NTNU
Innovation and Creativity

Modelling the endogenous and exogenous factors of the Nord Pool Spot price using state space models

Lina Ekern, Sebastian Brelin and Morten Adrian Lien

December 2012

PROJECT THESIS
Norwegian University of Science and Technology
Faculty of social Science
Department of Industrial Economics and Technology Management

Supervisor: Stein-Erik Fleten

(Aarrestad and Hatlen, 2015) (Spot, 2014) (Mirza and Bergland, 2012) (NordReg, 2014) (Unander et al., 2004) (Bergersen et al., 2004) (Henley and Peirson, 1997) (Seljom et al., 2011) (Spot, 2012) (DK, 2016) (Finland, 2016) (Povh et al., 2010) (ETS, 2016a) (ETS, 2016b) (Proietti, 2002) (Alexander, 2008b) (Harvey, 2006) (Kaaresen and Husby, 2000) (Haldrup and Nielsen, 2006) (Chen and Bunn, 2010) (Grull and kiesel, 2012) (Johnsen, 2007) (Johnsen, 2008) (Pettersen, 2012) (**?**) (Bunn et al., 2016) (Escribano A.and Pea, 2011) (Hagfors et al., 2016) (Haldrup and Nielsen, 2005) (Hamilton, 1989) (Harvey, 1989) (Higgs and Worthington, 2008) (Huurman et al., 2007) (Karakatsani and Bunn, 2008) (Knittel and Roberts, 2005) (Koopman et al., 2007a) (Koopman et al., 2007b) (Lucia and Schwartz, 2002) (Lland and Dimakos, 2006) (Weron and Misiorek, 2008) (Wei, 2006) (Welch and Bishop, 2006) (Hoover and Perez, 2002) (**?**)

Preface

This project thesis was conducted at the Norwegian University of Science and Technology (NTNU), Department of Industrial Economics and Technology Management.

We would specially like to thank our supervisor, Professor Stein-Erik Fleten for his time and guidance.

Abstract

The study of the behaviour of electricity prices provides the basis of which market participants make a variety of decisions. Hence, precise modelling of electricity prices is of high importance to market participants. However, interactions of several complexities, such as the non-storable nature of electricity, highly inelastic demand curves and exogenous effects from price drivers, implies that modelling the electricity spot price is not trivial. In this paper, we apply the structural time series model framework and decompose the spot price into unobservable components that are allowed to vary with time. These components include endogenous factors as well as sensitivities to exogenous factors. We find that modelling the electricity spot price in terms of fundamental price driver can capture specific dynamic variation the spot price. An interesting observation is that the exogenous effects of fundamental price drivers on the development of the price appear constant.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Since the early 1990s, when the Nordic countries began the deregulation of electricity markets, spot electricity prices have been characterized by seasonal fluctuations, volatility changes, mean-reversion, spikes, and jumps. This complexity arises from a convolution of market characteristics. Firstly, the non-storable nature of electricity requires a constant balance between supply and demand in order to have power system stability (Kaminski, 2013). Considering the diverse plant technologies available for dispatch, electricity markets thus have nonlinear and convex supply curves (Karakatsani and Bunn, 2008). Secondly, the price elasticity of demand is highly inelastic in the short-term, implying that there are few options available to the consumer in response to changes in the electricity price. As such, demand is a function of temperature, weather and consumer patterns (Mirza and Bergland, 2012). Thirdly, electricity must be delivered through a transmission network subject to complicated and interacting physical limitations. Thus, in times of scarcity and high demand, the producers with reserve margins have market power. These producers can set asking prices well above marginal costs, thereby contributing to occasional price jumps (Bunn et al., 2016). Lastly, there is strong evidence that exogenous factors have substantial impact on the electricity spot price (Chen et al., 2010). A clear implication of the interaction of all these complexities is that modelling the electricity spot price is not trivial.

The study of the behaviour of electricity prices provides the basis of which market participants make a variety of decisions. In the short run, modelling spot prices are important for optimization of day-to-day market operations, such as bidding. In the medium run, analyses are typically used for risk management and derivatives pricing, whereas in the long-run, pricing models are important for profitability analysis and investment planning. As such, precise modelling of electricity prices is of high importance to market participants.

Motivated by the above market characteristics, this paper presents a concise discussion on the modelling of both endogenous and exogenous factors of the Nord Pool spot electricity price in NO5. Applying the structural times series model framework introduced by Harvey (1989), we develop a model based on an interpretable decomposition of the electricity price into trend, seasonal, intervention and explanatory components. This approach does not require the differencing of the electricity spot price time series to a stationary process, because the non-stationary properties of the spot price time series are formulated explicitly by the selected components in the decomposition. The main empirical findings of our analysis can be summarised as follows: The salient features of the electricity spot price is well captured by directly formulating the model in terms of fixed seasonal patterns, fixed level and a slight slope. Further, our results suggest that modelling the electricity spot price in terms of fundamental price drivers, can capture specific dynamic variation the spot price. An observation of particular interest, is that the exogenous effects from most price drivers have been constant the last 14 years. This implies that allowing for time-varying sensitivities to exogenous price drivers not will capture additional dynamic variations in the price distribution.

The rest of this paper is organized as follows: Chapter 2 presents a literature review of previous research related to modelling electricity spot prices and places our paper in the context of two main research areas. Chapter 3 introduces the Nord Pool spot market and presents descriptive statistics for the electricity spot price in price area NO5. A comprehensive review of fundamental price drivers is considered in Chapter 4, together with descriptive statistics for the chosen explanatory variables. Chapter 5 introduces the structural time series model approach and presents the evaluation metrics used for in-sample testing. In Chapter 6 we apply structural time series models to our data and present an optimal model for the electricity spot price. Ultimately, our conclusion and suggestions for further research are presented in Chapter 7.

Chapter 2

Literature review

Our paper can be placed in the context of two main research areas in the field of electricity spot markets: (i) modelling of endogenous factors, and (ii) the impact of exogenous fundamental drivers on electricity prices.

During the past years, many papers have been dedicated to methods for capturing and modelling the endogenous factors of the electricity price. By endogenous factors, we consider intrinsic properties of the electricity price, such as its volatility and high-order moments, seasonality, jumps, and spikes, not taking any exogenous variables into consideration. Contributions of particular interest include Lucia and Schwartz (2002) and Knittel and Roberts (2005), who both propose mean reverting models with deterministic regular patterns. The results, having applications to Nordic electricity spot prices and Californian hourly electricity prices, respectively, highlight the importance of seasