is repeated twice, once for each area in Norway and therefore three different prices are found in Norway and another price in Denmark, Finland and Sweden. It then follows automatically that power is only transmitted from an area with lower price than the price in the receiving area.

3.1.2 Elbas

On Elspot, the time between bid and delivery is between 12-36 hours and during that period, both production and consumption schedules can deviate from the original plan. Consequently, market participants may find themselves in a situation where additional trading is necessary. This extra trading is done on the Elbas market. Elbas is a continuous power trading market, which is open at all times. Bids are made for individuals hours, gate closure is one hour before delivery and both hourly bids and block bids are accepted. Bids are prioritized by price (the lowest price has the highest priority) and if two bids have the same price, the bid which was received first is accepted first. Depending on areas, bids for the next day are opened at either 14:00 or 17:00 CET and therefore varies the upper limit on the time between trading and delivery from 8 to 32 hours. Since energy already traded on the Elspot market is higher prioritized than the energy traded on the Elbas market, transactions between areas where transmission capacities are already fully utilized are not allowed.

The Elbas market was founded by Nord Pool Finland Oy in 1999 and original members were Finland and Sweden. Eastern Denmark has been a member of the Elbas market since 2004 and in 2007, Western Denmark also joined. Finally, Norway is scheduled to join in the first half of 2008. In addition has Germany been a member of Elbas since 2006. [37, 38, 41]

3.1.3 The Regulating Market

When the bid by generators and consumers to the spot market are not fulfilled, the resulting imbalances must be leveled out in order to maintain equality between production and load in order to maintain power grid stability. This is where the regulating[2] market steps in. On the regulating market, bids for down and up regulation are stated $1 - 2$ hours before the production hour. Bids are of a quantity which can be delivered within 15 minutes notice. Settlement and pricing procedures on the regulating market differ between areas in Nord Pool and some efforts have been made in order to harmonize these procedures within the Nord Pool region. However the physical part of the market is identical in all areas. Despite the differences in the pricing procedures, the aim is always to ensure that prices reflect the production cost and to discourage planning of imbalances.

The need for regulating power depends on the original production plan, made after the collection of bids on Elspot, and the corrected version of the production plan. Since forcing the producers to keep their production plan 100% could compromise the grid stability, producers can request to alter their plan. In the

[2] Sometimes referred to as the real time market

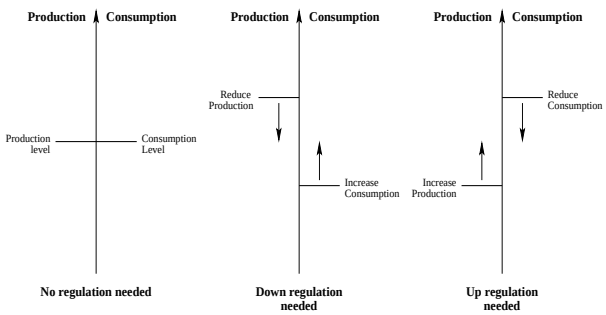

Figure 3.6: Possible regulation scenarios

event of producer's request to generate less power than agreed on, other producers have to regulate up, that is produce more energy, or consumption must be decreased in order to maintain system balance. Likewise if the same producer produces more than the bid states, other producers can be requested to reduce their production or consumption must be increased, see Figure 3.6. Same applies for consumers, if a consumer does not use the energy he has bought, down regulation is needed. If the consumer uses more energy, up regulation is needed.

Two examples of the physical function of the regulating market are shown in Figures 3.7 and 3.8. In the first example producer B has to change his production plan for some reason. The result will be an imbalance in the transmission grid. Therefore the TSO, which is responsible for the grid stability, accepts bids from the regulating market and thereby selects a producer to adjust his production so the system will be in balance despite producer B's deviation from the original plan. After the change, producer A has a changed plan while the producer who's regulation bid was accepted follows his original plan as well as selling regulating power. In the latter example, consumer B does not use the quantity, his bid states. Therefore the TSO accepts down regulation bid in order to keep the system in balance. Consequently, the producer who's bid is accepted reduces his generation and thereby, brings the system back in balance. The cost of this operation is carried by the consumer so practically, Consumer A has bought the quantity initially agreed on and then sold back the unused 20 MWh at a lower price.

The demand for regulating power, RD_t, at any given time, t is defined as the total difference between the original production plan and the updated production plan. Put mathematically,

$$RD_t = \sum_{p \in P} OP_{p,t} - UP_{p,t} \tag{3.1}$$

where P is the set of all producers in the system, $OP_{p,t}$ is the production origi-

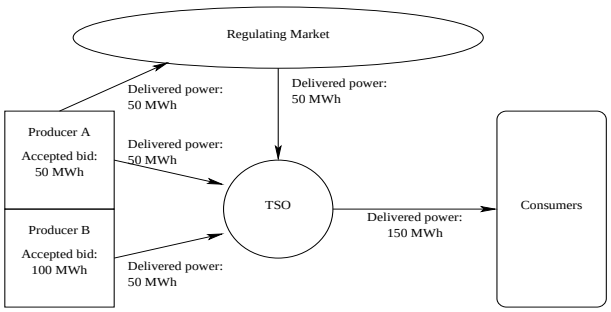

Figure 3.7: Demonstration of how the regulating market is used to correct imbalances caused by producers. Producer B only delivers 50 MWh of the 100 MWh he was supposed to. Therefore producer A is chosen to generate the missing 50 MWh since he has the best bid (lowest price) on the regulating market.

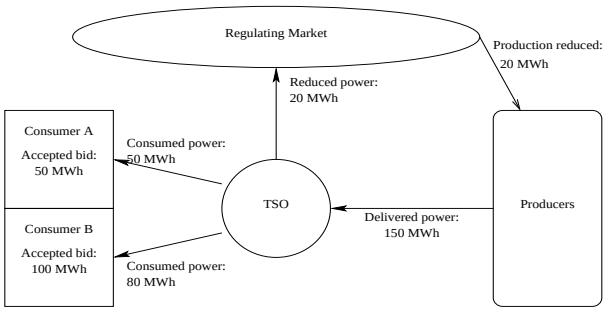

Figure 3.8: Demonstration of how the regulating market is used to correct imbalances caused by consumers. Consumer B only uses 80 MWh of the 100 MWh he was supposed to. The TSO therefore buys 20 MWh down regulation from the producers. The producer with the best bid (highest price) is selected to reduce his generation.

nally planned by producer p at time t and $UP_{p,t}$ is the updated production plan for producer p at time t. Hence, the possibility of there is no need for regulation even though the every producer has altered his production plan exists. The total demand for regulation power decides whether a producer is charged for his deviation. Producers are only charged for increasing the required regulating power, so if there is a need for up regulation and the producer is generating more energy than the plan states, the producer is bringing the system back towards balance and therefore he is not charged for regulation and receives the spot price for all his power.

Up regulation bids are made for a quantity which the producer can deliver within 15 minutes notice and the lowest price he is willing to receive for it.

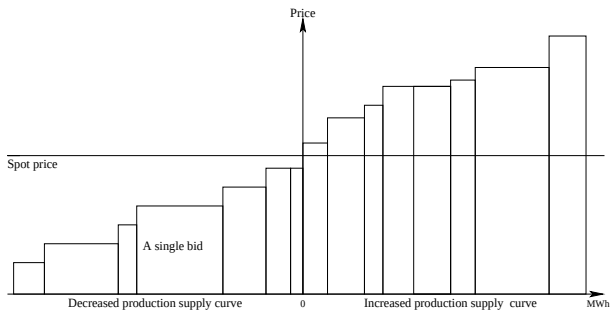

Figure 3.9: Bids on the regulation market are ordered so they form two different power curves, one for up regulation (right side of zero) and one for down regulation (left side of zero).

Down regulation bids are made for a quantity which the producer is ready to reduce his generation by and the payment he requests for stopping the production. The TSO accepts the bids, either for up or down regulation, depending on which is needed. When regulating up, the TSO accepts the bids in an increasing order and when regulating down, bids are accepted in an decreasing order. Put differently, if additional power is needed the most cost efficient production available is started and if generation reduction is needed, the least cost efficient generation is shut down. The two separate supply curves for the regulation power are shown in Figure 3.9. When regulation power is needed, bids are accepted for increased or decreased production with positive or negative sign respectively. So bids are accepted going from 0 and to the amount of power needed on the x-axis.[20, 27, 35, 42]

Prices at the regulating market are set in such a manner that nothing can be gained from being out of balance. In the two-price model used in Denmark, Finland and Sweden for pricing regulating power, each hour is either defined as an up regulating hour or a down regulating hour. Then prices are set as the price stated in the most expensive bid accepted in an up regulating hour and as the cheapest bid accepted in a down regulating hour. Regulating power is however never priced lower than the spot price in an up regulating hour and never priced higher than the spot price in a down regulation hour. So the regulation price for balance responsible parties having balance with opposite sign to the regulation demand is defined as

$$P_{r,o}(RD_t) = \begin{cases} \min\{P_{r,f}, P_{spot}\} & \text{if } RD_t < 0 \\ \max\{P_{r,f}, P_{spot}\} & \text{if } RD_t > 0 \end{cases} \tag{3.2}$$

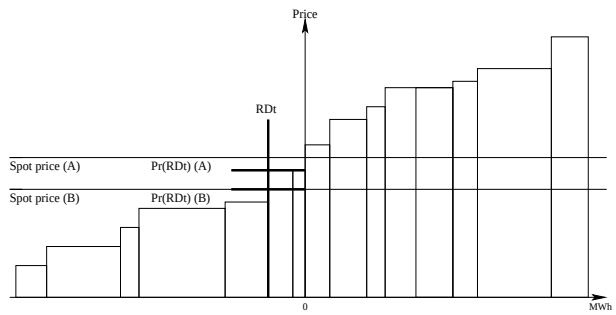

Figure 3.10: Determination of regulation prices in two different situations: A) Spot price is higher than the price of the bid accepted so $P_r(RD_t) = p^-$. B) Spot price is lower than the price of the bids accepted so $P_r(RD_t) = P_{spot}$

where $P_{r,f}$ is the free regulation price, defined as

$$P_{r,f} = \begin{cases} p^-(RD_t) & \text{if } RD_t < 0 \\ p^+(RD_t) & \text{if } RD_t > 0 \end{cases} \tag{3.3}$$

where p^- and p^+ are the prices from the down regulation supply curve and the up regulation supply curve respectively. However, since participants are not penalized for imbalances with the same sign as the regulation demand, the elspot price is used for balance responsible parties with positive balance during an up regulation hour or a negative balance during a down regulation hour, instead of the regulation price. So in the two price model, regulation prices are

$$P_r(RD_t) = \begin{cases} \begin{cases} P_{spot} & \text{if Balance} < 0 \\ P_{r,o} & \text{if Balance} > 0 \end{cases} & \text{if } RD_t < 0 \\ \\ \begin{cases} P_{spot} & \text{if Balance} > 0 \\ P_{r,o} & \text{if Balance} < 0 \end{cases} & \text{if } RD_t > 0 \end{cases} \tag{3.4}$$

In Norway a slightly different pricing model, a one price model, is used. Prices for regulation power are determined as the most expensive bid in an up regulation hour and as the cheapest bid in a down regulation hour as described in Equations 3.2 and 3.3. However, the price is valid for all regulating power, so only one price is defined for each hour, and the traders are responsible for balancing themselves. When no regulation longer than 10 minutes has been done, the Elspot price is used as the regulation power market price.[13, 42, 44]

3.2 The Financial Market

All financial contracts entered into at Nord Pool's financial market are cash settled and are therefore entered into without regard to technical conditions, such as grid congestion, access to capacity and other technical restrictions. Contract types currently traded on the market comprise of both power derivatives and electricity certificates. The derivatives are base load futures, forwards, options and Contracts for Difference, and the reference price for all of them is the system price of the total nordic market. Maximum trading time horizon is four years. Due to preferences of the market, trading time horizons for futures have been reduced over the past years from three years down to 8-9 weeks, while forwards are traded for the longer time horizons. These preferences are due to the daily mark-to-market settlements of the futures, which require a large amount of cash in pledged/non-pledged cash accounts. Financial settlement of forwards however involves no daily mark-to-market settlement and requires therefore only cash collateral during the delivery period, starting at the contracts maturity. In the following, a brief overview of the contract types traded on Nord Pool's Financial market is given.[33, 36]

3.2.1 Futures

Base load futures are traded on the Nord Pool financial market with a final settlement period of either 24 hours or 1 week. The contracts' ticker codes are written as

ENODXXXX-XX for day contracts with period of 24 hours

ENOWXX-XX for week contracts with period of 7 days

where E indicates the underlying commodity, electricity, and NO stands for the Nordic area.

Settlement of futures contracts involve both a daily mark-to-market settlement and a final spot reference cash settlement during the final settlement period which starts after the contract reaches it's due date. Mark-to-market settlement covers gains or losses from the day-to-day changes in the market price of each contract during the trading period of the contract. This means that throughout the trading period, a contract owner is debited/credited for the change in the closing market price between days. Then starting at the due date and throughout the delivery period, contract owners are debited/credited with the difference between the last closing price of the futures contract and the system price in the delivery period. [33, 36]

3.2.2 Forwards

Nord Pool's forward contract structure lists base load contracts for each calendar Month, Quarter and Year contracts:

ENOMmmm-yy: base load Month contract for month *mmm* and year *yy*

ENOQx-yy: base load Quarter contract for quarter *x* and year *yy*

ENOYR-yy: base load Year contract for year *yy*

Settlement of forward contracts involves no mark-to-market settlements and instead, the mark-to-market accumulated throughout the trading period. This means that during the delivery period, contract owners are debited/credited with the difference between the price of the contract when it was entered into and the system price in the delivery period. [33, 36]

3.2.3 Options

The options contracts at Nord Pool are European style with the underlying asset as season forward contracts. The ticker codes are written as

ENOtsssuu-yy: option of type *t* (Call/Put) with a strike price of *sss* and an underlying *uu* in year *yy*

[33, 36]

3.2.4 Contracts for Difference (CfD)

Since the reference price of for the forward and futures contracts is the Nord Pool Spot system price, the possibility of a perfect hedge exists only when there is no transmission grid congestion in the market area. That is, when the area price is equal to the system price. Hence, hedging in forwards or futures therefore implies a basis risk equal to the difference between the system price and the area price. In order to provide the possibility for a perfect hedge under the circumstances where the market is divided into more than one price areas, the Contracts for Difference (CfD) were introduced.

CfDs for the following area price differentials are offered by Nord Pool

Norway: Difference between Oslo area price and the System Price.

Sweden: Difference between Stockholm area price and the System Price.

Finland: Difference between Helsinki area price and the System Price.

Denmark West: Difference between Aarhus area price and the System Price.

Denmark East: Difference between Copenhagen area price and the System Price.

Germany: Difference between Germany area price and the System Price.

A CfD is a forward contract with reference to the difference between the area price and the Nord Pool Spot system price and therefore does the market price of a CfD reflect the market's prediction of the price difference during the delivery period. Since the nature of the area prices is such that if the price is higher than the system price in one area, it causes without exception the area price in some other area to be lower than the system price, the market price of a CfD can both be positive and negative. CfDs for a specific area trade at positive prices when expectations are that area price will be higher than the system price in that area during the delivery period. Correspondingly, CfDs for a specific area trade at negative prices if it is anticipated that area price will be lower than the system price in that area during the delivery period. [33, 36]

CHAPTER 4

The Mathematical Toolbox

Here follows a brief and general description of the mathematical tools used to carry out the analysis and the modeling, described in the chapters to come. The description is only meant to be superficial and therefore should references be looked up for proofs and more detailed descriptions.

4.1 Random Variables and Processes

A variable, who's value can not be stated with certainty before hand, at any time point is said to be a random variable. For example, if the variable D is defined as the number on the upwards turning face of a dice that is thrown, the value is not known before the dice is thrown and is therefore a random variable. Furthermore, when the value of the random variable is ever changing over time, the variable is said to be a random or a stochastic process.

A complete description of random variables, both continuous and discrete, is given in [14]. Here follows a short summary of the main tools to describe and evaluate the properties of random variables and processes.

4.1.1 Probability

Behind every random variable is a probability distribution, describing the probability of the variable taking a certain value or a range of values. A *probability density function*, PDF, is any function $f(x)$ that describes the probability density in terms of the input variable x. For any valid PDF, $f(x)$, the following equations hold

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)dx &= 1 \\ f(x) &\geq 0 \end{aligned} \tag{4.1}$$

The *cumulative distribution function*, CDF, completely describes the probability distribution of a random variable, X. The CDF is defined, in terms of the probability density function, as

$$F_X(a) = \mathbb{P}(X \leq a) = \int_{-\infty}^{a} f(x)dx \tag{4.2}$$

and describes the probability, $\mathbb{P}$, of X taking the real value less than or equal to a. The probability of X lying in the interval $(a,b), a < b$ is therefore found as $F(b) - F(a)$.

If the CDF of the random variable X is continuous, then X is a continuous random variable. Furthermore, if F is absolutely continuous, there exists a PDF such that

$$F_X(b) - F_X(a) = \mathbb{P}(a \leq X \leq b) = \int_{a}^{b} f(x)dx \tag{4.3}$$

Finally, for all CDFs,

$$\begin{aligned} \lim_{x \to -\infty} F(x) &= 0 \\ \lim_{x \to \infty} F(x) &= 1 \end{aligned} \tag{4.4}$$

For a discrete random variable, the probability that the variable takes exactly some value is given by the *probability mass function*, PMF. For a discrete random variable X, defined on the set $H = \{x_1, \ldots, x_N\}$, and described by the probability mass function $f(X)$, the following must hold for the function to be valid.

$$\begin{aligned} \sum_{i=1}^{N} f(x_i) &= 1 \\ f(x_i) &> 0 \quad \forall x_i \in H \end{aligned} \tag{4.5}$$

Therefore

$$\mathbb{P}(X = a) = f(a) \tag{4.6}$$

and equation (4.2) becomes

$$F_X(x_i \leq a) = \sum_{j=1}^{i} f(x_j) \tag{4.7}$$

4.1.2 Expectation and Higher Order Moments

For a random variable X, the mean outcome is called the expectation and is calculated as

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx \tag{4.8}$$

if X is continuous with a PDF $f(x)$. If X is discrete, defined on the set $H = \{x_1, \ldots, x_N\}$, the expectation is given by

$$\mathbb{E}(X) = \sum_{i=1}^{N} x_i \mathbb{P}(X = x_i) \tag{4.9}$$

The expectation is sometimes referred to as the first moment of X. Generally, the kth moment, m_k is defined as

$$m_k = \mathbb{E}(X^k) \tag{4.10}$$

where k is a positive integer. From the first moment, the central moment[1], μ_k is defined as

$$\mu_k = \mathbb{E}((X - m_i)^k) \tag{4.11}$$

The two most commonly used moments are m_1, describing the mean of X and μ_2 which is called the variance of X and measures the dispersion of X or the amount by which X tends to deviate from its average. The variance is often denoted as σ^2 and another important parameter derived from the variance is the standard deviation, $\sigma = \sqrt{\sigma^2}$.

Higher order central moments are often normalized with respect to the standard deviation and thereby dimensionless quantities, which represent the distribution independently of any linear change of scale are obtained. Two important measures of the distribution properties of X are the skewness, γ, and

[1] Also often referred to as the moments around the mean

kurtosis, κ of X, which are the 3rd and the 4th standardized moments respectively and are defined as

$$\begin{aligned} \gamma &= \frac{\mu_3}{\sigma^3} \\ \kappa &= \frac{\mu_4}{\sigma^4} \end{aligned} \tag{4.12}$$

The Skewness is a measure of asymmetry in the probability distribution. It has a positive sign and is said to be skewed to the right if the distribution has a longer tail in the positive direction from the mean. Kurtosis on the other hand is a measure of "peakedness" of the distribution. Higher kurtosis means that more of the variance is due to infrequent extreme deviations, as opposed to frequent modestly-sized deviations. Therefore, a distribution with a high kurtosis has a sharper peak and flatter tail than a distribution with a low kurtosis. Put in context with the well known standard normal distribution, that distribution has a skewness of 0 and kurtosis 3.[14, 53]

4.2 ARMA Models

The time varying process

$$Y_t + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} \tag{4.13}$$

where ε_t is white noise is called an $ARMA(p,q)$-process. The process is commonly written as

$$\phi(B) Y_t = \theta(B) \varepsilon_t \tag{4.14}$$

where $\phi(B)$ and $\theta(B)$ are polynomials of order p and q, respectively, and backward shift operator B is defined by

$$Bx_t = x_{t-1}, \quad B^j x_t = x_{t-j} \tag{4.15}$$

A class of models describing processes with periodic behavior has been proposed as well. The idea is to let Y_t from a time series with seasonal period s depend on $Y_{t-s}, Y_{t-2s}, \ldots$. If the process Y_t can be modeled by

$$\phi(B)\Phi(B^s) Y_t = \theta(B)\Theta(B^s)\varepsilon_t \tag{4.16}$$

it is said to follow a multiplicative $ARMA(p,q) \times (P,Q)_s$ seasonal model. If more than one periods can be detected in the model, the second seasonal behavior is added to the model in the same manner.

The $ARMA(p,q)$ process can be written on MA-form as $Y_t = \phi^{-1}(B)\theta(B)\varepsilon_t = \psi(B)\varepsilon_t$. From this the variance of the prediction error for the k-step ahead prediction can be derived as

$$\sigma_k^2 = (1 + \psi_1^2 + \ldots + \psi_{k-1}^2)\sigma_\varepsilon^2 \tag{4.17}$$

Assuming that ε_t is normally distributed, the $(1-\alpha)$-confidence interval for Y_{t+k} is found as

$$\widehat{Y}_{t+k} \pm u_{\alpha/2}\sigma_\varepsilon\sqrt{1 + \psi_1^2 + \ldots + \psi_{k-1}^2} \tag{4.18}$$

where $u_{\alpha/2}$ is the $\alpha/2$ fractile in the standard nomal distribution.

The model given by (4.13) can be written into a linear regression model as

$$Y_t = \mathbf{X}_t^T\boldsymbol{\beta} + \varepsilon_t \quad t = 1,\ldots,N \tag{4.19}$$

where

$$\begin{aligned} \mathbf{X}_t^T &= \left[Y_{t-1},\ldots,Y_{t-p},Y_{t-s},\ldots,Y_{t-Ps},\varepsilon_{t-1},\ldots,\varepsilon_{t-q},\varepsilon_{t-s},\ldots,\varepsilon_{t-Qs}\right] \\ \boldsymbol{\beta}^T &= \left[\phi_1,\ldots,\phi_p,\Phi_1,\ldots,\Phi_P,\theta_1,\ldots,\theta_q,\Theta_1,\ldots,\Theta_Q\right] \end{aligned} \tag{4.20}$$

A least squares estimation of the parameters in $\boldsymbol{\beta}$ is found as the solution to

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} S(\boldsymbol{\beta}) \tag{4.21}$$

where

$$S(\boldsymbol{\beta}) = \sum_{s=1}^{N}\left(Y_s - \mathbf{X}_s^T\boldsymbol{\beta}\right)^2 \tag{4.22}$$

By introducing

$$\mathbf{Y}^T = [Y_1,\ldots,Y_N], \tag{4.23}$$

$$\mathbf{X}^T = [\mathbf{X}_1,\ldots,\mathbf{X}_N], \tag{4.24}$$

$$\boldsymbol{\varepsilon} = [\varepsilon_1,\ldots,\varepsilon_N] \tag{4.25}$$

The model model (4.13) can be written in a compact linear form for all the observations as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{4.26}$$

and thereby Equation (4.22) can be rewritten as

$$S(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \tag{4.27}$$

to which the solution can be found as

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^\mathbf{T}\mathbf{X}\right)^{-1}\mathbf{X}^\mathbf{T}\mathbf{Y} \tag{4.28}$$

[22]

4.3 Recursive Least Squares Models With External Signals

Consider a process that can be described as a linear system in discrete time as

$$Y_t + \phi_1 Y_{t-1} + \ldots \phi_k Y_{t-k} = \omega_1 U_{t-1} + \ldots + \omega_m U_{t-m} + \varepsilon_t \tag{4.29}$$

where Y_t is the spot price (response), ε_t is white noise and U_t is an uncorrelated input signal. By introducing the vectors

$$\begin{aligned} \mathbf{X}_t &= [-Y_{t-1}, \ldots, -Y_{t-k}, U_{t-1}, \ldots, U_{t-m}]^T \\ \boldsymbol{\theta} &= [\phi_1, \ldots, \phi_k, \omega_1, \ldots, \omega_m]^T \end{aligned} \tag{4.30}$$

the system can be described with a linear regression model similar to (4.19) and therefore can the parameters in $\boldsymbol{\theta}_t$ be found with the same procedure as described in Equations (4.21)-(4.28).

Since this model is static, it does not account for any changes or seasonal variation that happen in the long run. One way of dealing with such effects is to estimate the parameters recursively so that $\boldsymbol{\theta}$ become time dependent ($\boldsymbol{\theta}_t$) and adapt to the relatively slow changes in the system. The recursive estimate of the parameters in $\boldsymbol{\theta}_t$ is found as the solution to

$$\widehat{\boldsymbol{\theta}}_t = \arg\min_{\theta} S_t(\boldsymbol{\theta}) \tag{4.31}$$

where

$$S_t(\boldsymbol{\theta}) = \sum_{s=1}^{t} \left(Y_s - \mathbf{X}_t^T \boldsymbol{\theta}_t \right)^2 \tag{4.32}$$

The *off-line* solution to this equation is found as

$$\widehat{\boldsymbol{\theta}}_t = \mathbf{R}_t^{-1} \mathbf{h}_t \tag{4.33}$$

where

$$\begin{aligned} \mathbf{R}_t &= \sum_{s=1}^{t} \mathbf{X}_s \mathbf{X}_s^T \\ \mathbf{h}_t &= \sum_{s=1}^{t} \mathbf{X}_s Y_s \end{aligned} \tag{4.34}$$

and from this the recursive updating formulas for $\mathbf{R_t}$ and $\mathbf{h}_t$ are easily derived

$$\begin{aligned} \mathbf{R}_t &= \sum_{s=1}^{t-1} \mathbf{X}_s \mathbf{X}_s^T + \mathbf{X}_t \mathbf{X}_t^T = \mathbf{R}_{t-1} + \mathbf{X}_t \mathbf{X}_t^T \\ \mathbf{h}_t &= \sum_{s=1}^{t-1} \mathbf{X}_s Y_s + \mathbf{X}_t Y_t = \mathbf{h}_{t-1} + \mathbf{X}_t Y_t \end{aligned} \tag{4.35}$$

Thereby, the updating formula for $\widehat{\boldsymbol{\theta}}$ is found to be

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} + \mathbf{R}_t^{-1}\mathbf{X}_t \left[Y_t - \mathbf{X}_t^T \widehat{\boldsymbol{\theta}}_{t-1} \right] \tag{4.36}$$

To avoid the expensive calculations of matrix inversion, the matrix $\mathbf{P}_t = \mathbf{R}_t^{-1}$ is introduced and by the matrix inversion lemma

$$[\mathbf{A} + \mathbf{BCD}]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left[\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}\right]^{-1}\mathbf{DA}^{-1} \tag{4.37}$$

an updating formula for $\mathbf{P}_t$ is found to be

$$\mathbf{P}_t = \mathbf{P}_{t-1} - \frac{\mathbf{P}_{t-1}\mathbf{X}_t\mathbf{X}_t^T\mathbf{P}_{t-1}}{1 + \mathbf{X}_t^T\mathbf{P}_{t-1}\mathbf{X}_t} \tag{4.38}$$

Finally the gain matrix, $\mathbf{K}_t$ is introduced as

$$\mathbf{K}_t = \mathbf{R}_t^{-1}\mathbf{X}_t = \frac{\mathbf{P}_{t-1}\mathbf{X}_t}{1 + \mathbf{X}_t^T\mathbf{P}_{t-1}\mathbf{X}_t} \tag{4.39}$$

Thus, the *RLS* algorithm can now be written as

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_t &= \widehat{\boldsymbol{\theta}}_{t-1} + \mathbf{K}_t \left[Y_t - \mathbf{X}_t^T \widehat{\boldsymbol{\theta}}_{t-1} \right] \\ \mathbf{K}_t &= \frac{\mathbf{P}_{t-1}\mathbf{X}_t}{1 + \mathbf{X}_t^T\mathbf{P}_{t-1}X_t} \\ \mathbf{P}_t &= \mathbf{P}_{t-1} - \frac{\mathbf{P}_{t-1}\mathbf{X}_t\mathbf{X}_t^T\mathbf{P}_{t-1}}{1 + \mathbf{X}_t^T\mathbf{P}_{t-1}\mathbf{X}_t} \end{aligned} \tag{4.40}$$

Although the *RLS* algorithm provides parameters that adapt to the changes in the system, it still assumes that the system behaves linearly on the time interval from $t = 0$ to $t = t$ and provides parameter estimates that are consistent in time and converge to a point in the parameter space. [22, 23]

4.4 Recursive Least Squares with forgetting factor

In many cases, it is desirable to have time varying parameters instead of the time consistent parameters provided by the *RLS* algorithm. The forgetting factor or exponential forgetting technique is a simple extension to the *RLS* algorithm and handles time varying parameters by discounting old prediction errors in the loss function, $S(\boldsymbol{\theta})$. In other words, the prediction error arisen from the most recent observations weights more than the older ones, when estimating the parameters. Although the change to the algorithm seems mall, this

often leads to drastic changes to the properties of the algorithm. The parameter estimates do not converge anymore and instead, the parameter sequences can be described as a stochastic process. This process is non-Gaussian but the parameter estimation error will still have a mean of 0.

The algorithm for *RLS* with exponential forgetting still involves finding a solution to the weighted least squares estimator

$$\widehat{\boldsymbol{\theta}}_t = \arg\min_{\theta} S_t(\boldsymbol{\theta}) \tag{4.41}$$

but now a weight function has been added to the loss function so

$$S_t(\boldsymbol{\theta}) = \sum_{s=1}^{t} \beta(t,s) \left(Y_s - \mathbf{X}_s^T \boldsymbol{\theta}\right)^2 \tag{4.42}$$

where it is assumed that the sequence of weights satisfies

$$\begin{aligned} &\beta(t,s) = \lambda(t)\beta(t-1,s) \quad 1 \leq s \leq t-1 \\ &\beta(t,t) = 1 \end{aligned} \tag{4.43}$$

which means that

$$\beta(t,s) = \prod_{j=s+1}^{t} \lambda(j) \tag{4.44}$$

The weight of the squared residual at time s in the computation of the parameter estimates at time t is the product of all intermediate weighting factors.

The solution for $\widehat{\boldsymbol{\theta}}_t$ is still found by Equation (4.33), but now

$$\begin{aligned} \mathbf{R}_t &= \sum_{s=1}^{t} \beta(t,s) \mathbf{X}_s \mathbf{X}_s^T \\ \mathbf{h}_t &= \sum_{s=1}^{t} \beta(t,s) \mathbf{X}_s Y_s \end{aligned} \tag{4.45}$$

so the updating formulas become

$$\begin{aligned} \mathbf{R}_t &= \lambda(t)\mathbf{R}_{t-1} + \mathbf{X}_t \mathbf{X}_t^T \\ \mathbf{h}_t &= \lambda(t)\mathbf{h}_{t-1} + \mathbf{X}_t Y_t \end{aligned} \tag{4.46}$$

From this a revised algorithm for *RLS* with a forgetting structure is derived

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_t &= \widehat{\boldsymbol{\theta}}_{t-1} + \mathbf{K}_t \left[Y_t - \mathbf{X}_t^T \widehat{\boldsymbol{\theta}}_{t-1}\right] \\ \mathbf{K}_t &= \frac{\mathbf{P}_{t-1} X_t}{1 + \mathbf{X}_t^T \mathbf{P}_{t-1} X_t} \\ \mathbf{P}_t &= \frac{1}{\lambda(t)} \left[\mathbf{P}_{t-1} - \frac{\mathbf{P}_{t-1} \mathbf{X}_t \mathbf{X}_t^T \mathbf{P}_{t-1}}{\lambda(t) + \mathbf{X}_t^T \mathbf{P}_{t-1} \mathbf{X}_t}\right] \end{aligned} \tag{4.47}$$

For $\lambda(t) = \lambda$, $\beta(t,s) = \lambda^{t-s}$ and if $0 < \lambda < 1$ then λ is called the forgetting factor. The squared errors are weighted exponentially and therefore the number of effective observations, T_0 can be found as

$$T_0 = \frac{1}{1-\lambda} \tag{4.48}$$

From this can be seen that forgetting factor $\lambda = 1$ is the same as not including a forgetting factor.[22, 23]

4.5 Locally Weighted Polynomial regression

Estimating a complex non-linear relationship between variables can be a difficult problem. However if the relationship is locally linear or described by a low order polynomial the problem is much more convenient to deal with. One would simply solve the weighted least squares problem

$$\arg\min_{\theta} \frac{1}{N} \sum_{s=1}^{N} w_s(\mathbf{x})(Y_s - \mathbf{X}_s^T \theta(\mathbf{x}))^2 \tag{4.49}$$

where w is a weight, that often is given by a kernel function, which weights the observations on the pre-specified bandwidth.

One way of estimating non-linear relationship between variables is to use Locally Weighted Polynomial Regression (LWPR)[23]. By assuming the relationship to be of low order on a small interval one can easily find the relationship on that interval by regression. The size of the interval or the bandwidth, h $(0 \leq h \leq 1)$, is given as the proportion of total observations used to estimate the relationship in each fitting point. In other words when a bandwidth of 0.2 is chosen it means that the 20% of the observations closest to the fitting point are taken into consideration when the relationship between the variables is estimated around that point. The bandwidth is to be chosen as one that is assumed appropriate for the problem at hand. For the models constructed in the thesis a bandwidth of $h = 0.3$ is used. This choice is made since it was desired to estimate the transfer function as locally as possible and $h = 0.3$ was the smallest bandwidth which did not result in singular matrices at any time. It ought to be mentioned that in order to get a good estimation from LWPR, fairly large data set has to be available. Fortunately this is the case here and therefore LWPR is used to estimate various relationship between variables in the model. In those analysis, a tri-cube kernel was used as a weight function i.e.

$$w(u) = \begin{cases} (1-u^3)^3 & u \in [0;1) \\ 0 & u \in [1;\infty) \end{cases} \tag{4.50}$$

where u is the relative distance between the point, which the relationship is to be estimated around and other points within the bandwidth.[22, 23]

4.6 k-step ahead prediction

In order to use the RLS-algorithm to predict more than one step ahead in time, the pseudo prediction error is used and it is defined as

$$\tilde{\varepsilon}^{pseudo}_{t+k|t} = Y_{t+k} - \mathbf{X}_t^T \widehat{\boldsymbol{\theta}}_{t+k-1} \tag{4.51}$$

And the k-step prediction is calculated as

$$\widehat{Y}_{t+k|t} = \mathbf{X}_t^T \widehat{\boldsymbol{\theta}}_t \tag{4.52}$$

So predictions are not made for 1 hour at a time and repeated k times, but one prediction is made for k time steps ahead. Therefore different parameters are found for every k.

4.7 Performance Estimation for Point Forecasting Models

Numerous ways exist to compare the performance of different models as described in [24]. Two of them are used here to compare the different models, the coefficient of determination (R^2) and the root mean squared error ($RMSE$). R^2 is a measurement of how much of the variability in the data is accounted for in the model and is calculated as

$$R^2 = 1 - \frac{SS_E}{SS_T} \tag{4.53}$$

where SS_T is the total sum of squares, and SS_E is the sum of squared residuals

$$\begin{aligned} SS_T &= \sum_{i=1}^{N} (y_i - \bar{y})^2 \\ SS_E &= \sum_{i=1}^{N} (y_i - \widehat{y})^2 \end{aligned} \tag{4.54}$$

$RMSE$ is a measure of how much the model deviates from the real values on average and therefore it decreases if a term is added to the model that provides

the estimator with information that otherwise would not be accounted for in the prediction. $RMSE$ is calculated as

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(Y_t - \widehat{Y}_t\right)^2} \tag{4.55}$$

4.8 Quantile Regression

The conventional way to estimate forecasting uncertainty directly from the point forecasting models presented is to make some assumptions about the distribution of the forecasting errors and estimate the expected uncertainty accordingly. However, for complicated processes, the distribution of the forecasting errors can both change depending on values of external factors and also vary over time. Therefore, it can be beneficial to model the uncertainty separately and without assuming any distribution properties, and thereby gain information about the distribution of the possible outcomes of the process. One method for doing such modeling is quantile regression.

Quantiles are a statistical measure, that quantifies a data set. The quantile lines separate the the data set in such a way that the number of observations below the line corresponds to a particular ratio which determines the quantiles. Quantiles are a general term for this separation, but for convenience, special commonly used cases have been given special names. Quartiles separate the data set into four areas, each containing equal share of the total observations and percentiles separates the observation space into 100 spaces, with an equal amount of observations within each segment. The nominal proportion of quantiles is often denoted as $\tau \in [0,1]$, so if the τth quantile is to be found, then $100\tau\%$ of the observations should have a lower value than the τ quantile.

In principle, the idea behind quantile regression is quite simple. When ordinary least squares regression is performed, the squared residuals of the resulting model are minimized and thereby a "best" line through the data points is obtained. Similarly in quantile regression, a "best" line is sought, only here the criteria is that a certain proportion of the observations are on each side of the line. This is done by introducing a piecewise linear and asymmetric loss function that penalizes the deviations from the line depending on the which side of the line they are placed. Minimizing the total penalty is then the same as minimizing the total error for the given quantile.

Let $\mathbf{Y}$ be a random variable with CDF F_Y and let τ be a real number, $0 \leq \tau \leq 1$. Then the τth quantile of F_Y is

$$q_Y(\tau) = F_Y^{-1}(\tau) = \inf\{y : F_Y(y) \geq \tau\} \tag{4.56}$$

i.e. $100\tau\%$ of the probability mass of Y is below $q_Y(\tau)$. This translates to the τ quantile, $q_Y(\tau)$, for Y being a function for which

$$F(q_Y(\tau)) = \mathbb{P}(Y < q_Y(\tau)) = \tau \tag{4.57}$$

Now assuming there is given a data set $(\mathbf{Y}, \mathbf{X})$ of N observations, where $\mathbf{Y}$ is $N \times 1$ and $\mathbf{X}$ is $N \times k$, from which the linear model

$$y_t = \widehat{Q}(\tau, x_t) + \varepsilon_t = \mathbf{x}_t \boldsymbol{\beta} + \varepsilon_t \quad t = 1, \ldots, N \tag{4.58}$$

has been constructed for the particular quantile of y_t. The piecewise linear and asymmetric penalty or loss function is defined as

$$\rho_\tau(\varepsilon) = \begin{cases} \tau\varepsilon & \text{if } \varepsilon \geq 0 \\ (\tau - 1)\varepsilon & \text{if } \varepsilon < 0 \end{cases} \tag{4.59}$$

And thereby the best τth quantile regression estimator of β can be found by minimizing the objective function

$$V_N(\beta; \tau) = \frac{1}{N} \sum_{t=1}^{N} \rho_\tau(y_t - \mathbf{x}_t \boldsymbol{\beta}) \tag{4.60}$$

Put differently

$$\widehat{\beta}_\tau = \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{N} \rho_\tau(\varepsilon_t) \tag{4.61}$$

In real life applications, completely linear processes are rare. It is therefore desirable to allow for some non-linearities in the model. One way of doing that is to consider $Q(x, \tau)$ as an additive model with splines. Splines are polynomials of degree m on intervals defined by a knot sequence and are $m - 1$ times differentiable over the knots. For each interval between two knots, a m order polynomial that minimizes the bending energy on the path between the knots is found. Several different classes of splines exist, all suited for different purposes. The distinguishing factor for basic splines is that their sum is 1 in the range they are defined in. Natural splines must have a constant 1st derivative outside the boundary knots and periodic splines can be joined continuously at the lower and upper boundary with matching derivatives.[9, 17]

When applying quantile regression to time series, in many cases some adaptivity is desired. For doing that, the optimization problem can be transformed into a linear programming problem and to which the simplex algorithm or the interior point method can be applied in every updating step. A forgetting factor of some kind is then used, such that only recent observations are included in the model estimation at every time. [16, 20, 21]

4.9 Quality Assessment of Probabilistic Forecasts

Since the forecasts made by quantile regression are no longer forecast of the mean in a distribution, methods for evaluating the quality of these forecasts differ fundamentally from the ones described in Section 4.7. A thorough description of the methods used in this thesis is given in [46] and will be introduced shortly in the following.

Forecasts obtained by non-parametric probabilistic methods are made for individual quantiles for which the nominal proportion in known. Consequently, *reliability* of probabilistic forecasts, is evaluated by assessing the reliability of each individual quantile prediction. Reliability refers to how much the predicted quantiles differ from the nominal quantile they represent and is typically presented in a so called reliability diagrams. To estimate the reliability of an individual quantile forecast, first an indicator variable, $\xi^{\tau}_{t,k}$ is introduced and defined as

$$\xi^{\tau}_{t,k} = \begin{cases} 1 & \text{if } y_{t+k} < \widehat{q}^{\tau}_{t+k|t} \\ 0 & \text{otherwise} \end{cases} \tag{4.62}$$

Then an estimate of the actual coverage is obtained by

$$\widehat{a}^{\tau}_{k} = \mathbb{E}[\xi^{\tau}_{t,k}] = \frac{1}{N}\sum_{i=1}^{N} \xi^{\tau}_{t,k} \tag{4.63}$$

and thereby the deviation from perfect reliability or the probabilistic bias, r^{τ}_{k} is found as

$$r^{\tau}_{k} = \tau_k - \widehat{a}^{\tau}_{k} \tag{4.64}$$

From these measures of empirical coverage, the reliability diagrams are constructed. Reliability diagrams are commonly given as a plot of the observed probabilities against the nominal ones for the nominal proportions estimated, making a perfect reliability appear as a line with a slope of one. Alternatively, reliability diagrams can give the deviation from the perfect reliability, in which case a perfect reliability will result in a straight line of zeros.

Another quality measure is the *sharpness*. Sharpness indicates how large the distance between two different quantiles is and therefore explains how sharp or flat the distribution is. Since sharpness is a measure of distance between two points, it can only be calculated for pairs of quantiles. Now for the central forecasting interval with nominal coverage rate $(1-\beta)$, the sharpness of an individual pair of quantiles is defined as

$$\delta^{\beta}_{t,k} = \widehat{q}^{(1-\beta/2)}_{t+k|t} - \widehat{q}^{\beta/2}_{t,k} \tag{4.65}$$

Then the overall measure of sharpness for this forecasting interval, for forecasting horizon k, is calculated as

$$\bar{\delta}_k^{\beta} = \frac{1}{N}\sum_{t=1}^{N}\delta_{t,k}^{\beta} = \frac{1}{N}\sum_{t=1}^{N}\left(\widehat{q}_{t+k|t}^{(1-\beta/2)} - \widehat{q}_{t,k}^{\beta/2}\right) \tag{4.66}$$

For evaluating the quality of a model, it can be very useful to define a unique skill score for the model. This makes comparison of different models much easier and provides a good overview of the model's quality. In [46], one such score is proposed and it will be used in this thesis. The unique skill score is defined as

$$Sc(\widehat{q}^{\tau}, y) = \sum_{i=1}^{m}(\xi^{\tau_i} - \tau_i)(y - \widehat{q}^{\tau_i}) \tag{4.67}$$

where m is the number of quantiles in the model. As can be seen from the definition of the skill score, it can be used both for assessing the quality of a single quantile model as well as for evaluating a number of quantiles. The skill score is a positively rewarding score, meaning that a higher score value indicates a higher skill of the model. Furthermore, the score is negative, so the score 0 would be awarded to the perfect forecast.

All these measures of performance can be calculated as an average of all forecasting horizons. However, since it is known that generally uncertainty is significantly influenced by the forecasting horizons, it is commonly accepted that a specific uncertainty model should be set up for each and every look-ahead time. Therefore, the quality of such models should also be evaluated for each horizon. Thereby do these measures also provide information about for which horizons the model is suitable - Information that is important as different models can be appropriate for modeling the same process on different horizons.

4.10 Classification Models

Classification of data is one of the oldest problem in statistics and has therefore been the subject of countless researches. Consequently, numerous methods for statistical classification exists. In this thesis, two popular classification methods are applied, all well described in [17].

4.10.1 Logistic Regression

The logistic regression model describes the posterior probabilities of K classes via linear function in x. At the same time it ensures that the probabilities sum

to one and remain in $[0,1]$. The logistic regression model for K classes has the form

$$\begin{aligned}
\ln \frac{\mathbb{P}(G=1|X=x)}{\mathbb{P}(G=K|X=x)} &= \beta_{10} + \beta_1^T x \\
\ln \frac{\mathbb{P}(G=2|X=x)}{\mathbb{P}(G=K|X=x)} &= \beta_{20} + \beta_2^T x \\
&\vdots \\
\ln \frac{\mathbb{P}(G=K-1|X=x)}{\mathbb{P}(G=K|X=x)} &= \beta_{(K-1)0} + \beta_{K-1}^T x
\end{aligned} \tag{4.68}$$

The model is specified in terms of $K-1$ log probabilities or logit transformations and simple calculation shows that

$$\begin{aligned}
\mathbb{P}(G=k|X=x) &= \frac{\exp(\beta_{k0} + \beta_k^T x)}{1 + \sum_{l=1}^{K-1} \exp(\beta_{l0} + \beta_l^T x)} \quad k = 1, \ldots, K-1 \\
\mathbb{P}(G=K|X=x) &= \frac{1}{1 + \sum_{l=1}^{K-1} \exp(\beta_{l0} + \beta_l^T x)}
\end{aligned} \tag{4.69}$$

From this, it is clear that the probabilities sum to one and furthermore, they are dependent on the parameters β. It is therefore common to denote the probabilities as $p_k(x;\beta) = \mathbb{P}(G=k|X=x)$.

In practice, logistic regression models are usually fitted by maximum likelihood, using the conditional distribution of $G|X$. $\mathbb{P}(G|X)$ completely specifies the conditional distribution, and therefore the multinomial distribution is appropriate in the general case. However, since the two-class case is only relevant in this thesis, fitting is only viewed in detail for that case.

The multinomial distribution now becomes a binomial distribution and by coding the two classes g_i via a 0/1 response y_i, where $y_i = 1$ when $g_i = 1$ and $y_i = 0$ when $g_i = 2$ and letting $p_1(x;\beta) = p(x;\beta)$ and $p_2(x;\beta) = 1 - p(x;\beta)$, the log-likelihood for N observations can be written as

$$\begin{aligned}
l(\beta) &= \sum_{i=1}^{N} (y_i \ln p(x_i;\beta) + (1-y_i)\ln(1-p(x_i;\beta))) \\
&= \sum_{i=1}^{N} \left(y_i \beta^T x_i - \ln(1 + e^{\beta^T x_i}) \right)
\end{aligned} \tag{4.70}$$

where $\beta = [\beta_{10}, \beta_1]$. The maximum likelihood estimate of β is found by maximizing the log-likelihood. This is done by solving the score equation

$$\frac{\delta l(\beta)}{\delta \beta} = \sum_{i=1}^{N} x_i (y_i - p(x_i;\beta)) = 0 \tag{4.71}$$

which is commonly done by applying the Newton-Raphson algorithm. [17]

A common way of deciding which explanatory variables should be included in a logistic regression model is to minimize the Akaike Information Criteria (AIC). The AIC is defined as

$$\text{AIC} = -\frac{2}{N}\text{loglik} + \frac{2d}{N} \tag{4.72}$$

where N is the number of observations, loglik is the maximized value of $l(\beta)$ for the estimated model and d is the number of parameters in the model. The parameters in the model are chosen by maximizing the loglik value and therefore does the negative sign on the first term in the definition make the AIC negative rewarding. The latter term then penalizes model with many explanatory variables by adding a positive number to the first term. To use AIC for model selection, the model giving the smallest AIC is considered as the optimal one.

4.10.2 Support Vector Machines

Two assumptions made when a logistic regression model is fitted to a data set, are that the classes are linearly separable and non-overlapping. However, these assumptions do not hold in general. A popular way of classifying data sets, for which these assumptions do not hold, is to use support vector machines (SVMs). Support vector learning is based on the idea of applying simple linear method to the data but in a high dimensional feature space non-linearly related to the input space and mapped by using basis expansions.

Once the basis functions $h_m(x)$, $m = 1,\ldots,M$ are selected, the SVM classifier is fitted using input features $h(x_i) = [h_1(x_i),\ldots,h_M(x_i)]$, $i = 1,\ldots,N$ and the non-linear function

$$\widehat{f}(x) = h(x)^T\widehat{\beta} + \widehat{\beta}_0 \tag{4.73}$$

for the separating hyperplane is produced. Then the decision function or classifier is defined as

$$\widehat{G}(x) = \text{sign}(\widehat{f}(x)) \tag{4.74}$$

It can be shown that the optimal hyperplane, in terms of classification performance, is the one with the maximal margin of separation between the two classes. It can be constructed by solving a constrained quadratic problem

$$\begin{aligned} \min \quad & \frac{1}{2}||\beta||^2 + \frac{C}{N}\sum_{i=1}^{N}\xi_i \\ s.t. \quad & y_i(h(x_i)^T\beta + \beta_0) \geq 1 - \xi_i \quad \forall i \\ & \xi_i \geq 0 \end{aligned} \tag{4.75}$$

whose solution β has an expansion

$$\beta = \sum_i \alpha_i h(x_i) \tag{4.76}$$

where non-zero coefficients (support vectors) occur when a point (x_i, y_i) meets the constraints. The coefficients α_i are found by solving the dual quadratic programming problem

$$\begin{aligned} \max \quad & W(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{i=j}^{N} \alpha_i \alpha_j y_i y_j \langle h(x_i), h(x_j) \rangle \\ s.t. \quad & 0 \leq \alpha_i \leq \frac{C}{N} \quad \forall i \\ & \sum_{i=1}^{N} \alpha_i y_i = 0 \end{aligned} \tag{4.77}$$

The cost parameter C of the SVM formulation in Equation 7 controls the penalty paid by the SVM for misclassifying a training point and thus the complexity of the prediction function. A high cost value C will force the SVM to create a complex enough prediction function to misclassify as few training points as possible, while a lower cost parameter will lead to a simpler prediction function.

Since both (4.73) and (4.77) involve $h(x)$ only through inner products, there is no need to specify the transformation $h(x)$ at all, but only knowledge of the kernel function

$$K(x, x') = \langle h(x), h(x') \rangle \tag{4.78}$$

that computes inner products in the transformed space. The kernel function K should be a symmetric and positive (semi-) definite function for which popular choices include:

dth Degree polynomial: $K(x, x') = (1 + \langle x, x' \rangle)^d$

Radial basis: $K(x, x') = \exp(-\sigma ||x - x'||^2)$

Neural network: $K(x, x') = \tanh(\kappa_1 \langle x, x' \rangle + \kappa_2)$

Finally, if class probabilities are desired instead of class labels, they can be obtained by fitting a logistic regression model to the estimated decision values.[17, 19]

CHAPTER 5

Static Seasonal Models of the Spot Price

To begin with, some linear time invariant models for the spot prices are considered. The construction of these models is not only to gain a better understanding of the price mechanism and its relationship with the past but will also be used as a benchmark for the more elaborate models.

5.1 Seasonal ARMA Models

5.1.1 Choice of Models

In some cases, transforming the time series with a non linear function can be helpful in order to obtain a time series that requires less effort to forecast. Furthermore, if the assumptions made when calculating the prediction interval in (4.18) are to hold, the range of predicted values must be independent of the mean. In Figure 5.1, daily and weekly means of the prices are plotted against the price range of the same day or week. The figure shows little correlation between the range and the mean. Therefore it is concluded that nothing will be gained by transforming the data.

A choice of proper order for the ARMA model is made by viewing the autocorrelation function (ACF) and the partial autocorrelation function (PACF), shown

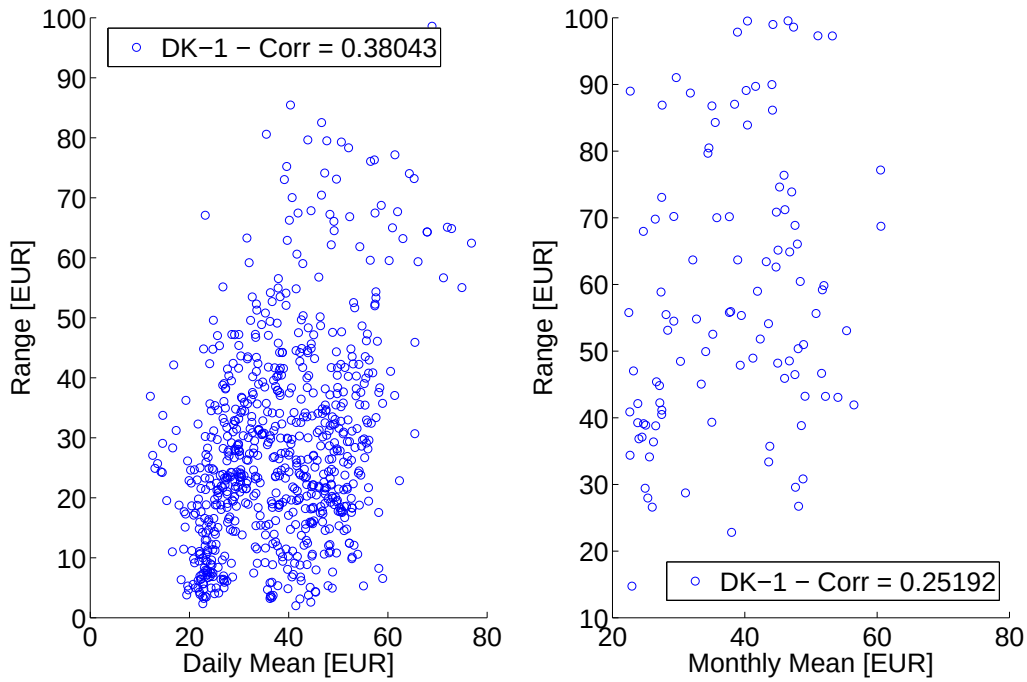

Figure 5.1: Range-Mean plot of the Price series, both daily mean (left) and weekly mean (right)

in Figure 5.2. From them, it is concluded that a model with seasonality on one or two time scales is appropriate. Furthermore, AR and MA parts of order 2 and 1, respectively, are deemed in order.

It is therefore decided to try out two models. The first model is a single seasonal model with a daily seasonal component, which according to Figure 5.2 will account for large share of the seasonal behavior. However, as the seasonal autocorrelation increases again after a week, a second model with two seasonal parts is made as well. This model is a double seasonal model with both daily and weekly seasonal parts. The k-step ahead prediction can then be found, using the notation explained in Section 1.3, as

$$\widehat{P}_{t+k|t+12} = \phi_0 + \phi_1 P_{t+12} + \phi_2 P_{t+12-1} + \Phi_{24} P_{t+k-24c} + \theta_1 \epsilon_{t+12} \tag{5.1}$$

for the single seasonal model and as

$$\widehat{P}_{t+k|t+12} = \phi_0 + \phi_1 P_{t+12} + \phi_2 P_{t+12-1} + \Phi_{24} P_{t+k-24c} + \Phi_{168} P_{t+k-168} + \theta_1 \epsilon_{t+12} \tag{5.2}$$

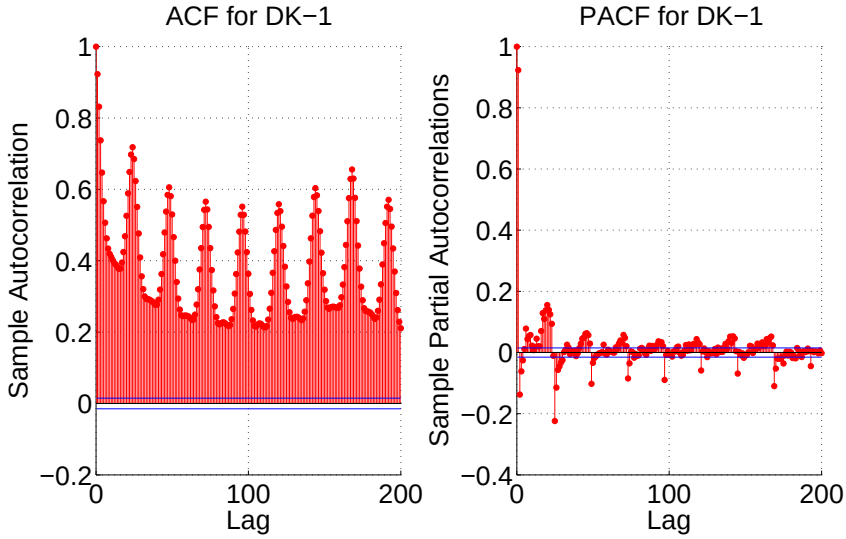

Figure 5.2: ACF (left) and PACF (right) for the price in DK-1

for the double seasonal model. In both models, c is defined by

$$c = \begin{cases} 1 & \text{if } 13 \leq k \leq 34 \\ 2 & \text{if } 35 \leq k \leq 58 \\ \vdots & \end{cases} \tag{5.3}$$

5.1.2 Evaluation of Forecasting Performance

Now the least squares estimate of the model parameters is found as described in Equations (4.21) - (4.28). For this parameter estimation, 80% of the available data, or around 18 months are used. The remaining 4 months are then taken for cross validation.

Figures 5.3 - 5.4 show the R^2 and the RMSE, introduced in Section 4.7, for the model when forecasting $13 - 44$ hours ahead in time. Judging from the figures, the performance of the model appears to be satisfying for the absolutely shortest prediction horizons. However, the performance falls quite rapidly as the prediction horizon becomes larger and model starts to rely more on the seasonal components. The decrease in performance then slows down for prediction horizons of $18 - 35$ hours as the predictions depend almost solely on the model's seasonal components. When the prediction horizon becomes larger than 36 hours, the performance falls quite significantly again, resulting in a

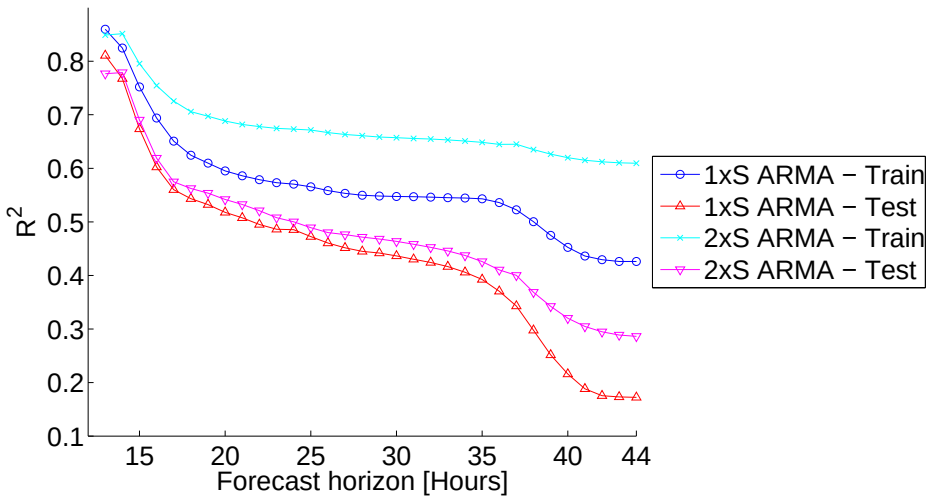

Figure 5.3: R^2 for different ARMA models

very poor performance for the largest horizons. This last performance drop is more severe for the single seasonal model as it is almost solely relying on a 48 hour old observation for its forecasting.

The double seasonal models seems to do better at all times and especially for the largest forecasting horizons. It must therefore be concluded that including the second weekly seasonality is worthwhile.

5.1.3 Forecasting

Due to the better performance of the double seasonal model, that model is chosen for further analysis. Now, forecasts are made, both for the training set and the test set, as defined by Equation (5.2) for $13 - 44$ hours ahead. The look-ahead times $13 - 44$ are chosen since $k = 13$ is the nearest hour where prices are unknown at noon, when the bids are sent in, and because $k = 44$ is the longest look-ahead time for which the wind power predictions, used in the models presented in Chapter 7, are available for every hour.

In Figure 5.5, residual histograms for 13 and 44 hour forecasts are shown. In the histograms, the residuals appear to be close to normally distributed around 0 as desired, but with somewhat heavier tails. What problems these heavy tails cause will be addressed later in the section. According to expectations, the distribution is more narrow for shorter forecast horizon and also for the

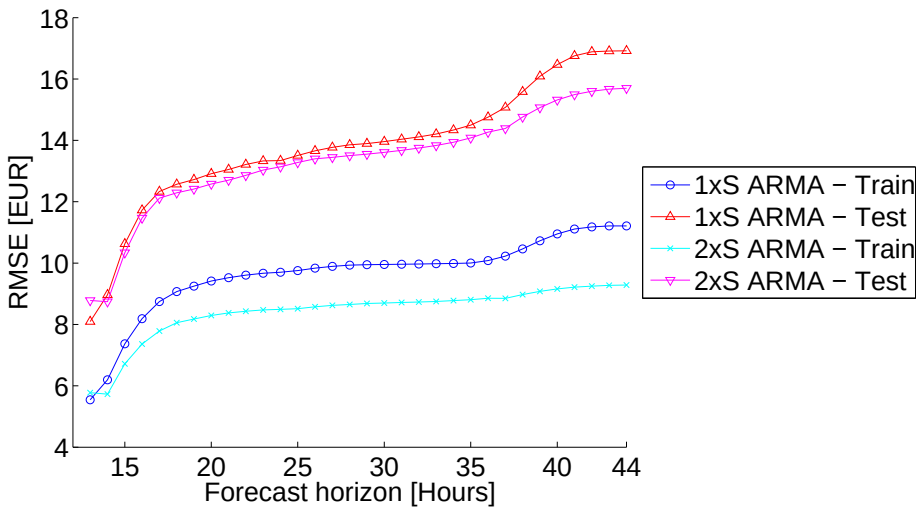

Figure 5.4: RMSE for different ARMA models

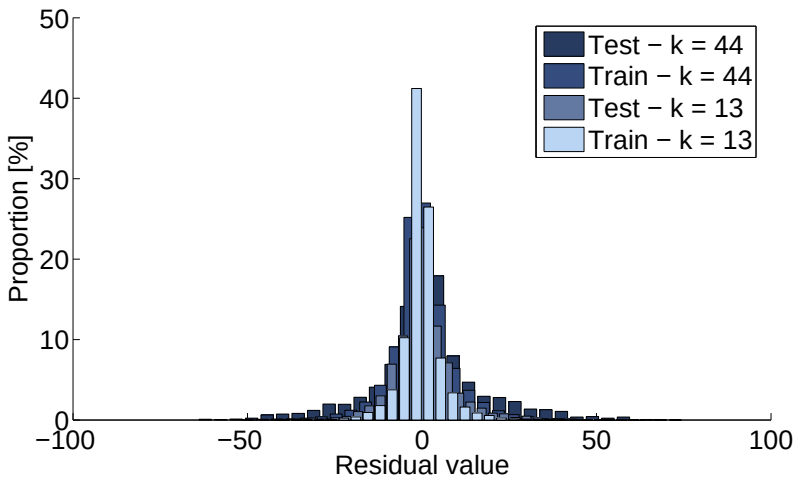

Figure 5.5: Residual histograms for 13 and 44 hour predictions in DK-1

training sets.

Figure 5.6 shows the residual ACFs for forecast, made with horizons of 13, 24, 36 and 44 hours. For the three larger horizons, everything seems to be in order as there is only autocorrelation detected in the lags shorter than the prediction

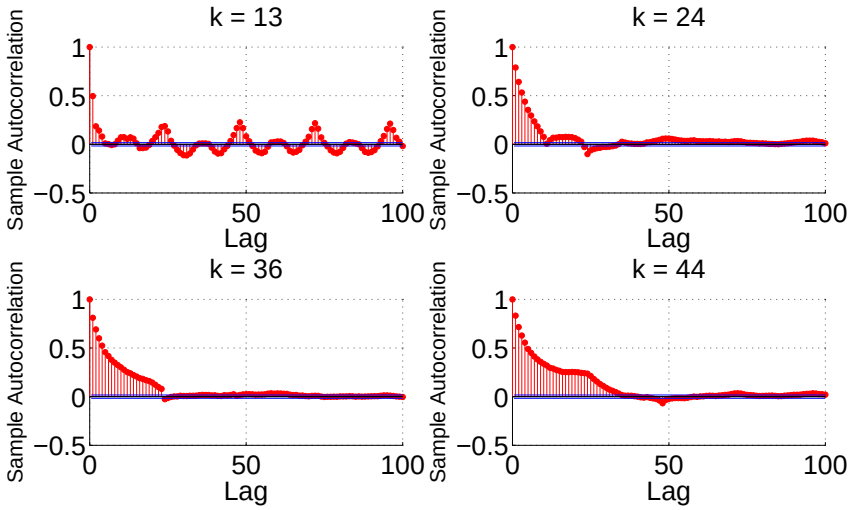

Figure 5.6: ACF of the residuals of the training set for 13, 24, 36 and 44 hour forecasts in DK-1

horizon. For the 1 hour horizon, a small daily seasonal autocorrelation is left. However, adding to the order of the model does not help to make this autocorrelation disappear and since it is so small, no efforts are made to eliminate it.

In Figure 5.7, predictions for a 32 hour period in September 2007 are shown along with prediction intervals, constructed by Equation (4.18), of 50%, 90 %, 95% and 99% certainty. According to expectations, the forecasts seem to be pretty good for the short prediction horizon and then get worse as the horizon gets larger, with few exceptions though. Prediction intervals get wider as predictions are made further ahead in time but do not seem to reflect the probability they are supposed to. However, nothing decisive can be concluded from these figures about the prediction intervals.

The prediction intervals are based on the assumption that the forecasted price is the mean in a normal distribution with the standard error as a standard deviation. To check whether this assumption holds, the reliability of chosen prediction intervals is examined. In Table 5.1 the proportion of observations inside the prediction intervals is shown.

From the table it is clear that the price distribution is much more heavy tailed than a normal distribution. The widest prediction interval is somewhat underestimated for the training set and heavily underestimated for the test set. Meanwhile, the 50% prediction interval is severely overestimated for the train-

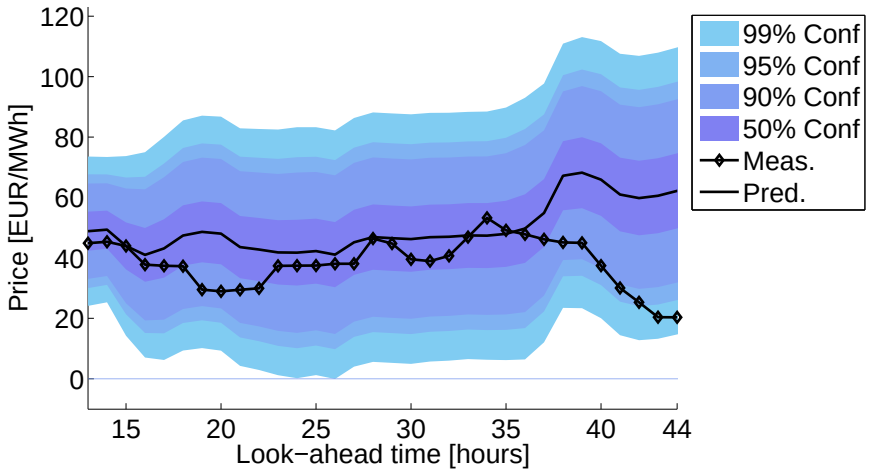

Figure 5.7: Forecasts 13 - 44 hrs ahead in DK-1

ing set while it is rather close to its nominal coverage for the test set. The two middle intervals are pretty well estimated for the training sets, but for the test sets, which have to be regarded as the more important one, the estimates are off by quite a lot. All this implies that the price distribution, and thereby the residual distribution, is more heavy tailed than a normal distribution. This conclusion is also supported by the fact that the reliability for each forecast horizon is not different from the simple one shown here.

An additional thing that catches the eye when viewing Figures 5.3 - 5.4 is the large difference between the in sample performance and the out of sample performance. Furthermore, when the forecasts made for the test set are visualized, the ones made late in the period seem to have larger errors than the ones at the beginning of the period. This indicates that a completely time invariant model, as this one, looses its ability to forecast as time goes. For testing this, the model is used to make forecasts for the first 500 hours in the test set and then again for the last 500 hours in the test set. In Figure 5.8 the difference in the performance between these two periods is plotted.

Table 5.1: Proportion of observations inside prediction intervals

	99%	95%	90%	50%
Training set	97.06%	93.73%	90.61%	67.58%
Test set	88.94%	82.28%	78.02%	52.01%

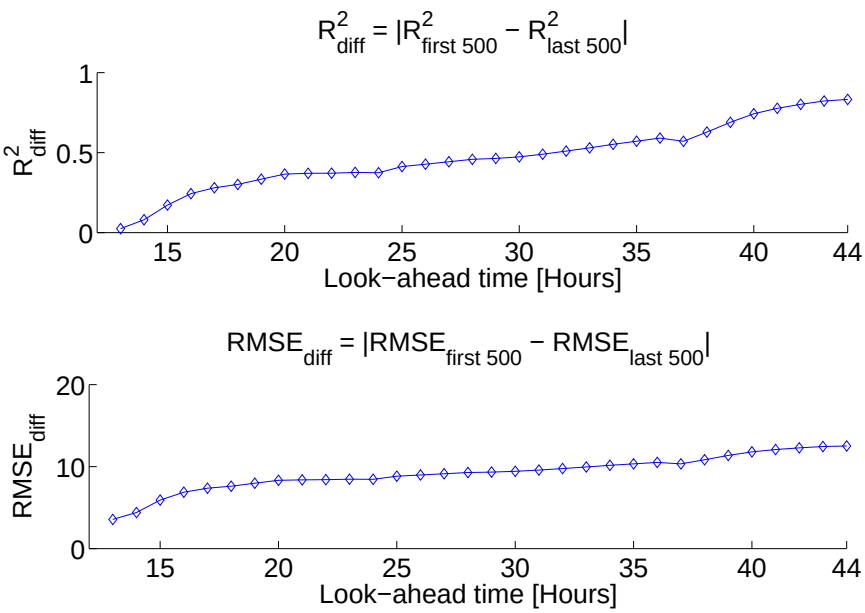

Figure 5.8: Difference in performance on the first 500 observations and the last 500 observations in the test set

The figure shows that for the absolutely shortest forecast horizons, the performance has decreased to some extend during this short period. On horizons larger than 4 hours, the model as completely lost its ability to forecast. Therefore, it must be concluded that if the spot prices are to be modeled over some period, at least some adaptivity has to be involved.

5.2 Holt-Winters model

Having concluded that adaptivity is necessary when modeling the spot prices, the double seasonal Holt-Winters model comes into mind as a candidate for improved forecasts, since it is adaptive by nature. Since Figure 5.1 shows very little correlation between the range and the mean, the additive form of the Holt Winters model is deemed in order. The additive form of the Holt-Winters

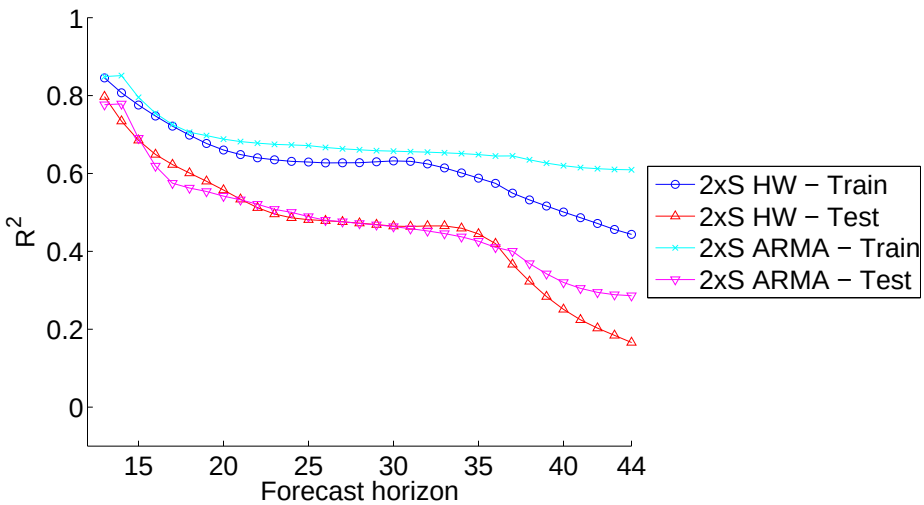

Figure 5.9: R^2 for the two Holt Winters models and the double seasonal ARMA model

model is

$$\mu_t = \alpha_1(P_t - (D_{t-24} + W_{t-168})) + (1-\alpha_1)(\mu_{t-1} + T_{t-1}) \tag{5.4}$$
$$T_t = \alpha_2(\mu_{t-1} - \mu_t) + (1-\alpha_2)T_{t-1} \tag{5.5}$$
$$D_t = \alpha_3(P_t - (\mu_t + W_{t-168})) + (1-\alpha_3)D_{t-24} \tag{5.6}$$
$$W_t = \alpha_4(P_t - (\mu_t + D_{t-24})) + (1-\alpha_4)W_{t-168} \tag{5.7}$$

where μ_t is the level, T_t is the trend, D_t is the daily seasonal part and W_t is the weekly seasonal part [22, 49]. The αs are smoothing parameters chosen as the parameters that minimize the residuals from 1 hour forecasts. For the spot prices, the k-step ahead prediction is then made by

$$\widehat{P}_{t+k|t+12} = \mu_{t+12} + (k-12)T_{t+12} + D_{t+(k-12)-24c} + W_{t+(k-12)-168} \tag{5.8}$$

where c is defined by (5.3)

The available data is divided into training and test sets in the same manner as before. The training set is used for estimating the smoothing parameters, α_i, $i = 1,2,3,4$. A secant version of the Levenberg-Marquardt method [18] is used for that estimation.

As can be seen from Figures 5.9 and 5.10, the performance of the double seasonal HW model is very similar to the performance of the double seasonal

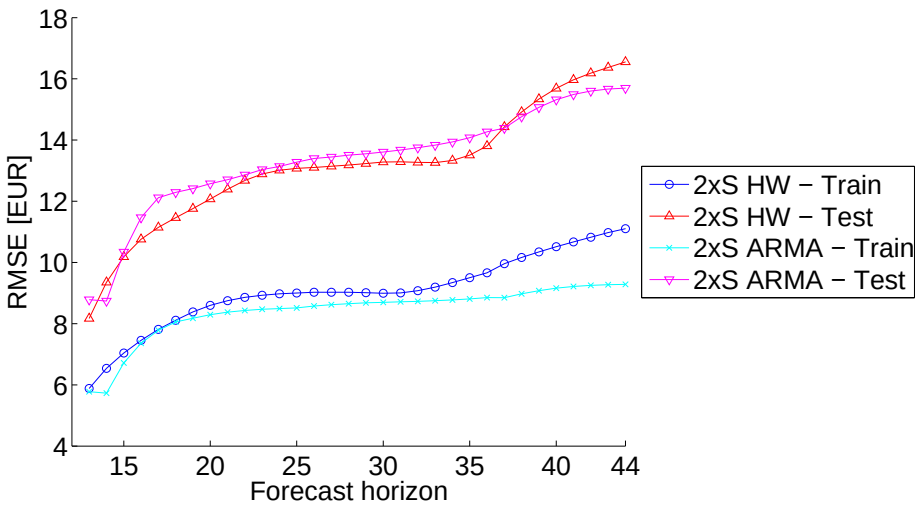

Figure 5.10: RMSE for the two Holt Winters models and the double seasonal ARMA model

ARMA model. However, since level, trend and seasonal components are constantly updated, the performance is pretty stable as time passes and would therefore require less effort to keep running.

Since it is clear that improved forecasts can not be obtained by the double seasonal Holt-Winters model, no further analysis is carried out. Furthermore, it must be concluded that more advanced techniques have to be utilized for improving the forecasts.

CHAPTER 6

External Factors Impacting the Spot Prices

It is a well known fact that prices on the Nord Pool spot market are mainly influenced by the water level in the reservoirs of the Norwegian and Swedish hydro power plants. However, changes in the water level happen on a much larger time scale than the prices are set on. Therefore, is the slowly varying water level more likely to set control the price level at each time, while other more rapidly changing factors are more likely to influence the intra day and intra week variations in the price. Furthermore, grid congestions and the price making structure on the Nord Pool spot market lead to these effects being local, to some extent. In the following, the aim is to come up with some explanatory variables for the electricity market price in Western Denmark and analyze how they affect the prices.

6.1 Static Analysis of the Effects of Wind Power Production Forecasts

In both of the Danish grid areas, wind power has a substantial share of the total production. The very low marginal cost of wind energy and the prioritization of the power produced by wind turbines, make produced wind power an influential variable in the price making process. In this context, as is stated

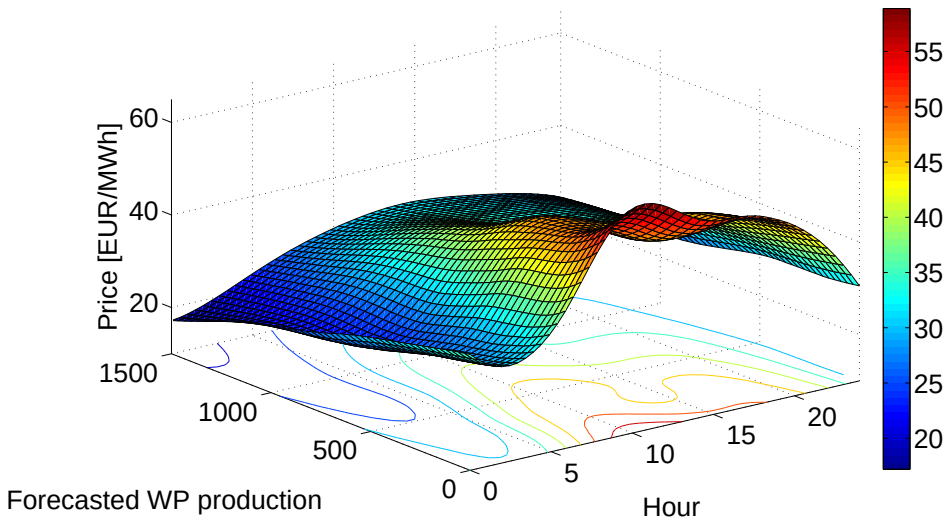

Figure 6.1: The dependance of the spot prices on forecasted wind power production and time of day

in [11], the right question to ask is: " What would the prices have been, if the wind hadn't blown?", but not: " What would the prices have been if there was no wind power in the Danish electricity system?". The fundamental difference between these two questions is that the first one relates to the situation of the wind power capacity in Denmark not being utilized at the time, while the second one relates to how the situation would be if no wind turbines were installed in Denmark. As the demand for electricity would be the same whether or not wind energy is produced, at least seen from a short term point of view, the energy would have to be produced in some other manner. However, wind power production is extremely volatile and therefore there is no way of knowing the actual production in a given hour when the spot prices are set. Traders bidding in the spot market therefore rely on forecasts of produced wind power in the price area when the make bids in the market. Hence, in this analysis which is intended to provide results suited for forecasting the question stated earlier is altered slightly to: "What will the prices be, if it is believed that the wind will/will not blow?"

In Figure 6.1 the average spot price in DK-1 is estimated as a function of both time of day and forecasted wind power production measured in MWh/h, using LWPR which was described in Section 4.5. From the figure, it is quite obvious that forecasts of large wind power production in a given hour will result,

on average, in a lower spot price in that hour, since when going along the y-axis, the mean price decreases. What is also very interesting is that the daily price raise, during the hours of the day where consumption reaches its daily peak, flatten out as wind power production in the system increases. The explanation is that the virtually nil marginal cost of the wind turbines shifts the supply curve to the right when more wind power is produced and therefore it takes more consumption to reach the steep end of it. In other words, the increased production of the turbines means that less cost efficient plants otherwise covering the base load, will be covering the peaks only. This prevents the even less cost efficient generators from being needed during the hours where demand peaks.

Since demand for electricity varies dramatically throughout the day and the week, the influence of the same amount of wind power, measured in MWh/h, can be different, depending on what time of the day and the week it is. More precisely, during evening and night hours the demand is generally lower and therefore is large wind power production during these hours bound to affect the price more, than it does during the day, since the supply and demand curves will cross each other more to the left during night. Although some of these effects can be accounted for by including time as one of the variables in the model, more of those effects come clear if forecasted wind penetration in the system is used for modeling rather than forecasted wind power production directly. This can be seen as the difference between Figures 6.1 and 6.2.

In Figure 6.2 a smooth estimate of the spot price, as a function of the time of the day and forecasted wind power penetration, is given. The figure shows the same type of effects as described before. However, the effects are more dramatic than in Figure 6.1 since the same actual production now has different effects depending on what time of the day and the week the production occurs.

Even though it is clear from Figure 6.2 that forecasted wind power does in fact have influence on the spot prices, it is hard to point out how big the actual effects are. In Figures 6.3 and 6.4 the extent of the effects are better illustrated. In Figure 6.3 the average spot price is shown for predicted wind power penetration on certain intervals in the period which the data set spans. It shows how the average spot price generally decreases as the share of wind power in the system increases.

In Figure 6.4, it is shown how much lower the electricity prices are on average if the forecasted penetration of wind power is within the intervals on the x-axis. Here, the assumption has been made that wind power penetration under 4%, corresponding to a production of approximately 150 MWh/h, has very little or no effects on the spot prices. The bars in Figure 6.4 represent how much lower, on average, the price is during periods of the given penetration interval, compared to the "no wind" situation. The plot clearly demonstrates that the spot prices tend to decrease as the forecasted wind power penetration

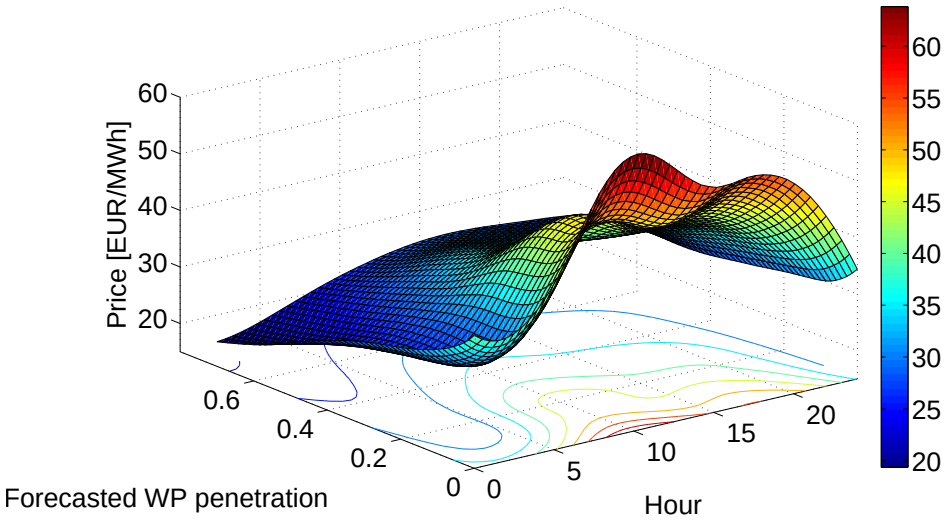

Figure 6.2: The dependance of the spot prices on forecasted wind power penetration and time of day

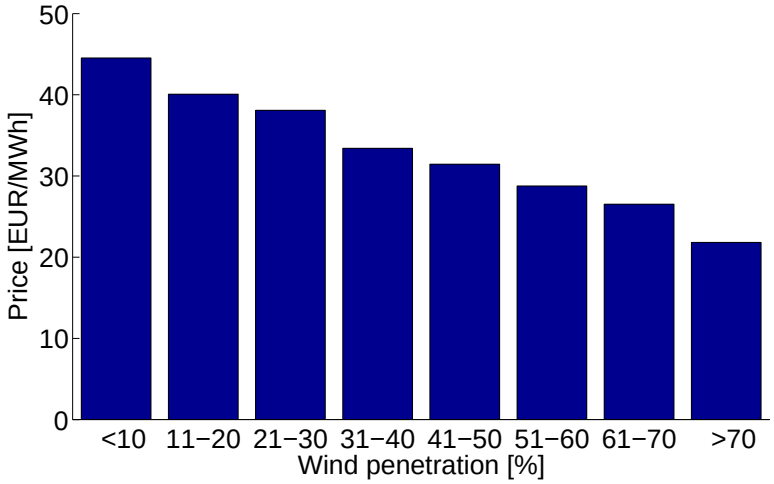

Figure 6.3: Average spot price, categorized by wind power penetration, in DK-1 in the period January 4th 2006 - October 31st 2007

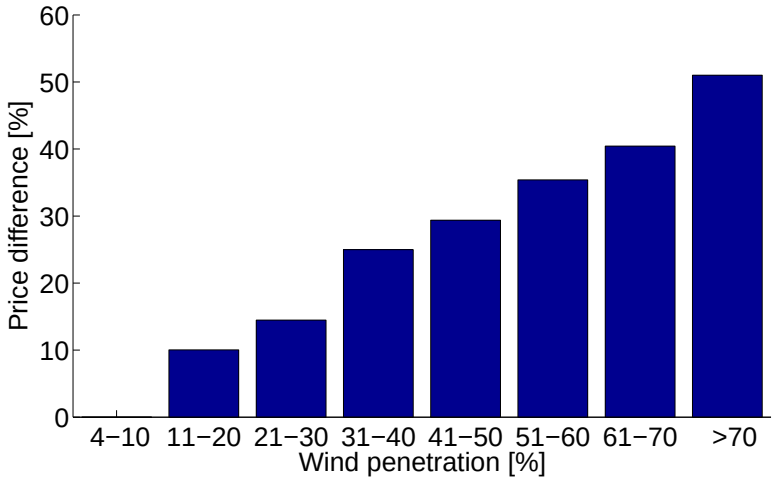

Figure 6.4: Decrease in average spot price, compared to the "no wind" situation, for different levels of wind power penetration in DK-1 in the period January 4th 2006 - October 31st 2007

increases.

Having established that forecasted wind power penetration certainly affects the mean spot price in DK-1, the question remains whether it affects the distribution of prices as well. Information about the price distribution is paramount, for reasonable choice of explanatory variables, when the uncertainty of the price forecasts is estimated in a non-parametric fashion. For carrying out the distribution analysis, the data set is divided into bins, according to forecasted wind power penetration, so that approximately 2500-3000 observations belong to each segment. Then the properties of the price distribution are estimated within each bin.

In Figure 6.5, histograms of the electricity prices are shown for different levels of forecasted wind power penetration. The figure shows, what already has been established, that the mean price shifts to the left as forecasted wind power penetration increases. Furthermore, the figure demonstrates the positive skewness of the price distribution and shows how the heavy tail of the price distribution becomes less severe as the forecasted wind power penetration increases. This translates to the probability of extremely high prices is much lower when the wind power penetration is predicted to be high.

The difference in distribution properties (see Section 4.1.2 for definitions) is summarized in Table 6.1. The first two lines in the table show the shift of mean and reduction in standard deviation, already discussed. Lines 3 and 4

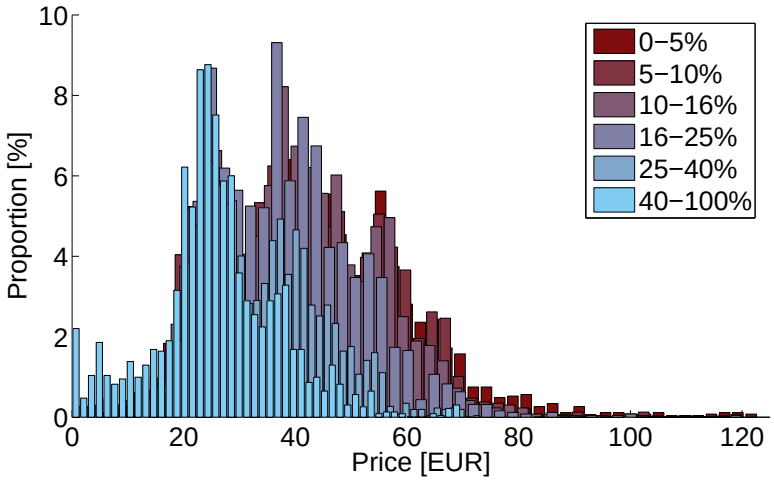

Figure 6.5: Distribution of prices for different levels of forecasted wind power penetration

show how the skewness and kurtosis of the distributions change for the different levels of penetration. Despite the fact that no obvious pattern can be detected when lines 3 and 4 are viewed separately, it can safely be stated that combined these numbers indicate different distribution properties. Taking the penetration intervals from left to right in the table, the first distribution is rather skewed and with high kurtosis, due to the heavy tail seen in the figure. Then as the wind power penetration increases in the 2nd and 3rd intervals, the tail gets less heavy, while the mean does not shift dramatically, explaining the decrease in both skewness and kurtosis. When the penetration reaches the level of the 4th interval, the mean has shifted more while the rather high prices still occur. Hence, the increase in skewness and kurtosis. For forecasted wind power penetration between 25-40% these extreme price situations no longer occur, re-

Table 6.1: Distribution properties of the price distribution for different levels of forecasted wind power penetration

	0-5%	5-10%	10-16%	16-25%	25-40%	40-100%
Mean	42.9807	41.1261	40.2579	38.0966	33.2420	26.0200
Std. Dev.	16.9512	15.3161	14.1772	13.0830	11.3455	11.2317
Skewness	0.8173	0.4775	0.3911	0.6270	0.3196	0.2937
Kurtosis	4.4137	3.3911	3.3962	4.0519	3.0388	4.0939

flected in a decrease both in skewness and kurtosis. Finally, for the highest forecasted penetration interval, the frequency of very low prices increases dramatically, explaining the reduction of skewness and increase in kurtosis.

An additional thing, caching the eye when Figure 6.5 is viewed is that the distributions seem to have two peaks. The different extent of these double peaks could contribute to the lack of pattern in the skewness and kurtosis. The cause of these double peaks is sought in Section 6.4.

6.2 Time Variation of the Distribution Properties

It was shown in Figures 6.1 and 6.2, that prices also vary seasonally with time. It is therefore natural to estimate the effects of time factors on the distributional properties of the spot prices.

First the attention is turned towards separation by the time of the day. The day is now categorized into three different categories. Night hours make up the hours from 00:00 - 07:00 when the majority of the population is asleep. Day hours are defined as the period from 07:00 - 19:00 - The hours when people is performing their daily tasks and using power consuming household equipment such as stoves and washing machines. In this period, electricity consumption reaches its daily peak and therefore are prices generally higher in this period. Finally, the evening hours are defined from 19:00-00:00. This period is a kind of a transition period between the other two periods, individual consumption decreases again during this period, yet it still makes up for significant share of the consumption.

In Figure 6.6, the price distribution for different periods of the day is shown and the same distributional properties as before as displayed in Table 6.2. Not surprisingly, the mean price is lowest during the night, and highest during the day. Also as expected, the standard deviation changes in the same way as the mean. During the night, the price distribution is quite sharp and very low prices are relatively frequent. Both of which, reflected by the low values of

Table 6.2: Distribution properties of the price distribution for different time of the day

	00:00 - 06:59	07:00 - 18:59	19:00 - 23:59
Mean	28.5679	42.2279	37.8245
Std. Dev.	11.4425	15.4165	12.8134
Skewness	0.0653	0.6149	0.897
Kurtosis	2.7665	3.9271	4.9001

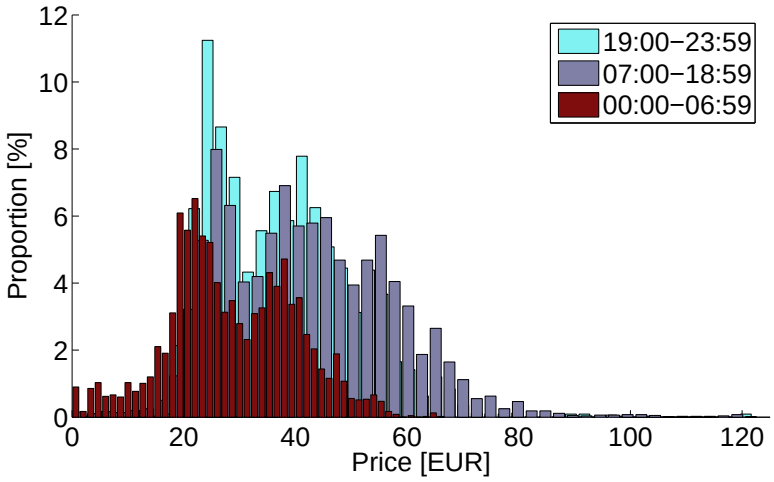

Figure 6.6: Distribution of prices for different time of the day

skewness and kurtosis in the distribution for that period. The very high skewness and kurtosis of the price distribution during the evening are due to the non-occurrence of low prices combined with a rather sharp distribution with a heavy positive tail. During the day, the distribution is rather flat so although the extreme price situations are most frequent in that period, the skewness and kurtosis are a bit lower than for the evening hours.

Apart from daily variations in the price, difference in consumption habits between working days and weekends are also likely to be reflected in the distribution of prices. The same analysis is therefore carried out on the data set divided according to whether it is a working day or a weekend and is summarized in Figure 6.7 and Table 6.3.

Again to no surprise, the lower consumption during the weekends is reflected

Table 6.3: Distributional properties of the price distribution for different days of the week

	Monday - Friday	Saturday - Sunday
Mean	39.4753	31.9257
Std. Dev.	15.4134	12.4493
Skewness	0.6681	0.2069
Kurtosis	4.0634	3.1563

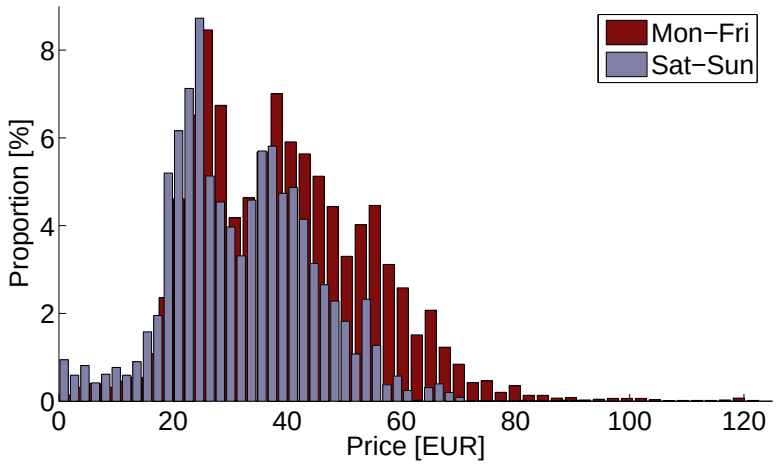

Figure 6.7: Distribution of prices for different days of the week

in a lower mean price and smaller standard deviation. Furthermore, the relatively lower frequency of extreme price situations during the weekends causes lower values of skewness and kurtosis for the weekends.

6.3 The Influence of Reservoir Water Level

It was stated, at the beginning of this chapter, that the level of water in the Norwegian and Swedish hydro power plant reservoirs is the driving force in the price making at Nord Pool's markets. Furthermore, it was stated that the changes in the water level happen so slowly that they are of very limited value, when price forecasts are to be made with short look-ahead times. Although the fact that the water level data are published with a resolution of one week should be a clear enough indication of this being the case, this is also demonstrated by carrying out similar analysis as was carried out for the forecasted wind power penetration in Section 6.1. The analysis is carried out on a data set covering the water level in Norwegian reservoirs only. However, the water level in Sweden is closely linked to the level in Norway, so despite Swedish data being missing, the analysis should give a good picture of the influence of hydro power in all of Scandinavia on the spot prices.

First by looking at Figure 6.8, where a smooth estimate of the price as a function of the water level and time of day, it can be seen that there is certainly a relationship between the water level and the price level. However, the water level

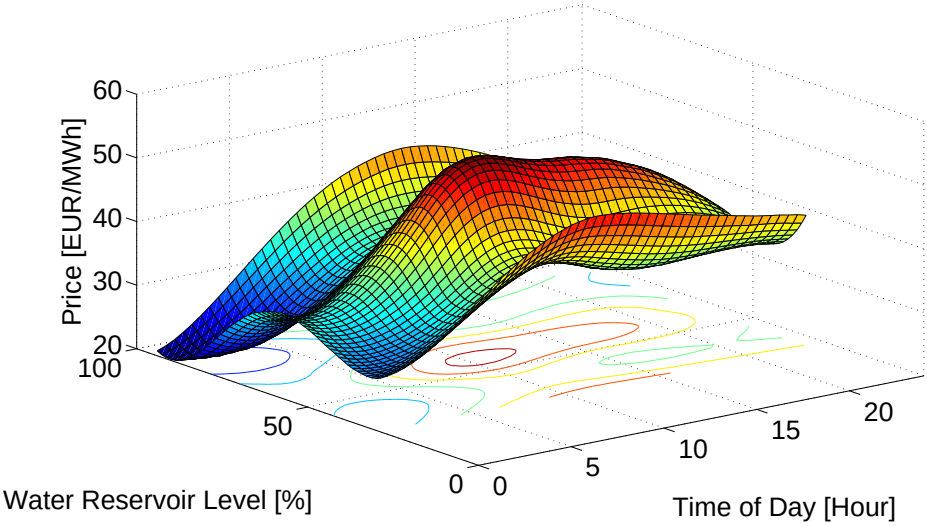

Figure 6.8: An estimation of the dependance of spot prices on water level and time of day

does not have any effects on the daily variation of the price level, as the wind penetration does, since the prices seem to fluctuate in the same manner for all levels of water. This is even clearer in Figure 6.9, where the price is estimated as a function of both water level and forecasted wind power penetration is given. Figure 6.9 shows how substantial the effects of forecasted wind power penetration are in comparison to the water level, since the mean varies on a much larger scale along the x-axis than it does on the y-axis.

Moreover, the extremely slow variation of the water level demonstrated in Figure 6.10, which shows the water level for the period from 1st of January 2006 throughout October 2007, make it unnecessary to account for the water level in a short-term forecasting model.

The same applies for the change in distributional properties. Even though there is a change of the price distribution, as seen in Figure 6.11 and Table 6.4, these changes happen on such a slow scale that they are irrelevant for the purposes of this thesis.

Finally, it seems strange that the price level seems to be higher for the intermediate level of water in the reservoirs. However, since the data set only covers little under 2 years, corresponding to around 95 observations, caution should be taken when drawing conclusions about this fact. For this reason, nothing

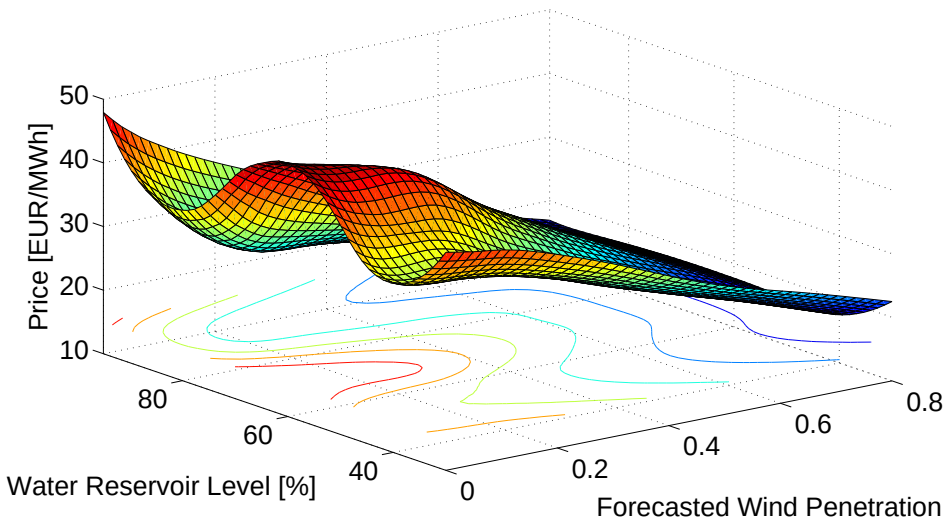

Figure 6.9: An estimation of the dependance of spot prices on water level and forecasted wind power penetration

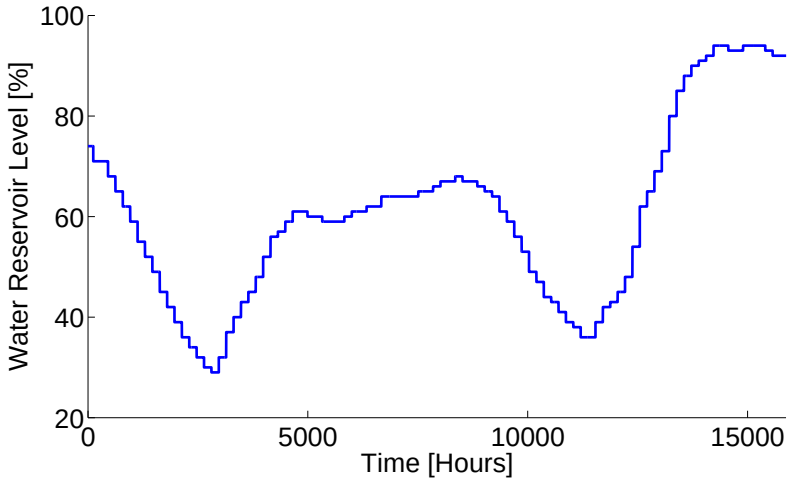

Figure 6.10: Time series plot of the water level in Norwegian hydro reservoirs

will be concluded about the dependance of price on water level directly, apart from that it exists.

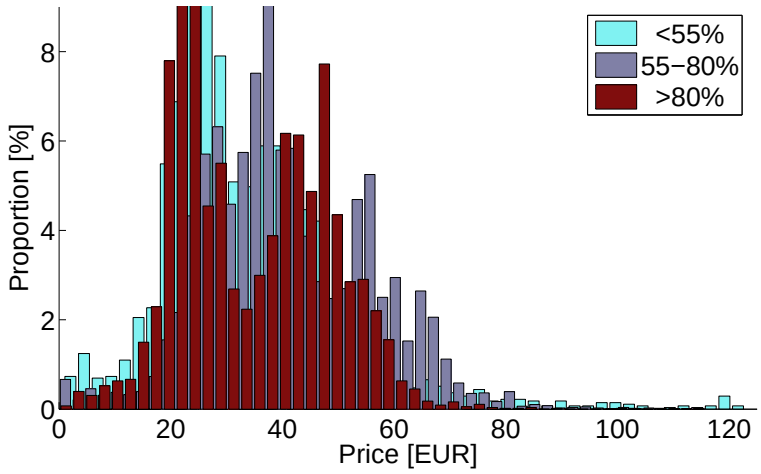

Figure 6.11: Distribution of prices for different days of the week

6.4 Double Peaks in Price Distributions

In the majority of the histograms presented in this chapter, one can't help noticing the double peaks of the distributions. None of the factors which was analyzed completely eliminated these double peaks, leading to the conclusion that they are caused by a combination of all these factors. To summarize, the double peaks could be the result of a combination of these factors:

- Change in consumption habits between summer and winter makes the mean shift to the right during the summers.
- The data set covers almost two years that were very different in the meteorological sense. The year 2006 was the warmest and driest one in years, while circumstances were very good for both hydro power plants and

Table 6.4: Distributional properties of the price distribution for different levels in Norwegian hydro reservoirs

	0 - 55%	55 - 80%	80 - 100%
Mean	34.4479	40.8623	33.1780
Std. Dev.	13.4070	14.8996	16.1143
Skewness	0.3853	0.3643	1.7524
Kurtosis	2.8444	3.3366	8.6745

wind power production in 2007.

- Where time factors have not been excluded, the daily and weekly variation of the mean price can cause the distribution to be double peaked.

The double peaks will therefore diminish to some extend in the adaptive time series analysis framework used in the following chapters. Furthermore, it might be worth mentioning that the probabilistic methods used in these chapters has no problems with handling double peaks, as it makes no assumptions about the distribution.

CHAPTER 7

Adaptive and Non-Linear Point Forecasting of the Spot Price

The main problem with the models presented in Chapter 5 is that their forecasting capability falls rapidly as the forecasting horizon becomes larger. It is obvious that some external signals, that are not affected by the horizon length, have to be used. Furthermore, the non linearities in the relationship to these external signals must be accounted for.

In addition, the rapidly falling forecasting ability of the model in Chapter 5, as time passes, suggests that adaptivity is a desired property of the model. By making the model adaptive, the model will adjust to the slow variations in the price mechanism, regardless of what causes the variations, and therefore not loose its forecasting ability over time.

7.1 Model Structure

The general idea behind the models now described is that, LWPR, introduced in Section 4.5 is used to create conditional parametric model of the spot price as a function of external variables. The model can be said to calculate function value for a function of the same type as displayed in Figure 6.2, where the

function itself is updated as new observations become available. The resulting smooth estimate of the spot price is then included in an additional layer of a dynamic model, which is added on top of the conditional parametric model.

Conditional parametric models of two different dimensions are constructed. One models the price as a function of time of day and forecasted wind power penetration and the other one has an additional input variable, indicating the weekday for which the estimate applies. For the dynamic part, Recursive Pseudo Linear Regression (RPLR) is used, an extension to the RLS-algorithm, described in Sections 4.3 and 4.4, that allows for MA parts to be included in the model as well. The algorithm is identical to the RLS-algorithm in terms of implementation but the theoretical extension is made since the design matrix $\mathbf{X_t}$ is now dependent on $\grave{}_{\mathbf{t}}$ [22, 23].

The choice of model order for the RPLR model is made from the ACF presented in Figure 5.2 in Chapter 5 and after examination of a few different model orders, not displayed here, a model for which the k-step ahead forecast is found by

$$\begin{aligned}\widehat{P}_{t+k|t+12} = \omega_t \widehat{P}^{cp}_{t+k|t+12} + \phi_{1,t} P_{t+12} + \phi_{2,t} P_{t+12-1} + \Phi_{s_1,t} P_{t+k-24c} + \Phi_{s_2,t} P_{t+k-168} \cdots \\ + \theta_{1,t}\epsilon_{t+12} + \Theta_{s_1,t}\epsilon_{t+k-24c} + \Theta_{s_2,t}\epsilon_{t-24(c+1)}\end{aligned} \tag{7.1}$$

is chosen, where

$$\widehat{P}^{cp}_t = f(\text{Time of day, Wind power penetration}) \tag{7.2}$$

if the two dimensional conditional parametric model is used (Model A) and

$$\widehat{P}^{cp}_t = f(\text{Time of day, Day of week,Wind power penetration}) \tag{7.3}$$

if the three dimensional conditional parametric model is used (Model B). As before c is defined by

$$c = \begin{cases} 1 & \text{if } 13 \leq k \leq 34 \\ 2 & \text{if } 35 \leq k \leq 58 \\ \vdots & \end{cases} \tag{7.4}$$

The purpose of increasing the dimensions of the conditional parametric model is to account for more of the seasonal behavior in the process and therefore leave less variations for the regression model to catch. The performance of the conditional parametric model can be measured alone and then gives R^2 of 0.1787 for the 2D version in (7.2) and $R^2 = 0.2993$ for the 3D version in (7.3). Although this performance numbers are not impressive by themselves, an input to a model that describes this much of the process' variation is of great

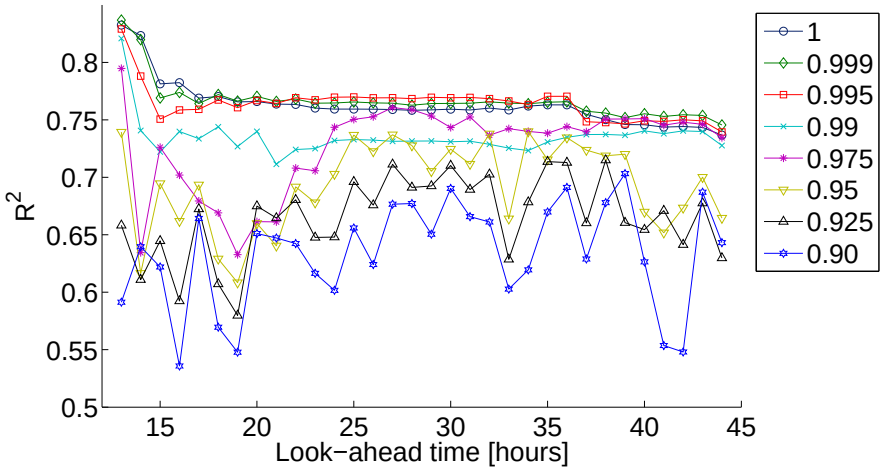

Figure 7.1: R^2 of model B for different values of λ

value. Another characteristic of the model that is of great value is that since the dependence is only estimated for variables that are known in advance, the performance of the model does not fall when the look-ahead time is increased. This explains why the performance measures in Figure 7.3 and 7.4 do not show decreased forecast quality when k is increased as it does for the static models.

The appropriate forgetting factor, λ, has to be chosen as well. The forgetting factor is commonly chosen as a value between 0.9 and 1 and simply found by trying out different values [23]. The search for the proper value of λ is therefore executed by running the models with $\lambda \in [1.0, 0.999, 0.99, 0.975, 0.95, 0.925, 0.90]$ and then the forgetting factor which leads to the best performance, both in terms of R^2 and $RMSE$ is concluded to be the optimal one. Figures 7.1 and 7.2 show the R^2 and $RMSE$ (see Section 4.7 for definition) for model B and for all horizons. The difference between the performance measures for the higher values of λ is very small, and therefore they all can be taken as a proper choice. Nevertheless, $\lambda = 0.999$ is chosen, both because it results in the most even performance values and also because the model will then have some forgetting, and thereby increase the model's chances of maintaining its forecasting ability past the scope of the data set used here. As described in Section 4.4, the number of effective observations is derived from Equation (4.48) and with $\lambda = 0.999$, the number of effective observations is 1000. This translates to that data from approximately the past 6 weeks is used for parameter estimation at each time.

The behavior of the performance measures for the four lowest λ-values is quite strange. The performance is very unstable and in some cases, significantly

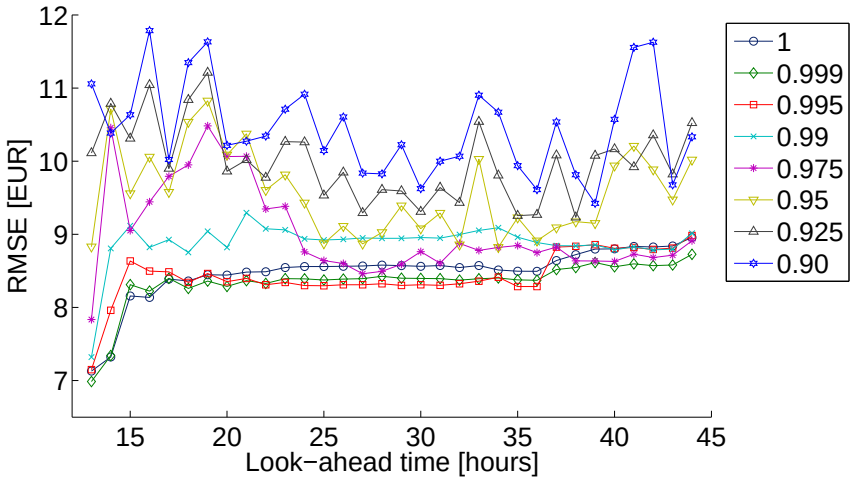

Figure 7.2: RMSE of model B for different values of λ

better performance is detected in the larger horizons than for the shorter ones. The explanation for this is most likely that the model forgets too quick and therefore are rarely occurring extreme situations worse handled by these model than models with more memory.

7.2 Model Quality

The performance measures, R^2 and $RMSE$, for the model with 3 explanatory variables in the conditional parametric model (model B) has already been displayed in Figures 7.1 and 7.2. The figures show that the performance of the model is quite satisfactory for all horizons tried. Furthermore, they show that after the first two look-ahead times, for which the model gives the very best forecasts, the performance is rather stable and only decreases a bit when k exceeds 36 and observations are no longer available for the preceding day.

Figures 7.3 and 7.4 compare the performance of models A and B to the models presented in Chapter 5. Including the third dimension in the conditional parametric model has very little effects on the shortest horizons. However, as the prediction horizon gets larger, the performance of Model B does not fall quite as much as it does for Model A. Furthermore, it has to be assumed very likely that this development continues beyond the 44 hour scope presented here. Including the performance measures of the more simpler model earlier presented, clearly demonstrates the value of adaptivity and accounting for the non-linear

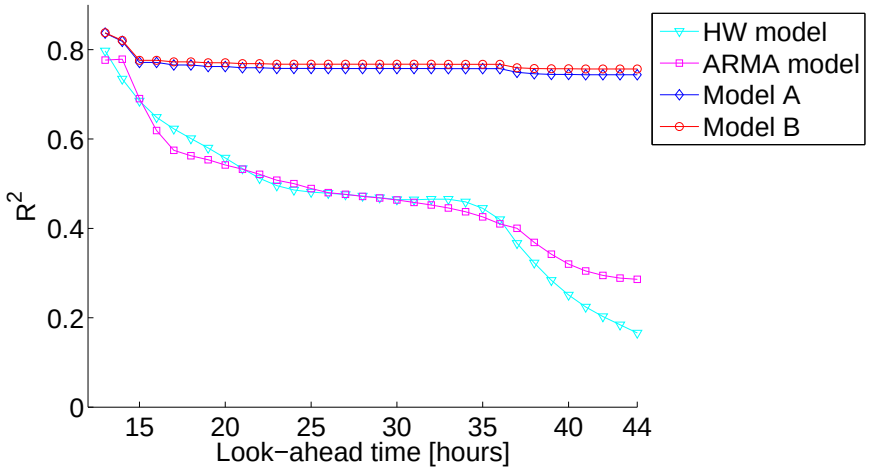

Figure 7.3: R^2 of the 4 different point forecasting models tried

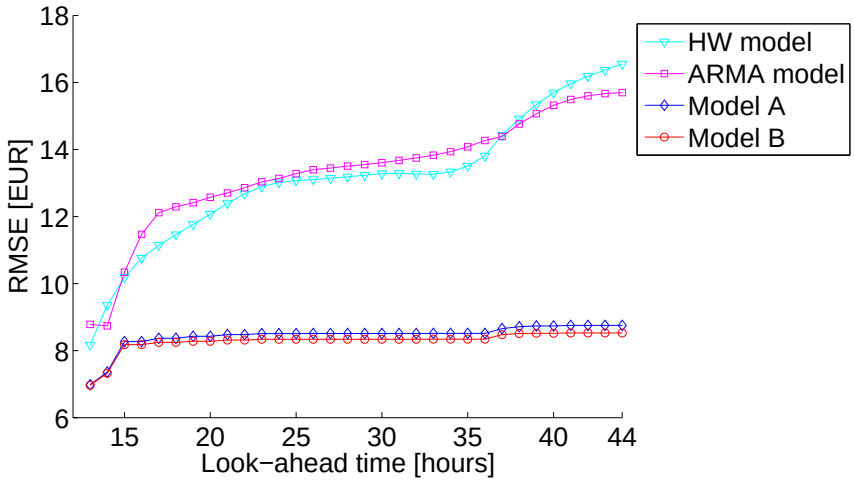

Figure 7.4: RMSE of the 4 different point forecasting models tried

relationship between the spot prices and the external inputs, especially for the larger look-ahead times. The more complex models perform steadily throughout all horizons while the more simpler ones completely loose their forecasting ability.

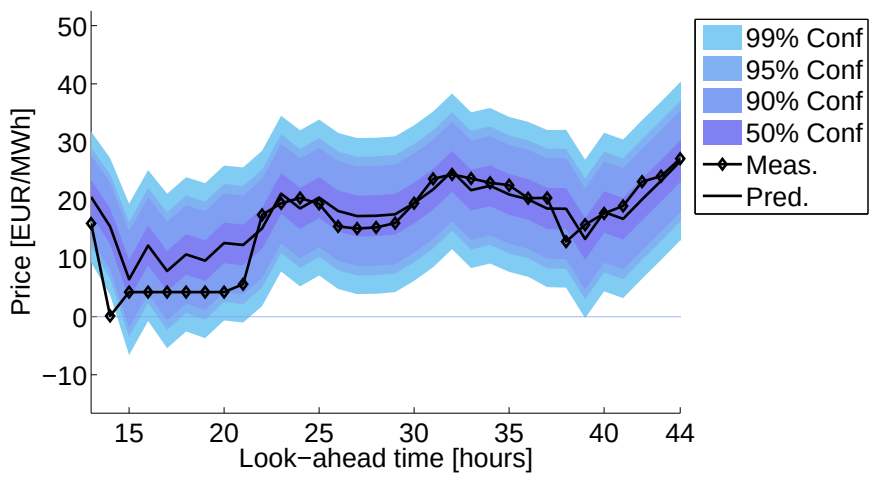

Figure 7.5: Predictions made at 00:00 on January 14th 2007 for 13 - 44 hours ahead in time

7.3 Forecasting With the Adaptive Model

In Figure 7.5, predictions made by model B for $13 - 44$ hours ahead in time are shown for a randomly chosen date along with the 50%, 90%, 95% and 99% prediction intervals. The prediction intervals are found by the classical time series analysis approach

$$\widehat{Y}_{t+k|t+12} \pm u_{\alpha/2}\sigma_k \tag{7.5}$$

where $u_{\alpha/2}$ is the $\alpha/2$ fractile in the standard normal distribution and σ_k^2 is the variance of the prediction error, estimated from all available data at each time point.

In the small example the figure shows, the predictions seem quite satisfactory. However, as can be seen from Table 7.1, the prediction intervals are, on average, way too narrow for 3 out of 4 intervals. The table shows the proportion of actual observations that fall within each prediction interval. Although, the table only shows these results averaged over all prediction horizons, same results are obtained for each horizon individually. It is therefore quite clear that the assumption of the prices being normally distributed around the forecasted mean does not hold. Consequently, one must conclude that if information on forecasting uncertainty is desired, more advanced forecasting techniques like for example time adaptive quantile regression [29] must be used.

Histograms of the residuals from the forecasts made by model B, with horizons

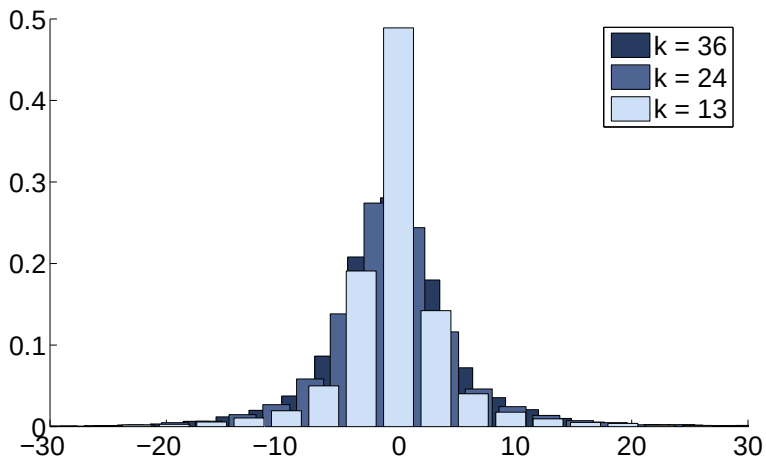

Figure 7.6: Residual histograms for forecasts with look-ahead times 13, 24 and 36

13, 24 and 36, are presented in Figure 7.6. The residuals are distributed around 0 as desired. However, the distribution seems to be more heavy tailed than a normal distribution and therefore the residual can not be completely regarded as white noise. The residual distribution is sharpest for the shortest look-ahead time and then slightly flattens out as the forecasting horizon gets larger. This is according to expectations and reflects the slightly decreasing quality of the forecast with increased look-ahead time. Although the histograms are only shown for three forecasting horizons here, they were viewed for every horizons, all showing very similar distributions.

The heavy tails of the residual distribution are also demonstrated by the Q-Q plots presented in Figure 7.7. The deviation of the blue crosses, representing the real quantiles, from the ideal red line show that the distribution of the residuals is somewhat more heavy tailed than a normal distribution. The heaviness of the tails however, does not seem to increase with larger k. The most likely explanation for these heavy tails is therefore rarely occurring extreme prices, caused by unforeseeable breakdowns in the grid and inter-connections or other

Table 7.1: Evaluation of the reliability of prediction intervals for Model B

	99%	95%	90%	50%
Model B	89.05%	83.40%	79.09%	49.28%

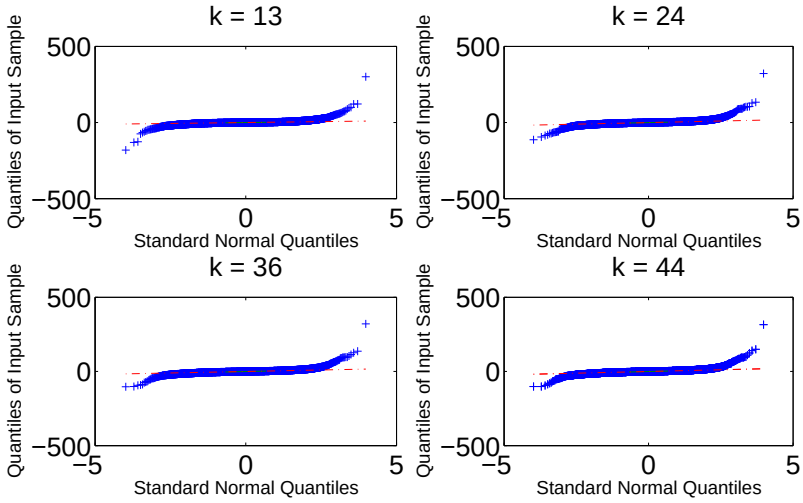

Figure 7.7: Q-Q plots of the residuals for forecasting horizons 13, 24, 36 and 44

events that prevent normal function of the system.

The plots of the autocorrelation functions of the residuals, in Figure 7.8, show that there is only very little or no autocorrelation in the residuals beyond the mathematical prediction horizon. So the required property of the residuals that they are not correlated seems to be fulfilled or close to that.

The importance of the recursive parameter estimation is clearly demonstrated by Figure 7.9, where the evolution of the model parameters, for 13 hour forecasts, throughout the period is displayed. As the changing dynamics of the system and the seasonal behavior of the market contribute to a considerable change in some of the parameters while others are more stable. Yet all parameters are subject to some changes during these 22 months. This confirms what was previously said, that estimating the model parameters recursively, definitely is worth the effort. The evolution of the model parameters for other horizons are not shown here, since the changes are somewhat similar for all horizons, despite different parameter values.

All in all, an appropriate point forecasting model for the spot prices seems to have been found. The quality of the forecasts is satisfying and the residuals seem to be uncorrelated. However, if a reliable uncertainty estimation is wished for, more advanced methods than the classic normality assumption have to be put in force.

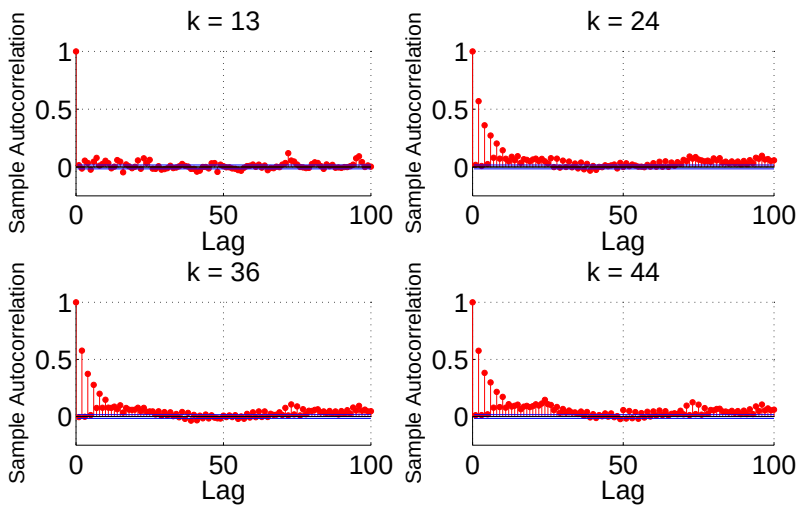

Figure 7.8: Autocorrelation functions of the residuals for look-ahead times 13, 24, 36 and 44

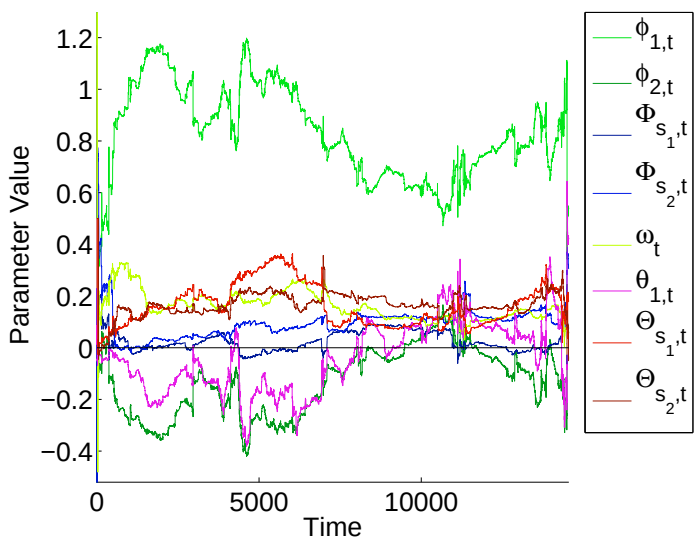

Figure 7.9: Evolution of the model parameters throughout the period which the data spans for forecast horizon 13 hours

CHAPTER 8

Quantile Regression Model for the Uncertainty

It was established in the previous chapter that the prediction errors from the point forecasting model are not normally distributed. Hence, the classic approach to constructing a prediction intervals by assuming gaussianity of the residuals does not result in prediction intervals that represent the real uncertainty of the price forecasts properly. Furthermore in Chapter 6 it was shown that the distribution of the spot prices depends on external factors. This fact is likely to be reflected in the dependence of the uncertainty on the same external factors. Therefore, if the actual uncertainty is to be estimated properly, some more elaborate method for uncertainty estimation have to be used. In this chapter quantile regression is used to estimate the uncertainty of the price forecasts. As previously described in Section 4.8, quantile regression can be used to estimate the distribution of some output variable (in this case the prediction errors from the price forecasts) conditioned on a set of input variables.

8.1 Implementation of Quantile Regression

Due to ever changing dynamics in the system and based on the experience from constructing the point forecasting model, it is concluded that time adaptivity is necessary to carry this task out properly. Implementation of time adaptive non-

linear quantile regression can be extremely complicated and time consuming, if the implementation is to be both numerically stable and efficient. Proper implementation of time adaptive quantile regression has been the topic of at least two master thesis at DTU. It is therefore decided to use the implementation fully described in [16] for the analysis presented here.

The implementation is done in C++. It uses splines to handle non-linearities and the user can choose whether the splines are natural cubic splines, without a constant term, or periodic splines. The possibility of segmenting the explanatory variables into bins is given so the data either be considered as a whole or divided into bins for a more local estimate of the quantiles.

The purpose of having multiple bins is to provide a better description of the process in the tails of the input variables. It goes without saying that the observations in the tails are less frequent than the ones closer to the mean. Therefore, when these infrequent events actually occur, an estimate of the output variable will be obtained by parameters, mainly estimated by explanatory values that are very different from the ones in the tails, if the data is not segmented. Although this is more of a problem for time adaptive estimation, the problem still exists for a static estimation. By separating the data into bins, an estimate of the exploratory variable can be obtained primarily based on older observations of similar circumstances.

However, segmenting the data like this is not problem free when time adaptivity is also necessary. Time adaptivity of the model parameters becomes a desirable property when the dynamics of the process which the model describes, is evolving in time. Under these circumstances, a forecast solely based on old values will be of limited quality. So one must find the appropriate balance between the number of bins and the updating frequency of the bins. This aside, the number of bins can be anywhere from 1, corresponds to no division of the data, to as many as the data allows for without running into singularity problem.

How fast/slow the model forgets is controlled by the size of the bins. The size of the bins is defined by the user and then a ”sliding window” of that size is used to decide which observations should be included for the estimation at each time. So every time new observation becomes available, it is put in the bin it belongs to and the oldest observation in that same bin is disregarded and therefore, the quantile estimates become more flexible and local in time as the bin size is decreased. The rate of parameter updating is also controlled by the user, so for processes with only slow seasonal variations, this feature can be used to speed up the calculation process by not updating every time a new observation is available.

8.2 Model Setup

The flexibility of the quantile regression program leaves the user with numerous possibilities regarding model setup. Therefore, it is important to realize, in which direction each of the model setup parameters should be taken for obtaining the optimal model setup. For doing that, various model setups are tested and two measures of their overall performance viewed for determining the best model setup.

The main performance measure, viewed for this purpose is the average skill score, over all look-ahead times the model is estimated for. The average skill score is defined as

$$\overline{SSc} = \frac{1}{N_k} \sum_{k=1}^{N_k} SSc(\widehat{q}_k^{\tau}, \widehat{P}_k) \tag{8.1}$$

where $SSc(\widehat{q}_k^{\tau}, \widehat{P}_k)$ is defined by (4.67) and N_k is the number of horizons for which the model is estimated for. As a support for the skill score, the mean absolute deviation from the nominal quantile, defined as

$$\overline{r} = \frac{1}{N_k \cdot N_\tau} \sum_{k=1}^{N_k} \sum_{i=\tau_1}^{\tau_{N_\tau}} |r_k^i| \tag{8.2}$$

is also viewed. Here, N_τ is the number of quantiles estimated. The primary objective is to find the best performing model regardless of calculation time and complexity of the model. However, as shorter calculation time and less complex model is always desirable, these things are also kept in mind when finding the best setup. So when two models perform nearly equally well, the less complex model, or the one with the less frequent updates, is considered as the better one.

Bin sizes are varied from 300 to 1000 and updating rates of 168, 24 and 1, corresponding to a weekly, daily and hourly updates respectively, are tried. Furthermore, the number of bins is varied from 1 to 6. The spline knots are placed at the edges and at the medians of the bins, as recommended by [16]. An exception of that is made near the global maximums as the amount of data around those points is not sufficient for including them as knots. Explanatory variables considered are the forecasted wind power penetration, $\widehat{WP}$, and the spot price forecasts, $\widehat{P}$, and bin range and knot placement is decided on by viewing the values presented in Table 8.1.

In Table 8.1, 10 different quantiles of both variables are listed along with the their extreme values. It is evident from the table that both variables have a heavy tailed distribution and for that reason the decision is made to segment the data by equal proportion of data in each bin, instead of equal bin range. So when two bins are used for an explanatory variable, bin edges are set at

the extreme values and at the median. Additional knots are then placed at the other quartiles. In the four bin setup, bin edges are set at the quartiles and additional knots on the medians of each bin (0.125, 0.375, etc). Some exceptions from this choosing system are made though. When only one bin is used, the knots are placed at all three quartiles instead of at the minimum and at the median. Furthermore, due to the difficulties regarding knot placement near the maximum mentioned earlier, no knot is placed above the 0.875 quantile in the 6 bin setup in which the values from the analysis presented in Chapter 6 are used as bin edges.

Bin size above 500 often leads to singularity problems and furthermore, those models who were successfully estimated with large bin sizes performed quite similarly or worse than models with the same setup but smaller bins. Therefore no model with bin size above 500 is included in Table 8.2, where some of the setups tried and the corresponding performance are summarized. During the work on the thesis, numerous other setups were also tried out, but the ones shown in the table were chosen as representatives as they are thought to give a good overview of what was established regarding model setup.

From the data in Table 8.2 it is concluded that a bin size of 500 is appropriate since the skill score get lower, going from model 2 to 3 where the only difference is reduction in bin size. The table also shows that more frequent updates add significantly to the model performance. It is interesting to see that dividing the data into more bins results in reduced overall quality of the predictions. This contradicts what would naturally be expected and what has been shown for wind power predictions, where splitting the data into more bins increases the quality of the forecasts. The reasons for this will be discussed later on. Adding

Table 8.1: Quantification of the forecasted wind power penetration data set and the spot price point forecasts

τ	$\widehat{WP}$ [%]	$\widehat{P}$ [EUR]
Min	0.59	0
0.01	1.32	12.05
0.125	4.08	22.38
0.25	6.49	26.17
0.375	9.92	30.48
0.5	14.46	35.71
0.625	20.80	40.93
0.75	29.48	46.55
0.875	42.66	53.62
0.99	78.11	74.40
Max	111.18	503.27

Table 8.2: Setup and performance of various model setups of the QR model for the uncertainty of the spot price forecasts

Model no.	Update Rate	Explanatory Variables	No. of Bins	Bin Size	No. of Knots	Skill Score	$\overline{r}$
1	168	$\widehat{WP}$	6	500	10	-36.06	2.6433
2	24	$\widehat{WP}$	2	500	4	-33.31	1.8138
3	24	$\widehat{WP}$	2	300	4	-33.52	2.0651
4	1	$\widehat{WP}$	2	500	4	-32.74	1.6669
5	24	$\widehat{WP}$	4	500	5	-33.62	2.1507
6	24	$\widehat{WP}$	4	500	7	-33.67	2.1994
7	24	$\widehat{WP}$	6	500	10	-33.91	2.3822
8	1	$\widehat{WP}$	4	500	5	-33.25	2.0814
9	24	$\widehat{WP}, \widehat{P}$	4,1	500	5,3	-32.67	1.3724
10	24	$\widehat{WP}, \widehat{P}$	4,2	500	5,4	-33.02	1.1960
11	24	$\widehat{WP}, \widehat{P}$	2,1	500	4,3	-32.25	1.2880
12	1	$\widehat{WP}, \widehat{P}$	2,1	500	4,3	-31.37	0.8952

more knots to the model while keeping the number of bins the same does not help either. Finally it is clear that adding the explanatory variable $\widehat{P}$ improves the forecasting skill of the model.

Figure 8.1 shows the skill scores for individual forecasting horizons for the three last models in Table 8.2. First of all, the figure demonstrates that forecasts made with model 12 the most accurate ones for all look-ahead times since the line representing that model never crosses the other two. Secondly, the fact that all three lines behave in exactly the same manner reflects that although the level of the skill score is controlled by the setup of the model, the difference in the skill scores between horizons is, bounded by the quality of the point forecasts the output variable is based on. This is maybe stating the obvious, however it is important to realize that improving the point forecasts will also improve the quality of the uncertainty estimation. At last, the figure shows that reduced quality of the models with more bins is due to different performance for different horizons and therefore the reason must lie in the estimation of the quantiles.

A consequence of having more bins is that updates in each bin becomes less frequent. This would not be a problem if the range of values spanned by the bins was similar for all bins. However in the case of both wind power penetration predictions and spot price forecasts this is far from being true. In fact the positive skewness and the high kurtosis of the distributions of these two

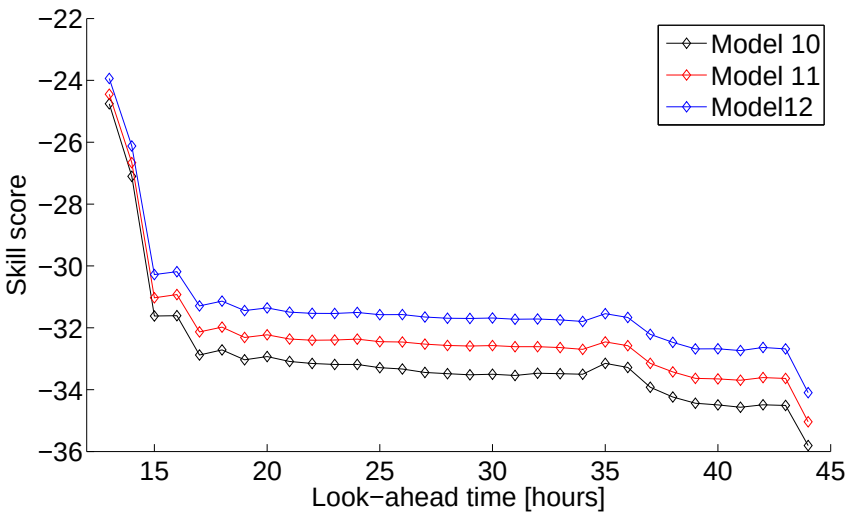

Figure 8.1: Skill scores for models 10 - 12 at each look-ahead time

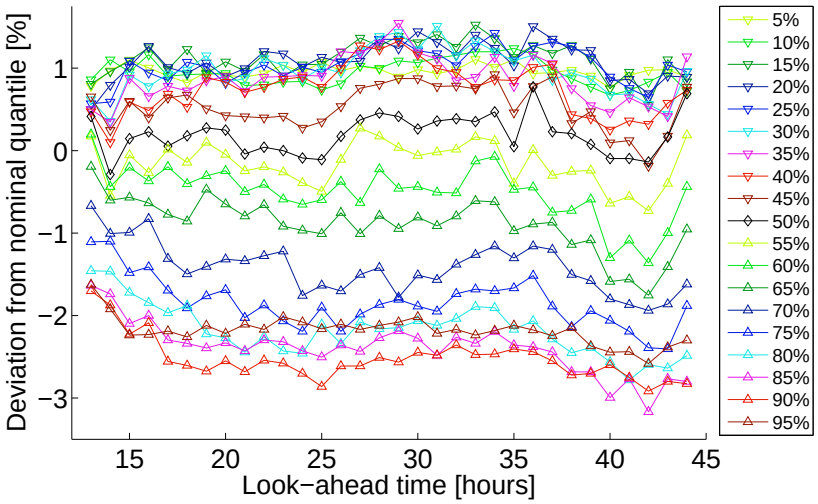

Figure 8.2: Reliability diagram for model 10 in Table 8.2

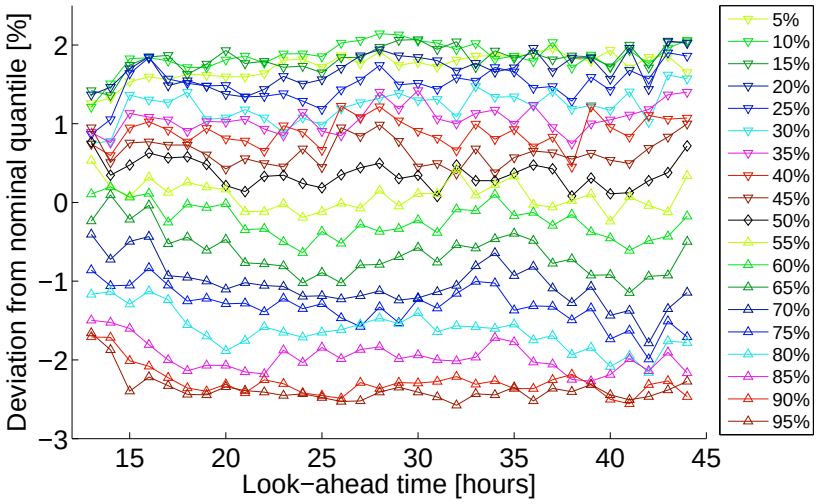

Figure 8.3: Reliability diagram for model 11 in Table 8.2

variables cause the "right" most bin to span more than two times the range of values than the other three do combined, in the four bin setting. Moreover, both variables are very correlated to their previous values causing updating of the bins often to happen in clusters, and especially this is the case for the wide spread bin with the highest values. Naturally, this large positive skewness and kurtosis will mainly affect the quantiles further from the mean. However, the fact that the prices are bounded by zero makes the impact much less severe for the lower end quantiles. By looking at Figures 8.2 and 8.3 it can be seen that this is the reason for the reduced prediction quality when the number of bins are increased. The low end quantiles benefit from the more concentrated bins close to zero, while the estimates of the high end quantiles become worse. The improvement of the estimates for the lower end quantiles is less than the quality reduction for the upper end quantiles and therefore the skill score becomes lower.

8.3 Model Analysis

Since both its skill score and average absolute reliability are the best ones, model 12 is concluded to the the best one and therefore chosen for further analysis. Uncertainty is known to always be underestimated [3] and therefore are the results of Figure 8.4 not surprising. The figure shows that all quantiles are

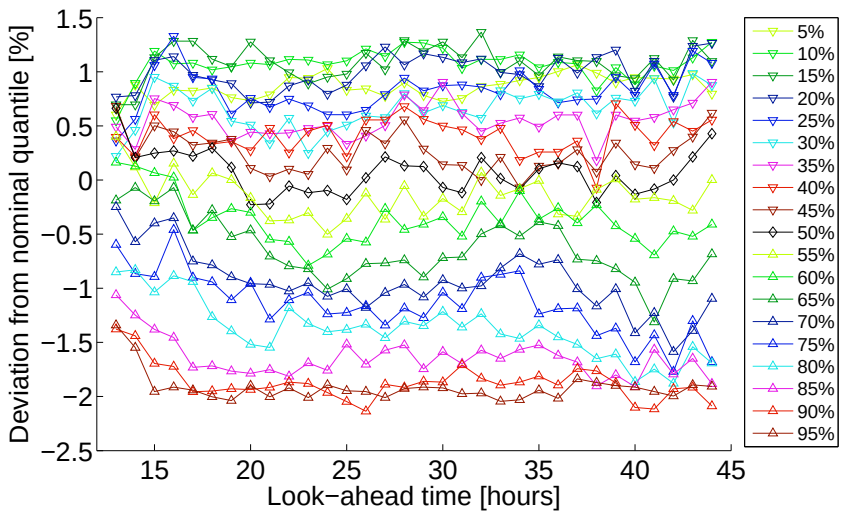

Figure 8.4: Reliability diagram for model 12 in Table 8.2

biased in the direction towards the mean, since all lines representing quantiles below 0.5 are above 0 and all lines representing quantiles above 0.5 are below 0. The deviation from the nominal quantiles is smallest for the shortest horizons as was to be expected from the skill scores presented in Figure 8.1 and then is quite steady for look-ahead times above 16 hours.

Also according to expectations is the slightly decreased sharpness, displayed in Figure 8.5, of the uncertainty estimates as look-ahead time is increased. In the plot at the upper left corner is the overall sharpness displayed and by comparing the overall sharpness to the sharpness of chosen individual horizons it can be seen that the sharpness for the shortest horizon deviates most from the overall sharpness, being more sharp, while the others show less deviation in the other direction. This reflects the superior performance of the point forecasting model for the shortest look-ahead times and the steady performance of the following horizons.

Figure 8.6, which shows real quantiles plotted against the Nominal ones, confirms what already has been stated, that the uncertainty is underestimated at all times since the black line is above the ideal red one for quantiles lower than 0.5 and below the ideal line for quantile values larger than 0.5 for all look-ahead times. Furthermore, the figure shows that the quantile forecasts seem to be very accurate when looked at with a more macroscopic view.

In Figure 8.7 forecasts made 12:00 on January 11th 2007 for the spot prices and

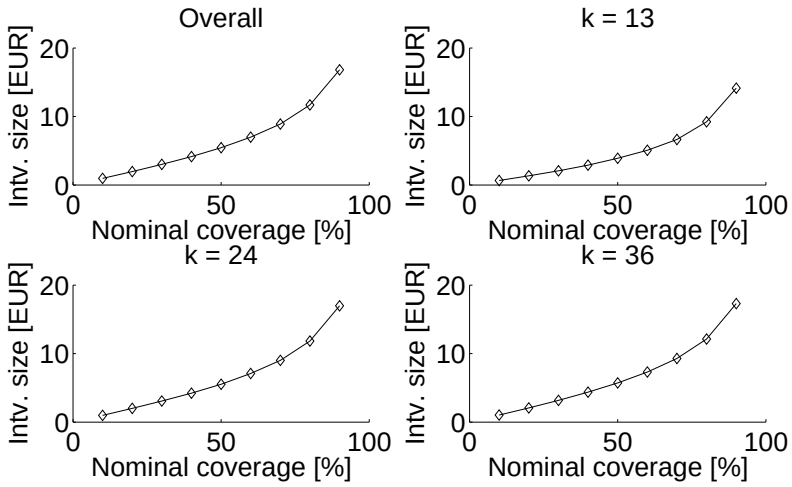

Figure 8.5: Overall sharpness of model 12 along with the sharpness for look-ahead times 13, 24 and 36 hours

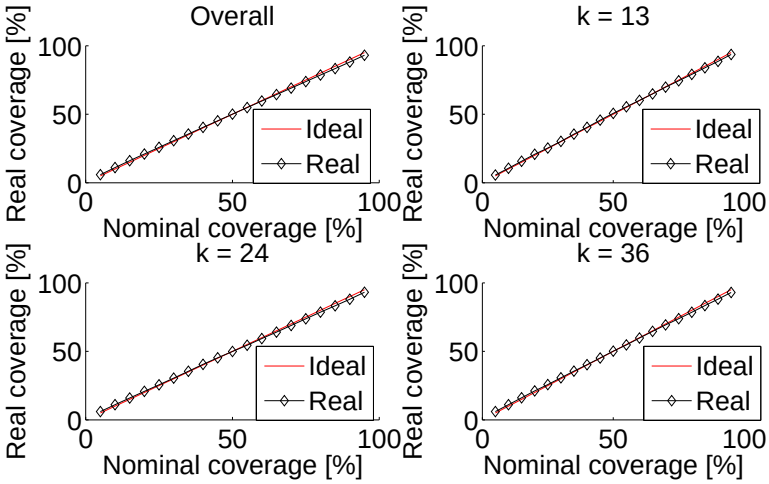

Figure 8.6: Real quantiles plotted against the Nominal ones for different forecast horizons and overall

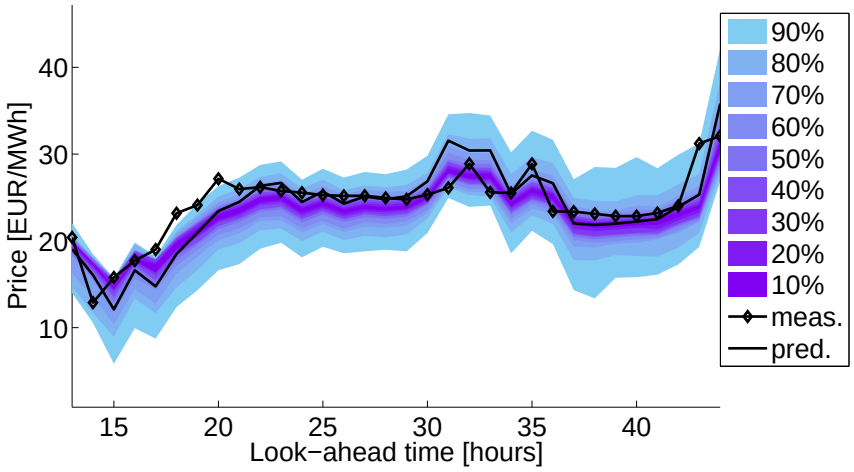

Figure 8.7: Predictions made at 12:00 on January 11th 2007 for 13 - 44 hours ahead in time

the uncertainty in the hours from 00:00 on January 12th and until 08:00 on January 13th 2007 are displayed. Compared to the prediction intervals presented in Figure 7.5 in previous chapter, the uncertainty estimates are now much more realistic. They are not symmetric around the mean anymore and vary not only according to the look-ahead time as they did before. Furthermore, the mean is not equal to the median any more.

8.4 Remarks About the Uncertainty Model

Overall, the quality of the uncertainty estimates is acceptable and should at the very least qualify as a good pointer towards reliable uncertainty estimates. It is clear that using quantile regression to construct prediction intervals for the spot prices lead to a much better estimation of the uncertainty than is obtained by the classic assumption of gaussianity. Quantile regression's lack of ability to handle heavy tailed distributions can clearly be seen from the poorer performance in the higher end quantiles estimated where improved estimates are needed. Some better estimates could be obtained by optimizing the model setup for each individual quantile instead of viewing the performance over all quantiles simultaneously.

Since quantile estimates are constructed independently of each other, the problem of quantiles crossing is known to exist [28]. However visualizations of the

forecasts made by the models in this chapter do not indicate that this problem is big here. For this reason, no actual investigation of the existence or extent of the quantile crossing problem has been conducted for the forecasts presented in this thesis.

CHAPTER 9

Forecasts for the Regulating Market

Now the attention is turned towards the regulating market. Having access to forecasts of both spot prices and regulation prices will enable the market players to design their bidding strategies in a more integrated way, taking into account more of the possibilities the market provides. However, forecasting on the regulation market is a whole different ball game than forecasting the spot prices.

The regulation price is in fact two separate processes[1], the up regulation price and the down regulation price. In theory, only one of them can be different from the spot price at each time. However, occasional breakdowns in the system, cause both up and down regulation prices are different from the spot price in the down period. Furthermore, since the regulation need is defined by overall imbalance in the system, as described in Chapter 3, the occurrence of no regulation need is more frequent than one might think.

Since the regulation price is bounded by the spot price in one direction, it is decided to model the difference between the two instead of the actual regulation price. This results in all negative values for the down regulation price

[1]Currently this is true for both producers and consumers in Sweden and Denmark. However from January 1st 2009 the one price system, described in Chapter 3, will be adopted for the consumption side in Sweden and Denmark, while the two price system still be used on the production side and adopted to the production side in Norway, resulting in a harmonized system for all three countries.

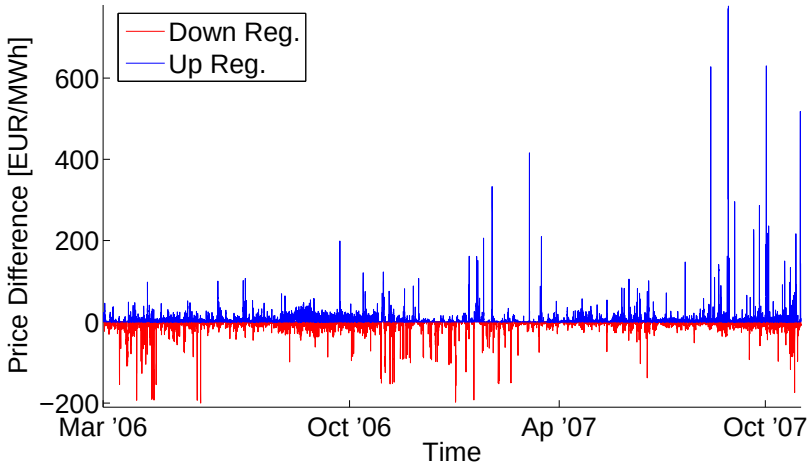

Figure 9.1: Time series plot of the difference between the Spot price and the regulation prices

and all positive values for the up regulation price, as can be seen on Figure 9.1, where the two time series are plotted for the time interval indicated on the x-axis. The plot shows that there is some seasonal variation in the regulation prices, especially in the down regulation price. The increase in the price difference, and regulation quantity, is mainly caused by the increased share of wind power in the overall generation during winters. However other factors such as more variance in the consumption and increased consumption contribute to the increased imbalances as well.

Because regulation in each direction occurs only for some hours, it is not possible to treat the regulation prices as a time series, since the time since last available observation varies from observation to observation. Therefore, the regulation prices have to be modeled separately, conditioned on the occurrence of regulation in that direction. Consequently the first step towards a model for the regulation price is to gain some knowledge of whether regulation will be needed and if so in which direction. Then price models are made for each direction, only based on a data set that contains observations for the corresponding regulation direction.

Before going any further it should be said that the analysis presented in this chapter is only intended to be superficial and preliminary. The goal is only to come up with some ideas about in which direction the modeling should be taken, for obtaining a reasonable model for the regulation market. All models presented in the following need some improvement before they performance

can be considered to be applicable in practice.

9.1 Classification of Regulation Direction

In order to gain some information about in which direction, if any, the regulation in a given hour will be, some classification models for the regulating direction are examined. Therefore, the binary response variables $I_{D,t}$ and $I_{U,t}$ are introduced and defined as

$$\begin{aligned} I_{D,t} &= \begin{cases} 1 & \text{if } \widehat{RP}_{D,t} < 0 \\ 0 & \text{otherwise} \end{cases} \\ I_{U,t} &= \begin{cases} 1 & \text{if } \widehat{RP}_{U,t} > 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{9.1}$$

where $\widehat{RP}_D$ the down regulation price and $\widehat{RP}_U$ is the up regulation price. So $I_{D,t}$ indicates whether there will be regulated down in hour t and $I_{U,t}$ indicates the same for up regulation. The occurrence of regulation is therefore defined in terms of deviations in price and not by regulation power bought in a given hour.

In [47], choices of explanatory variables in a logistic regression model for the regulation direction are discussed. There it is concluded, that an appropriate model will be on the form

$$I_{Dir,t} \sim M + W + H + T_G + T_N + P_t + WG_t + I_{Dir,t-1} \tag{9.2}$$

where $Dir \in \{D, U\}$ and the explanatory variables are as follows:

- M: A series of 12 indicator variables where the ith variable takes the value 1 if the observation is from the ith month of the year and 0 otherwise.
- W: A series of 7 indicator variables where the ith variable takes the value 1 if the observation is from the ith day of the week and 0 otherwise.
- H: A series of 24 indicator variables where the ith variable takes the value 1 if the observation is from the ith hour of the day and 0 otherwise.
- T: Quantity of electricity traded over the interconnections to Germany, T_G and to the Nordic countries, T_N.
- P_t: The spot price at hour t
- WG_t: The generated wind power in DK-1 at time t

As it is the intention here to come up with ideas of a model construction that will enable one to forecast the regulation direction before the spot prices are set, it is clear that the variables T, P_t and WG_t can not be used in the model derived here. Furthermore, it is very unlikely that the regulation direction so far in the past as 12 hour or more will influence the regulation direction significantly. Variables P_t and WG_t can easily be replaced by the corresponding forecasts of the spot price and the wind power penetration, $\widehat{P}_{t|t-k}$ and $\widehat{WP}_{t|t-k}$, which are available when the model is intended to be used. The problem remains that no forecasts of the traded quantity on the interconnections are available for such a large look-ahead time. What however is available are the trading schedules made after bids have been collected at noon each day. Although made with a slightly shorter horizon than are of interest here, the schedules share some characteristics with forecasts in general as they are estimates from which actual observations can deviate in either direction. Furthermore, since driving forces behind the trading on the interconnections are the spot prices in the connected areas (or the difference between them), it has to be deemed extremely likely that if price forecasts for Norway, Sweden and Germany, of similar quality as seen in Chapter 7 for DK-1, were available, reliable forecasts of the trading schedule could be obtained. It is therefore decided to use the trading schedules as variables T_G and T_N, denoted $\widehat{T}_G$ and $\widehat{T}_N$, and to compare it to a model where no trading is accounted for but otherwise the same. Thereby, some indication of the value of forecasts for the interconnection trading will be gained. Furthermore, a lower bound on the performance of a classification model accounting for such forecasts will be obtained.

This being said, it is concluded that a good initial model construction for a classification model of regulation occurrence is

$$\widehat{I}_{Dir,t+13} \sim M + W + H + \widehat{T}_G + \widehat{T}_N + \widehat{P}_{t+13|t} + \widehat{WP}_{t+13|t}. \tag{9.3}$$

where the variables $\widehat{T}_G$ and $\widehat{T}_N$ are excluded from the model when trading schedules are not used as explanatory variables.

The statistical software R is used for the model estimation and to begin with, the choice of explanatory variables is revised by using backwards elimination to minimize the AIC (see Section 4.10.1). The revision of the explanatory variables is carried out by optimizing the AIC for a logistic regression model (introduced in Section 4.10.1) estimated for the first 75% of the data set. Regardless of whether the trading schedules are used or not, the backwards elimination leads to the exclusion of both the month indicator, M, and the hour indicator, H. So a revised form of the model is

$$\widehat{I}_{Dir,t+13} \sim W + \widehat{T}_G + \widehat{T}_N + \widehat{P}_{t+13|t} + \widehat{WP}_{t+13|t} \tag{9.4}$$

where still the variables $\widehat{T}_G$ and $\widehat{T}_N$ are excluded from the model when trading schedules are not used as explanatory variables.

When the model is applied to the remaining 25% of the data set, the resulting hit rate is 61.1% and 58.6% for the down and up regulation respectively for the model including the trading schedule. So in 61.1% of the time, the model predicts correctly whether there will be regulated down in that hour or not. Correspondingly, 58.6% of the time, the model successfully tells whether up regulation occurs or not. For the model not including the trading schedule the hit rates for down and up regulation are 59.7% and 56.7% respectively. So in this setting, including the trading schedule adds around 2% to the hit rate. Regardless of whether the trading schedule is included or not, the resulting performance is obviously far from being acceptable. It should be stressed though that the model returns probabilities, as shown in Section 4.10.1, so in some cases the predictions are more certain than in others.

Time adaptivity has proven to be a useful property of the models constructed for the spot prices and therefore, a sliding window version of the logistic regression model is attempted as well. The model works in such a way that for a chosen window size N_w, the model parameters are estimated from the last N_w observations and then used to forecast one step ahead. Then the newest observation is added to the training set, the oldest one discarded and the model parameters are re-estimated. Window sizes from 168 to 1000 are tried and the one with the highest resulting hit rate concluded to be the best one. This has led to the choice of a window size 300 and the following results all refer to a model with that window size.

Using the sliding window technique, significantly improves the hit rate of the model which for this setting becomes, 67.8% and 69.2% for the down and up regulation respectively, when the trading schedule is included. The performance of the model is now significantly better and of course probabilities of either up or down regulation can still be obtained, allowing one to estimate how certain the classification values are. When the trading schedule is not included as an explanatory variable, the hit rates fall down to 64.5% for the down regulation and 66.3% for the up regulation.

One of the drawbacks of logistic regression is that it assumes linearly separated, non-overlapping classes. However, in the case of regulation direction, this assumption does not hold. Therefore, a support vector machine (SVM) model is constructed for the occurrence of regulation. SVMs handle overlapping and non-linearly separated classes by mapping them into a high dimensional space, where distance between the observations is larger, as described in Section 4.10.2.

For the construction of the SVM model, the `R` package `kernlab` [19] is used. In the models presented here, a radial basis kernel was used for which the value of the σ parameter is optimized, with respect to minimal cross validation error in the training set. The value of the cost parameter C however has to be determined and since the objective function is rather insensitive for changes

in C is usually varied on a logarithmic scale in stead of linearly. The C parameter is therefore chosen as the one giving the best hit rate out of $C = 10^x$, $x \in \{-1,0,1,2,3\}$. This has led to the choice of $C = 10$.

Both a static and a sliding window version of the SVM model is constructed. The model has the same explanatory variables as described by Equation 9.4 and a window size 300 is used. The static version is estimated with the same division of the data set as before and gives a hit rate that is around the same as the static version of the logistic regression model, or 57.3% for the down regulation and 58.4% for the up regulation when the trading schedule is used. When the trading schedule is not used, the resulting hit rate is 54.8% and 56.9% for the down-and up regulation respectively. So going from a static version of a logistic regression model to a static version of a SVM model results in a slight decrease in performance. However, the sliding window version gives a hit rates that are quite better than the ones obtained by logistic regression, as can be seen from the results summarized in Table 9.1. The hit rate of the SVM model with a sliding window of size 300 is 75.5% for the down regulation and 76.4% for the up regulation when the trading schedule is included in the model. A performance which has to be regarded as acceptable at least. What the table also shows is difference between the actual proportion of hours that have regulation and proportion of hours that are predicted to have regulation. The bias is calculated as

$$Bias_{Dir} = \frac{\sum_{t=1}^{N} I_{Dir,t}}{N} - \frac{\sum_{t=1}^{N} \widehat{I}_{Dir,t}}{N} \tag{9.5}$$

and therefore the small bias indicates that both types of misclassifications are equally probable. Not including the trading schedule in the model leads to a decrease in hit rate of 4 - 5%. Such a large decrease in hit rate indicates that constructing a forecasting model for the trading on the interconnections would be worth the effort. By including the trading schedules, the model's bias towards one response is reduced, also supporting the suggestion that modeling the interconnection trading would be worthwhile.

What the bias percentages translate to in terms of number of true/false classifications is what Tables 9.2 and 9.3 demonstrate where the actual amounts of true and false classifications for each predicted response are shown for down-and

Table 9.1: Hit rate and bias in predictions from the SVM model

	With Trading Schedule		Without Trading Schedule	
	Down	Up	Down	Up
Hit Rate	75.5%	76.4%	71.0%	71.2%
Bias	1.17%	-0.20%	3.66%	-1.41%

up regulation respectively.

The smaller bias of the model with trading schedule is reflected by the much more similarity between the row sums and the column sums for the models where the trading schedules are accounted for. This is especially the case for the down regulation model, but can be seen for the up regulation as well.

The SVM model therefore seems to be quite good. Although the best performance numbers were obtained by using an explanatory variable that is not available itself at the time of forecasting, the replacement of them with a reliable forecast of it is not likely to cause the performance to decrease a lot. Furthermore, there is a huge potential for improving the model by fine tuning the cost parameter, C, and varying the model structure over time. Both aspects that were not addressed here in great detail.

Table 9.2: Number of true and false classifications for both predicted responses for down regulation

		With Trading Schedule			Without Trading Schedule		
		True			True		
		$I_{D,t}=0$	$I_{D,t}=1$	$\sum$	$I_{D,t}=0$	$I_{D,t}=1$	$\sum$
Forecasted	$\widehat{I}_{D,t}=0$	6976	1829	8805	6994	2166	9160
	$\widehat{I}_{D,t}=1$	1662	3777	5439	1644	3440	5084
	$\sum$	8638	5606		8638	5606	

Table 9.3: Number of true and false classifications for both predicted responses for up regulation

		With Trading Schedule			Without Trading Schedule		
		True			True		
		$I_{U,t}=0$	$I_{U,t}=1$	$\sum$	$I_{U,t}=0$	$I_{U,t}=1$	$\sum$
Forecasted	$\widehat{I}_{U,t}=0$	5504	1669	7173	5273	1727	7000
	$\widehat{I}_{U,t}=1$	1697	5374	7071	1928	5316	7244
	$\sum$	7201	7043		7201	7043	

9.2 Quantile Regression Model for the Distribution of Regulation Prices

Along with information about the probability of regulation being done, some information about the associated price is desired as well. Due to the irregular occurrences of the observations, the time series analysis approach taken for the spot price can not be imposed directly on the regulation prices. Furthermore, for most possible applications of the forecasts of both regulation prices and spot prices, the forecasting uncertainty will play a more important role than the forecasted mean. It is therefore decided to model the distribution of the difference between the spot price and regulation price directly using quantile regression. Thereby advantage is taken off the rather reliable spot price forecast and regulation prices are modeled as a distribution of deviations from the spot price.

Given the previous experience of choosing explanatory variables, the spot price forecast and the predicted wind power penetration are chosen as inputs. The k-step ahead prediction for the τth quantile can therefore be written as

$$\widehat{RP}^{\tau}_{Dir,t+k|t} \sim \widehat{P}_{t+k|t} + \widehat{WP}_{t+k|t} \tag{9.6}$$

In order to save computation time and since the exploratory variable is the same for all horizons, it is decided only to try out the model for $k = 13$. Quality of the predictions are however expected to degrade in harmony with what was observed for the spot price uncertainty in the previous chapter.

Three different model setups are tried, and the resulting performance measures (the same ones as in Table 8.2) are displayed in Table 9.4. Bin edges and knot placements are determined in the same manner as before.

Compared to the setups tried in the previous chapter, the similar reduction in quality is detected when the number of bins is increased for the same reasons

Table 9.4: Setup and performance of the three model setups tried

	Bin Size	No. of Bins	No. of Knots	Skill Score	$\overline{r}$
Down	500	2,1	5,3	-75.19	2.3367
	500	4,2	7,5	-77.71	2.9489
	300	2,1	5,3	-74.24	1.9037
Up	500	2,1	5,3	-71.60	2.0272
	500	4,2	7,5	-72.88	2.4567
	300	2,1	5,3	-70.58	1.6963

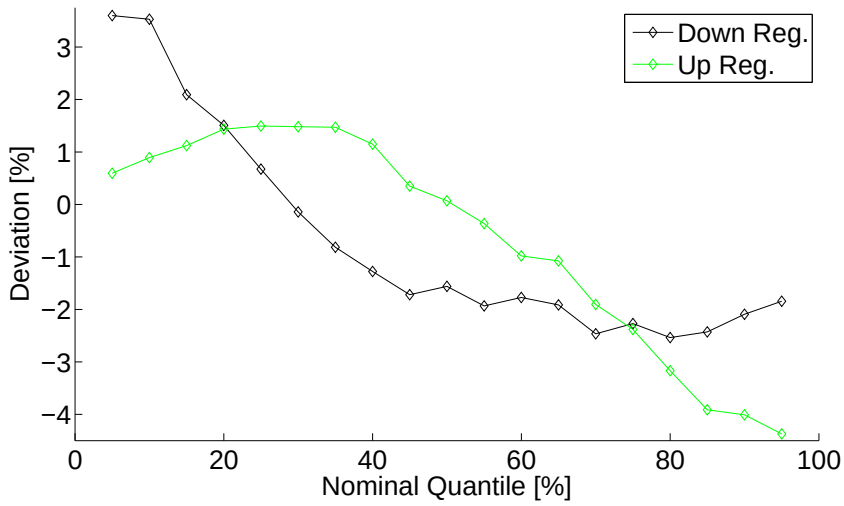

Figure 9.2: Reliability of the quantile estimates for down and up regulation

as discussed there. As for the bin size goes, the smaller bin size of 300 does somewhat better than the larger one. The best performance is obtained by the same model setup for both directions, although the quality of the forecasts made for the up regulation price is slightly better. The forecasts made with the model setup with bin size 300 are therefore taken for further analysis

In Figure 9.2 the reliability of the quantile estimates is displayed as the deviation from the nominal quantile plotted against the nominal quantile. The fact that the majority of the observations are quite close to zero, for both regulating directions, contributes to a satisfactory quantile reliability for the high end quantiles of the down regulation prices and the lower end quantiles of the up regulation prices. However, the heavy tails of the distribution of the regulation prices result in a quite severe underestimation of the quantiles at the opposite ends. The reliability of the intermediate quantiles seems to be quite satisfactory and as for the quantile forecasts presented in previous chapter, the uncertainty is underestimated most of the time. An exception to that are the quantiles between 25 and 50% for the down regulation which is due the the extreme heavy tails of the distribution of down regulation.

The sharpness of the estimated distributions is shown in Figure 9.3 and as expected the intervals with low nominal coverage are quite sharp while the intervals with the large nominal coverage are much wider, reflecting the extreme tails of the price distributions.

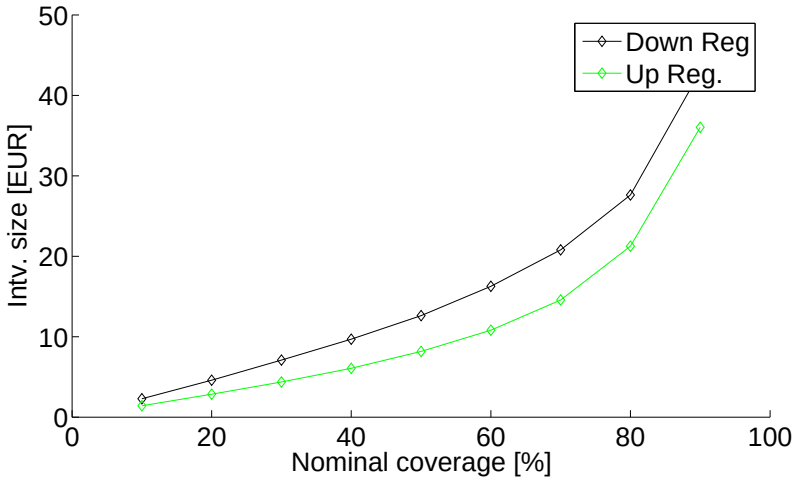

Figure 9.3: Predictions made at 00:00 on January 12th 2007 for 13 - 44 hours ahead in time

An example of the heavy tails estimated for the distribution can be seen in Figures 9.4 and 9.5 where 50 consecutive forecasts for the deviation of the down and up regulation price respectively, from the spot price are shown.

9.3 General Remarks About Modeling the Regulation Market

The quality measures estimated for the models presented in this chapter indicate that the modeling approaches applied could give satisfactory results. However, as stated in the beginning of the chapter, some further analysis has to be carried out before this can be stated with certain. Nevertheless, the analysis presented here can without doubt be used as a pointer towards a successful modeling of the regulation market.

Obviously, many aspects of the models could be improved significantly. Improved forecasts could presumably be obtained by tuning the penalty parameters in the SVM model and the model setup in the quantile regression model. Furthermore, the classification models assume independence between the occurrences of regulation in different direction. However since the breakdowns causing regulation in both directions to happen are relatively rare, this assumption is not likely to hold.

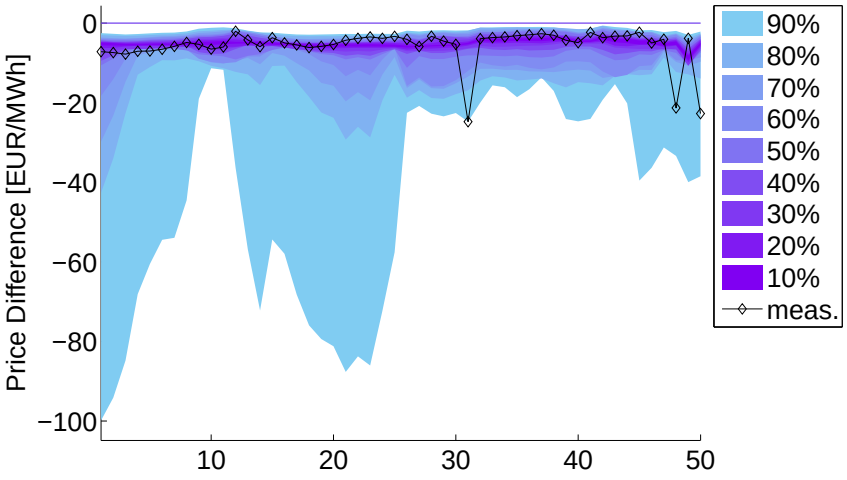

Figure 9.4: 50 consecutive forecast of the distribution of deviation of the down regulation prices from the spot prices

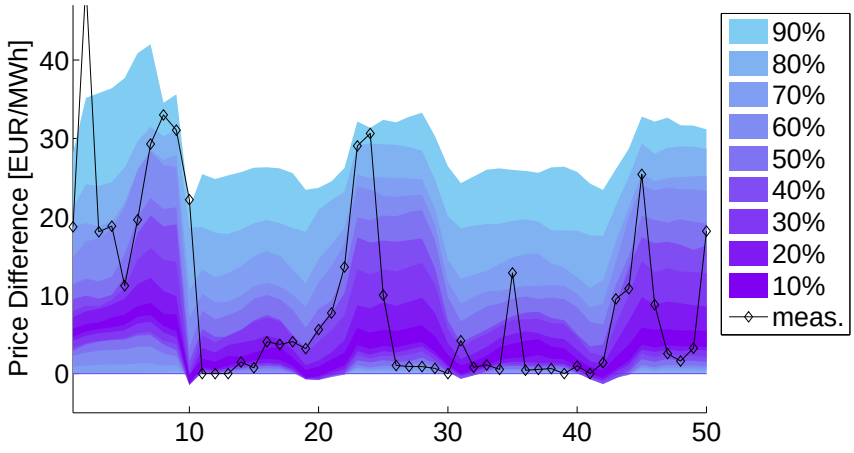

Figure 9.5: 50 consecutive forecast of the distribution of deviation of the up regulation prices from the spot prices

A potential for improving the quantile regression model is to construct a point forecasting model of the regulation prices and thereby transform the problem

into a one similar to the problem analyzed in the previous chapter. Furthermore, since available transmission capacity on the interconnections to Germany, Norway and Sweden leaves the TSO with a lot more possibilities when it comes to regulation, some knowledge of this availability could be of great help when predicting both the regulation direction and also the price.

CHAPTER 10

Conclusion

Players on deregulated energy markets have to account for numerous factors when bids are made for the upcoming day. Throughout this thesis, the goal has been to come up with a model which enables one to gain information about one of these factors - The spot price during the period bids are about to be placed for. The road taken towards this goal is to combine different types of statistical models into a total model, which captures the non-linear and volatile behavior of the prices and the dependence of the prices on external factors and previous prices.

By analyzing the forecasting ability of the static seasonal models described in Chapter 5, it becomes clear that both recursive parameter estimation and reliance on external variables is a desired property of the model. By introducing recursive parameter estimation, the model will slowly adapt to the ever changing market characteristics and thereby maintain its forecasting ability as time passes. Furthermore, since dependence on the most recent known values fades rapidly, accounting for the non-linear relationship between the spot prices and external factors, allows one to obtain quality forecasts for a much larger look-ahead time than otherwise possible.

The analysis in Chapter 6, of the relation between external factors and the spot prices demonstrates the severe impact of predicted penetration of wind power in the system on not only the level of the spot prices but also the distributional properties of the spot prices. The spot price is shown to decrease with increased wind power penetration and intra-day price variations diminish to

some extent. From the analysis, the effects of time factors is also evident, both in form of shift in mean and distributional properties. The conditional parametric approach used to capture the level variation of the prices turns out to be of great value as it caches the dependence on both wind power penetration and time factors well.

When the effects of wind power penetration and time factors are combined with the recursive parameter estimation, the forecasting ability of the resulting model has to be regarded as satisfactory, as shown in Chapter 7, making it very interesting to investigate both its performance and value in practice. The model's performance is quite stable for all forecasting horizons from 13 to 44 hours apart from the superior performance of the two shortest horizons. It can therefore be deemed plausible that the horizon can be extended with similar performance as long as reliable wind power forecasts exists.

When it comes to estimating the uncertainty of the forecasts made by the recursive model, the quality of prediction intervals, derived using quantile regression, is superior to the quality of the ones derived under the classic assumption of gaussianity. The use of conditional distributions for estimation of prediction intervals must therefore considered as essential. The crucial role of wind power forecasts when constructing such conditional distributions is clearly demonstrated by the analysis in Chapter 6 and by the quality of the prediction intervals derived in Chapter 8. Despite the paramount role of wind power, other variables must be considered as well for obtaining a complete description of the uncertainty. Here the forecasted mean of the spot prices was used with good results, however the use of other factors, such as generation combination in surrounding areas and current possibilities in trading between areas, should not be excluded as a ground for improved quantile forecasts. Furthermore, when the uncertainty is to be modeled on market with a wider generation mix, the impact of wind power on the uncertainty could very well decrease to some extent. However, the effects will never completely diminish and therefore it must be concluded that assumptions made in [26] and [45] about independence between wind power generation and the price do not hold.

Despite the very interesting results the use of quantile regression leads to, the method's shortcomings of when it comes to model variables with heavy tailed distributions is apparent in all applications presented in this thesis. A potential way of obtaining a better quality of the individual quantile estimates would be to adjust the model setup for each quantile instead of aiming at the overall quality of the quantiles. Furthermore, the fact that sharpness and reliability is bounded by the quality of the point forecasts, from which the explanatory variable originates make it desirable to improve the point forecasts as well.

In this thesis, the focus has primarily been on the Western Danish price area (DK-1) at Nord Pool. An area which has many characteristics of the future in

deregulated electricity markets. Although the share of wind power in DK-1 is much larger than the average, it is well plausible that the models derived here could give satisfactory performance in other areas, penetrated by wind power to some extent, as well - Due to the varying availability of the fuel and the extremely low marginal costs. As shown in Appendix A, the static seasonal model performs very similar in DK-2, and despite somewhat less penetration of wind in that area, the penetration is still considerable. So performance improvements could presumably be obtained by accounting for wind power penetration and by adaptive parameter estimation.

The severe impact of wind power forecasts on all behavior of the electricity prices is also interesting to consider with the future generation mix in Denmark in mind, where the goal is to increase the share of wind power significantly. With current market structure of marginal bidding, the frequency of hours where the spot price is zero is bound to increase in harmony with increased wind power penetration in the system in the same manner as has been demonstrated here. This will then not only result in increased volatility of the already very volatile spot prices, but also make investments in new energy generation capacity less attractive due to longer repayment of initial capital. The impact of wind power on the price making at the power markets should therefore be given careful thought when the possibility of having 50% of Denmark's energy consumption generated by wind turbines is discussed. In this context it will also be interesting to monitor the market's response to energy generated by other renewable sources reaching the status of having significant share of the energy generation. Whether they will level out the effect of increased wind power or magnify them will play an important role in future development of the market's structure

The preliminary analysis presented in Chapter 9, on how modeling of the regulation market looks promising. Although it is clear that a lot of tuning can be done for gaining better performance of the models, they seem to be quite accurate. However, more analysis is required before any concrete conclusions can be drawn about the predictability of the regulation market and how forecasts for the regulation market should be obtained in detail. Nevertheless, it can safely be stated that this analysis could be a useful pointer towards a reliable and complete model of the regulating market.

Considering all aspects of the thesis, the models and analysis presented here are an important step towards a complete model of the energy market in Denmark and towards a general framework for modeling the price making at deregulated energy markets. Combined, the two layer point forecasting model presented in Chapter 7 and the quantile regression model in Chapter 8, provide reliable information about the full distribution of the spot prices several hours ahead. However, a lot remains to be done, not only in model development but also do prospective applications have to be investigated.

10.1 Future Work

Since energy markets in Europe are strongly connected and since very few market players only participate in one market area, extending the spot price model to cover more areas would definitely be of interest. In this context beginning with covering both Danish ares and the adjacent Nord Pool areas would be a reasonable first step and later on the model could be expanded to the whole Nord Pool region and towards other markets, such as EEX in Germany and ENDEX in the Netherlands, as well. Furthermore, once individual models of more than one adjacent price area exists, the opportunity of taking a more integrated approach to the modeling the market behavior by taking into consideration the interdependencies of the markets could give interesting and improved results.

All aspects of modeling the regulation market need more thorough examination. Even though it has been said that the models presented in Chapter 9 look promising, other approaches should definitely not be disregarded. For instance, constructing a forward curve-like model for the deviation of the regulation prices from the spot prices could be interesting. Thereby the regulation price could be determined by the spot price and the regulation quantity. Obviously, many other approaches can be taken, however under all circumstances, a reasonable first step will be to define properly in terms of what the regulation prices should be modeled, e.g. in terms of the actual regulation price, the difference from the actual spot price or in terms of the deviation from the forecasted spot price.

In this thesis the main focus has been on the spot market and some attention has been paid to the regulation market. However market players have other possibilities as well, Elbas works as a kind of an intermediate step between the spot market and the regulation market and the so called availability market is open between 09.30 and 10:00 every day. By having models of them along with forecasts of the spot and regulation market, the market players would have a better possibility of utilizing all trading opportunities in an optimal way. Furthermore when models of these markets have been found, the horizons shorter than 13 hours become relevant. Expanding the horizon beyond the 44 hour upper limit here would be of interest.

Potential grounds for improving the models presented in this thesis include a revision of the conditional parametric model and better uncertainty models. Although in theory, the conditional parametric model could be extended to an infinite number dimensions, this is not the case in practical applications since distance between data points increases with every dimension added. Replacing the current variables or combining them into one to make way for new ones would therefore be a more plausible road to success. In addition to the possible ways of increasing the quality of the uncertainty estimation for the spot market, mentioned in the previous section, a fully described method for

deriving joint conditional probability densities would presumably give a great opportunity for improvement of all the models. In the form of a joint conditional density of wind power penetration and prices for the spot model and in form of the joint conditional density of the spot prices and the regulation prices for the regulation market model. However, methods for deriving such joint densities have not been fully described yet.

Another aspect not touched at all in this thesis it the potential usage of these forecasts. Since the value of information depends entirely on how it is used, a thorough investigation of how the forecasts will serve best as a bidding aid must be conducted. In this process an essential first step will be to formulate an optimization problem, tailor made for each end user. Only then the actual potential value of the information can be assessed.

APPENDIX A

Static Seasonal Model for the Spot Prices in DK-2

As mentioned earlier, it is desirable to expand the model to cover more areas. In parallel to creating the static seasonal model for the spot prices in DK-1 presented in Chapter 5, the same types of models were constructed for the spot prices in DK-2. Figures showing the same as the ones in Chapter 5 are displayed, however since all results are extremely similar to the ones obtained for DK-1, comments are kept at minimum since everything that was said regarding DK-1 in Chapter 5 applies for DK-2 as well.

A.1 Seasonal ARMA Models

A.1.1 Choice of Models

Despite that a bit higher correlation is detected between the range and the mean for DK-2 (see Figure A.1), it is still concluded that transforming the data is not required. Then by viewing the ACF and PACF of the spot prices, displayed in Figure A.2, it is concluded that the models described by Equations 5.1 and 5.2 are also appropriate for modeling the spot price in DK-2.

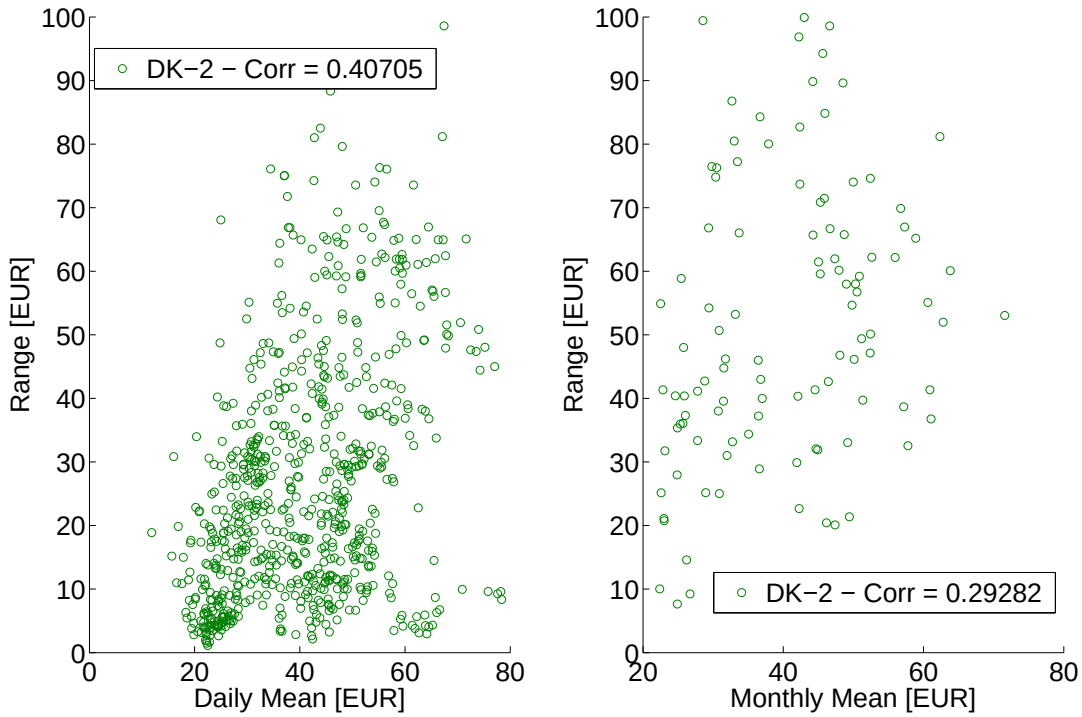

Figure A.1: Range-Mean plot of the Price series for DK-2, both daily mean (left) and weekly mean (right)

A.1.2 Evaluation of Forecasting Performance

The forecasting performance of the model when it is used in DK-2, displayed in Figures A.3 and A.4, is almost identical to what was seen for DK-1. Therefore, everything said in Section 5.1.2 applies for forecast in DK-2 as well.

Furthermore, Figures A.5 and A.6 show that the residuals behave in the same manner on both areas. This is all according to expectations, since the prices in the two areas are very related.

A.1.3 Forecasting

Evaluation of the same estimated prediction intervals as were constructed for DK-1, shown in Table A.1 leads to the same conclusions as were drawn in

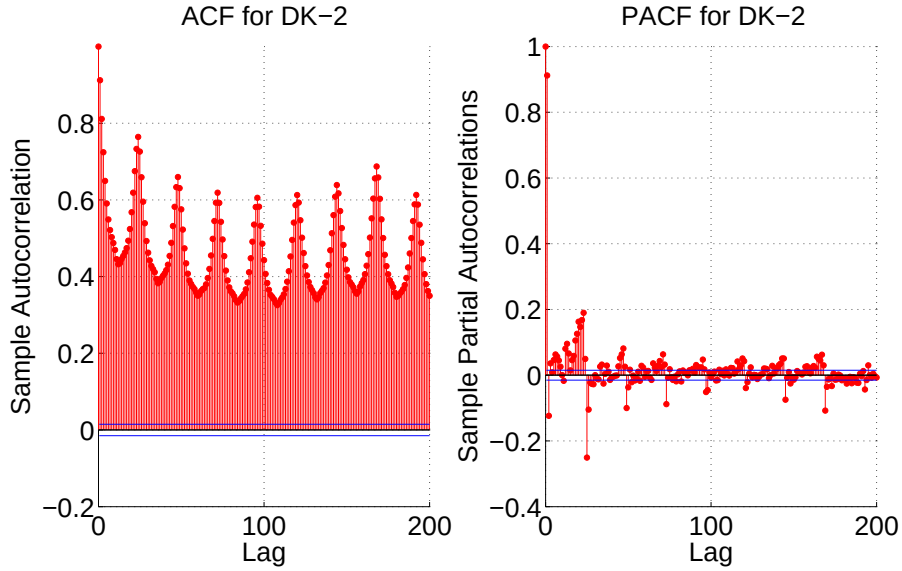

Figure A.2: ACF (left) and PACF (right) for the price in DK-1

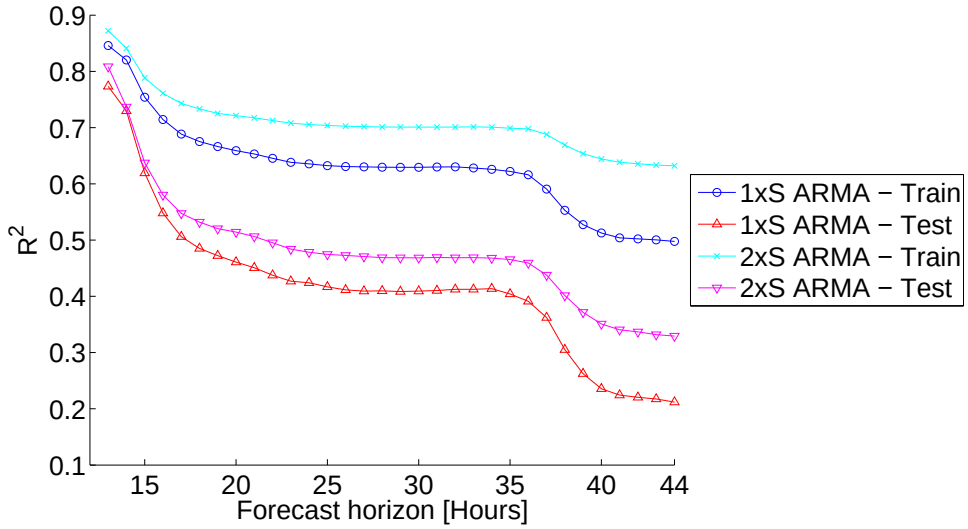

Figure A.3: R^2 for different ARMA models of the spot price in DK-2

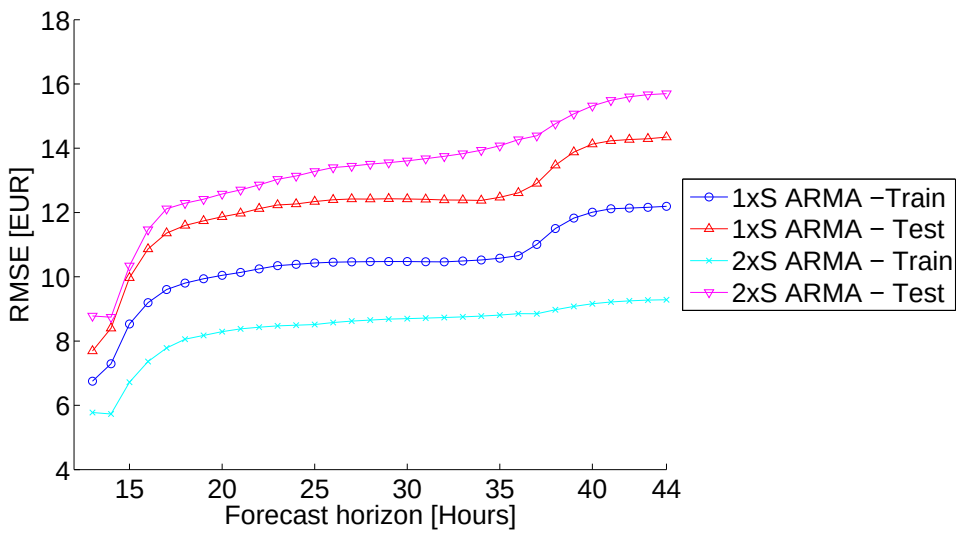

Figure A.4: RMSE for different ARMA models of the spot price in DK-2

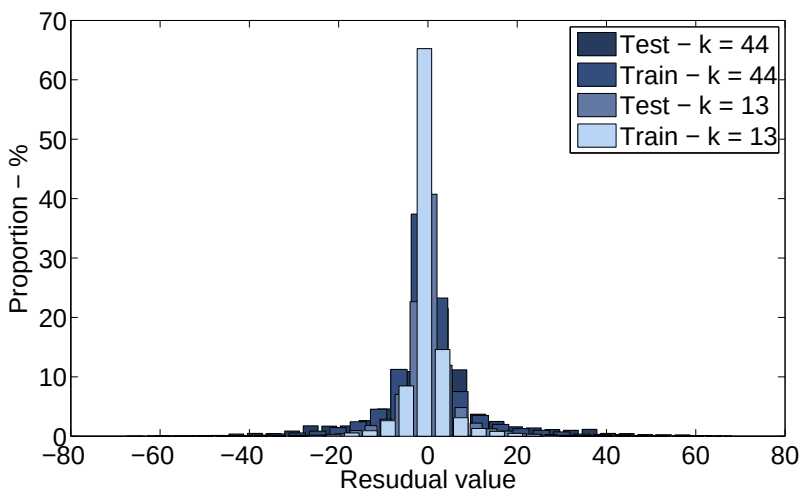

Figure A.5: Residual histograms for 13 and 44 hour predictions in DK-2

Section 5.1.3. The difference between the prediction intervals for the areas is mainly that estimates for DK-2 reflect the real uncertainty somewhat worse than was seen for DK-1.

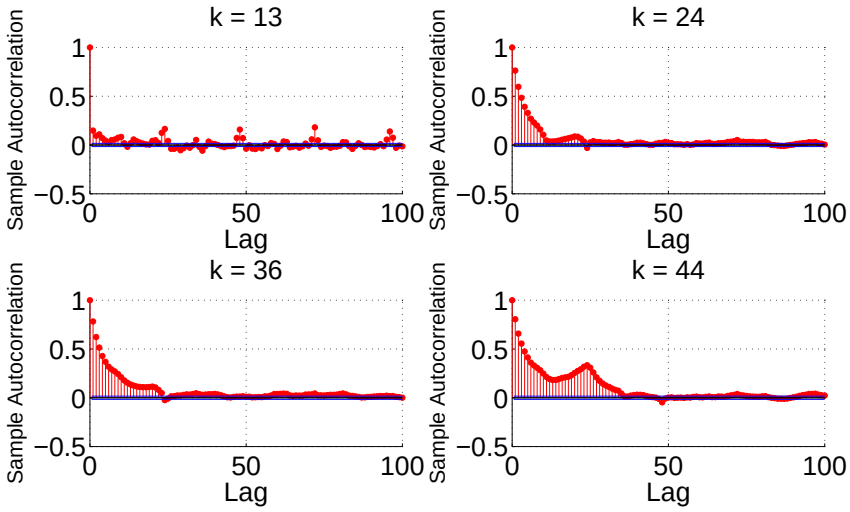

Figure A.6: ACF of the residuals of the training set for 13, 24, 36 and 44 hour forecasts in DK-2

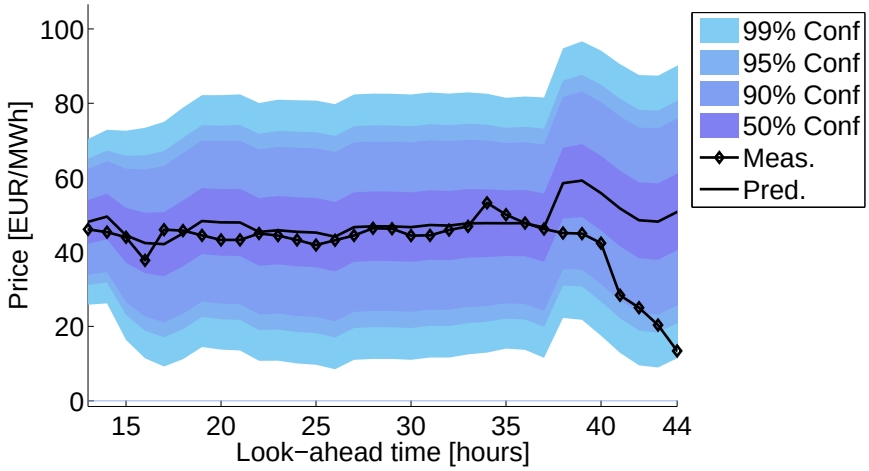

Figure A.7: Forecasts 13 - 44 hrs ahead in DK-2

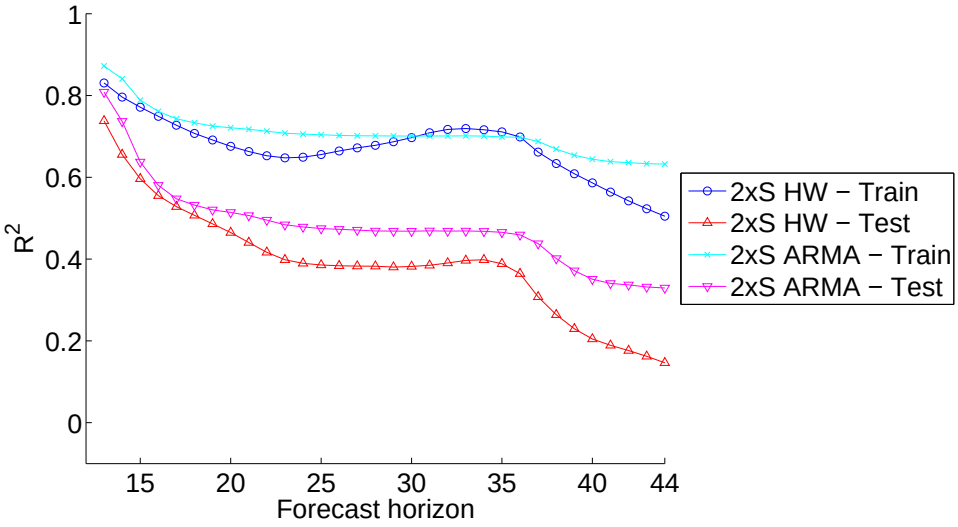

Figure A.8: R^2 for the two Holt Winters models and the double seasonal ARMA model of the spot price in DK-2

A.2 Holt-Winters model

Finally, the Holt-Winters model described by Equations 5.4 - 5.8 is applied to the DK-2 data set. As can be seen from Figures A.8 and A.9, this leas to similar results as were obtained in the final section of Chapter 5.

Table A.1: Proportion of observations inside prediction intervals for DK-2

	99%	95%	90%	50%
Training set	96.37%	93.74%	91.61%	75.03%
Test set	93.31%	89.12%	86.10%	67.02%

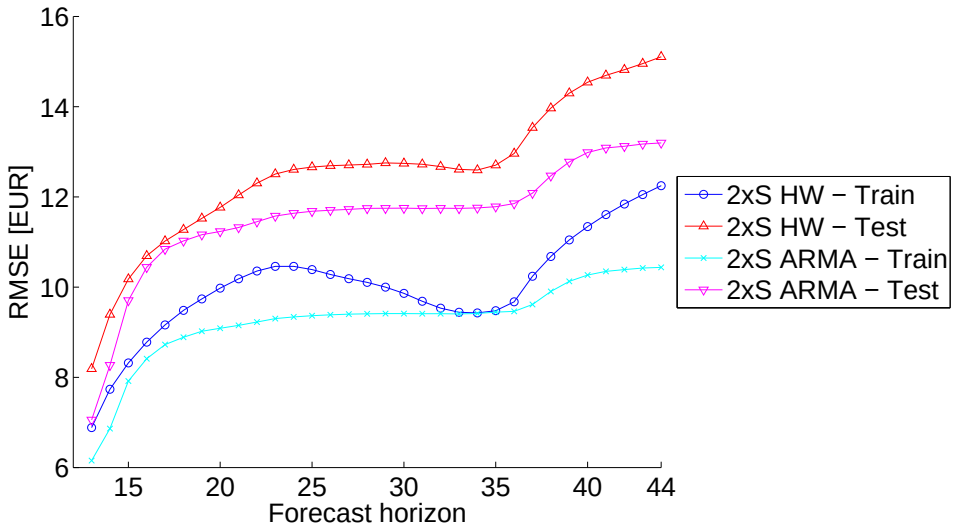

Figure A.9: RMSE for the two Holt Winters models and the double seasonal ARMA model of the spot price in DK-2

Appendix B

Example of Codes

In this chapter, examples of the code written in Matlab and R are given. One from each program. Since the purpose of displaying the code is only to give examples of how the programs are used, the remaining code is omitted. It is however available upon request.

B.1 RPLR model i Matlab

```
function [Yt,Yt_hat, theta, res, x_ext,conf] = Amodele(price, external,lambda,k,n)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% RPLR model:
% Inputs:
%         - price: A matrix with the price series in the first column and
%         the series lagged as explanatory variables in the remaining
%         columns
%         - external: A matrix with the external factors
%         - lambda: The forgetting factor
%         - k: length of forecastin horizon
%         - n: no. of forecasts to be made
%Outputs:
%         - Yt: The actual time series
%         - Yt_hat: The forecasts of Yt
%         - theta: A matrix with the model parameters at each time
```

```
%         - res: The residuals
%         - x_ext: Output from LWPR
%         - conf: Confidence intervals from Yt_hat. So Yt_hat +/- conf are
%         the actual confidence intervals
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% First two months of the data set taken for initial estimation of the LWPR
s = size(external,1) - size(price,1);
endP = 1416 - s;

priceI = price(1:endP,:);
priceII = price(endP+1:end,:);

extI = external(s+1:endP+s,:);
extII = external(endP+s+1:end,:);

%Initializing
if nargin == 2
    n = size(priceII,1);
    lambda = 1;
elseif nargin == 3
    n = size(priceII,1);
    lambda = 1;
elseif nargin == 4
    n = size(priceII,1);
end

if k<24
    lag = [23 47];
else
    lag = [47 71];
end

noPar = size(priceI,2) + 3;
theta = 0.5*ones(noPar,1);
Pt = 1000*eye(noPar);
Xt = [];
Yt = [];%priceII(1,1);
et = zeros(1,48+k);
tVal = [2.5758 1.96 1.6449 0.6745];
x_ext = zeros(n,1);
Yt_hat = zeros(n,1);
res = zeros(n,1);
conf = zeros(n,4);

for t = 1:n
    % Conditional parametric model
```

```
    x_ext(t) = lwpr3D(extII(t,1), extII(t,2),extII(t,3),...
        [extI, priceI(:,1)],2,0.5);

    % Construct design matrix
    Xt = [Xt; priceII(t,2:end),x_ext(t), et(end-k+1),...
        et(end-lag(1)), et(end-lag(2))];

    %Forecast
    Yt_hat(t) = Xt(t,:)*theta(:,t);
    Yt = [Yt;priceII(t,1)];

    % Calculate prediction error
    et = [et, Yt(t)-Yt_hat(t)];
    res(t,1) = Yt(t) - Yt_hat(t);

    % Calculate the condidence intervals
    sigma = std(res);
    conf(t,:) = tVal.*sigma;

    % Update matrix for LWPR
    priceI = [priceI;priceII(t,:)];
    extI = [extI;extII(t,:)];

    % Update theta
    Kt = (Pt*Xt(t,:)')/(1+Xt(t,:)*Pt*Xt(t,:)');
    Pt = 1/lambda.*(Pt - (Pt*Xt(t,:)'*Xt(t,:)*Pt)/(lambda+Xt(t,:)*Pt*Xt(t,:)'));
    theta(:,t+1) = theta(:,t) + Kt*(Yt(t) - Xt(t,:)*theta(:,t));

end
```

B.2 SVM model in R

```
library('kernlab')

# Data import and construction
ph <- read.table(file="planhand.txt",sep=",",dec=".",col.names=c("hour","DK-D","DK-N","DK-SP"),colClasses=c("facto
logd <- read.table(file="logregdata1.txt",sep=",",dec=".",col.names=c("M","W","H","Sp","RN","RO","RS"),colClasses=
prCn <- c("H.1","H.2","H.3","H.4","H.5","H.6","H.7","H.8","H.9","H.10","H.11","H.12","H.13","H.14","H.15","H.16","
prf <- read.table(file="prf.txt",sep=",",dec=".",col.names=prCn)
wppt <- read.table(file="wpp1_32_2.txt",sep=",",dec=".",col.names=prCn)

RN <- as.factor(sign(logd$RN-0.5))
RO <- as.factor(sign(logd$RO-0.5))

DataSet12 <- data.frame(RN,RO,M=logd$M,W=logd$W,H=logd$H,DK.D=ph$DK.D,DK.N=ph$DK.N,DK.SP=ph$DK.SP,wppt12=wppt$H.1,
DataSet112 <- data.frame(RN,M=logd$M,W=logd$W,DK.D=ph$DK.D,DK.N=ph$DK.N,DK.SP=ph$DK.SP,wppt12=wppt$H.1,prf12=prf$H
```

```
d <- dim(DataSet12)
nn <- round(0.5*d[1])
train <- DataSet12[1:nn,]
test <- DataSet12[(nn+1):d[1],]

# SVM using the kernlab package - Static models
modelKsvmRN <- ksvm(RN~M + W + wppt12 + prf12 + DK.D + DK.N + DK.SP,
  data = train, method = "C-svc", kernel = "rbfdot", kpar="automatic",
  C = 10, prob.model=TRUE)

modelKsvmRO <- ksvm(RO~M + W + wppt12 + prf12 + DK.D + DK.N + DK.SP,
  data = train, method = "C-svc", kernel = "rbfdot", kpar="automatic",
  C = 10, prob.model=TRUE)

prKsvmRN <- predict(modelKsvmRN,test)#,type='probabilities'
perfKsvmRN <- sum(prKsvmRN==test$RN)/nn
prKsvmRO <- predict(modelKsvmRO,test)#,type='probabilities'
perfKsvmRO <- sum(prKsvmRO==test$RO)/nn


# SVM using the kernlab package - Sliding window

# Window size: 300 obs.
nTrain <- 300
nTest <- 1

pTrSvmRNw300 <- matrix(0,(d[1]-nTrain),1)
pTrSvmROw300 <- matrix(0,(d[1]-nTrain),1)

prSvmRN300 <- matrix(0,(d[1]-nTrain),1)
prSvmRO300 <- matrix(0,(d[1]-nTrain),1)

pprSvmRN300 <- matrix(0,(d[1]-nTrain),2)
pprSvmRO300 <- matrix(0,(d[1]-nTrain),2)


for(i in 1:(d[1]-nTrain)){
  traintmp <- DataSet12[i:(nTrain+i-1),]
  testtmp <- DataSet12[(i+nTrain),]

  modelKsvmRN <- ksvm(RN~W + wppt12 + prf12 + DK.D + DK.N + DK.SP,
    data = traintmp, method = "C-svc", kernel = "rbfdot", kpar="automatic",
    C = 10, prob.model=TRUE)

  modelKsvmRO <- ksvm(RO~W  + wppt12 + prf12 + DK.D + DK.N + DK.SP,
    data = traintmp, method = "C-svc", kernel = "rbfdot", kpar="automatic",
    C = 10, prob.model=TRUE)
```

```
  predTrRN <- fitted(modelKsvmRN)
  predTrRO <- fitted(modelKsvmRO)

  pTrSvmRNw300[i] <- sum(traintmp$RN==predTrRN)/nTrain
  pTrSvmROw300[i] <- sum(traintmp$RO==predTrRO)/nTrain

  prSvmRN300[i] <- predict(modelKsvmRN,new=testtmp)
  prSvmRO300[i] <- predict(modelKsvmRO,new=testtmp)

  pprSvmRN300[i,] <- predict(modelKsvmRN,new=testtmp,type='probabilities')
  pprSvmRO300[i,] <- predict(modelKsvmRO,new=testtmp,type='probabilities')
}

idx1RN <- which(prSvmRN300==1)
idx2RN <- which(prSvmRN300==2)

prSvmRN300[idx1RN] <- -1
prSvmRN300[idx2RN] <- 1

idx1RO <- which(prSvmRO300==1)
idx2RO <- which(prSvmRO300==2)

prSvmRO300[idx1RO] <- -1
prSvmRO300[idx2RO] <- 1

perfWRN300 <- sum(DataSet12$RN[(nTrain+1):d[1]]==prSvmRN300)/(d[1]-nTrain)
perfWRO300 <- sum(DataSet12$RO[(nTrain+1):d[1]]==prSvmRO300)/(d[1]-nTrain)

preds300 <- data.frame(RN=prSvmRN300,RO=prSvmRO300)
write.table(preds300,file="preds300Svm.txt")

prSvmRO300_H <- prSvmRO300
prSvmRN300_H <- prSvmRN300
pprSvmRO300_H <- pprSvmRO300
pprSvmRN300_H <- pprSvmRN300

perfWRN300_H <- perfWRN300
perfWRO300_H <- perfWRO300
```

Bibliography

1. Dereck W. Bunn, editor. *Modelling prices in competitive electricty markets*. John Wiley & Sons, 2004.

2. Lennart Carlsson. International power trade - the nordic power pool. Technical report, The World Bank Group, January 1999. Note no. 171.

3. Christopher Chatfield. *Time Series Forecasting*. Chapman & Hall / CRC, London, UK, 2000.

4. Antonio J. Conejo, Javier Contreras, Rosa Espínola, and Miguel A. Plazas. Forecasting electricity prices for a day-ahead pool-based electric energy market. *International Journal of Forecasting*, 21:435–462, 2005.

5. Jesús Crespo Cuaresma, Jaroslava Hlouskova, Stephan Kossmeier, and Michael Obersteiner. Forecasting electricity spot-prices using linear univariate time-series models. *Applied Energy*, 77:87–106, 2004.

6. Shi-Jie Deng and Wenjiang Jiang. *Modelling prices in competitive electricty markets*, chapter 7 - Quantile-Based Probabilistic Models for Electricity Prices, pages 161–176. John Wiley & Sons, 2004.

7. Earth Policy Institude. *Earth Policy Institude's website*, September 2007. URL http://www.earth-policy.org.

8. Haukur Eggertsson. The scandinavian elecricity power market and market power. Master's thesis, Technical University of Denmark, DTU, Kgs. Lyngby, Denmark, 2003.

9. Lars Elden, Linde Wittmeyer-Koch, and Hans Bruun Nielsen. *Introduction to Numerical Computation*. Studentlitteratur, Lund, Sweden, 2004.

10. Energinet. *Energinet's website*, August 2007. URL http://www.energinet.dk.

11. Svend W. Enevoldson, Poul Alberg Østergaard, Poul Erik Morthorst, and Rune Moesgaard. Vindkraftens betydning for elprisen i danmark. Technical report, IBT-Wind, 2006.

12. European Nuclear Society. *European Nuclear Society - Encyclopedia*, September 2007. URL http://www.euronuclear.org/info/encyclopedia.

13. Fingrid Oyj. *Fingrid's website*, September 2007. URL http://www.fingrid.fi.

14. Geoffrey Gimmett and David Stirzaker. *Probability and Random Processes*. Oxford University Press, Oxford, UK, 2001.

15. Alicia Mateo González, Antonio Mu noz San Roque, and Javier García-González. Modeling and forecasting electricity prices with input/output hidden markov models. *IEEE Transactions on Power Systems*, 20, 2005.

16. Christian Viller Hansen. Efficient numerical methods for adaptive quantile regression. Master's thesis, Technical University of Denmark, DTU, Kgs. Lyngby, Denmark, 2007.

17. Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning - Data Mining, Inference and Prediction*. Springer, New York, USA, 2001.

18. John Bagterp Jørgensen, Kaj Madsen, Hans Bruun Nielsen, and Marielba Rojas. *Introduction to Optimization and Data Fitting*. IMM, DTU, Kgs. Lyngby, Denmark, 2006.

19. Alexandros Karatzoglou, David Meyer, and Kurt Hornik. Support vector machines in r. *Journal of Statistical Software*, 15, 2006.

20. Úlfar Linnet. Tools supporting wind energy trade in deregulated markets. Master's thesis, Technical University of Denmark, DTU, Kgs. Lyngby, Denmark, 2005.

21. Henrik Madsen. *Quantile Regression - Slide show*, November 2006.

22. Henrik Madsen. *Time Series Analysis*. Chapman & Hall / CRC, London, UK, 2007.

23. Henrik Madsen and Jan Holst. *Modelling Non-Linear and Non-Stationary Time Series*. IMM, DTU, Kgs. Lyngby, Denmark, 2000.

24. Henrik Madsen, Pierre Pinson, Torben S. Nielsen, Henrik Aa. Nielsen, and George Kariniotakis. Standardizing the performance evaluation of short-term wind power prediction models. *Wind Engineering*, 29(6):475–489, 2005.

25. Paras Mandal, Tomonobu Senjyu, and Toshihisa Funabashi. Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated markets. *Energy Conversion & Management*, 47:2128–2142, 2006.

26. Julija Matevosyan and Lennart Söder. Minimization of imbalance cost trading wind power on the short-term power market. *IEEE Transaction on Power Systems*, 21:1396–1404, 2006.

27. Peter Meibom, Poul Erik Morthorst, Lars Henrik Nielsen, Christoph Weber, Kai Sander, Derk Swider, and Hans Ravn. Power system models, a description of power markets and outline of market modelling in wilmar. Technical report, Risø National Laboratory, December 2003. R-1441.

28. Jan Kloppenborg Møller. Modeling of uncertainty in wind energy forecast. Master's thesis, Technical University of Denmark, DTU, Kgs. Lyngby, Denmark, 2006.

29. Jan Kloppenborg Møller, Henrik Aalborg Nielsen, and Henrik Madsen. Time-adaptive quantile regression. *Computational Statistics and Data Analysis*, 51:1292–1303, 2008.

30. Rune Moesgaard and Poul Erik Morhorst. The impact of wind power on electricity prices in denmark. In *EWEC 2008, European Wind Energy Conference, Business and Policy Track*, Brussels, Belgium, 2008.

31. Torben Skov Nielsen, Henrik Aalborg Nielsen, and Henrik Madsen. Prediction of wind power using time-varying coefficient functions. In *15th IFAC World Congress*, Barcelona, Spain, 2002.

32. F.J. Nogales and Antonio J. Conejo. Electricity price forecasting through transfer function models. *Journal of Operational Research Society*, 57:350–356, 2006.

33. Nord Pool ASA. *Trade at Nord Pool's Financial Market*, 2004.

34. Nord Pool ASA. *Nord Pool´s website*, August 2007. URL http://www.nordpool.com.

35. Nord Pool ASA. *The Nordic Power Market: Electricity Power Exchange Across National Borders*, 2004.

36. Nord Pool ASA. *Trading Rules for Financial Electricity Contracts and Certificate Contracts*, 2007.

37. Nord Pool Spot AS. *Continuous trading at Nord Pool Spot's Elbas market*, .

38. Nord Pool Spot AS. *Introduction: Block bidding in the Elbas trading system*, .

39. Nord Pool Spot AS. *Bidding in Nord Pool Spot's Elspot Market*, 2006.

40. Nord Pool Spot AS. *Calculation of System- and Area prices*, 2006.

41. Nord Pool Spot AS. *Nord Pool Spot's website*, August 2007. URL http://www.nordpoolspot.com.

42. Nord Pool Spot AS. *The Nordic Power Exchange and the Nordic Model for a Liberalised Power Market*, July 2006.

43. Nordel. *Nordel's website*, August 2007. URL http://www.nordel.org.

44. NordREG - Nordic Energy Regulators. *Development of a Common Nordic Balance Settlement*, February 2006.

45. Pierre Pinson, Christophe Chevallier, and George N. Kariniotakis. Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE Transaction on Power Systems*, 22 (3):1148–1156, 2007.

46. Pierre Pinson, Henrik Aalborg Nielsen, and George Kariniotakis Jan K. Møller, Henrik Madsen. Nonparametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy*, 10 (6):497–516, 2007.

47. Philip Bendorff Røpcke. Forecast af handel på dk1 balancemarkedet. Technical report, Markedskraft Danmark A/S, April 2008.

48. Claudio M. Ruibal and Mainak Mazumdar. Forecasting the mean and the variance of electricity prices in deregulated markets. *IEEE Transactions on Power Systems*, 23 (1):25–32, 2008.

49. James W. Taylor. Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of Operational Research Society*, 54:799–805, 2003.

50. James W. Taylor and Patrick E. McSharry. Short-term load forecasting methods: An evaluation based on european data. *IEEE Transactions on Power Systems*, 22:2213 2219, 2008.

51. Uranium Information Centre. *Uranium Information Centre's website*, September 2007. URL http://www.uic.com.au.

52. Iivo Vehviläinen and Tuomas Pyykkönen. Stochastic factor model for electricity spot price - the case of the nordic market. *Energy Economics*, 27:351–367, 2005.

53. Wikipedia. *Wikipedia - The Free Encyclopedia*, September 2007. URL http://www.wikipedia.org.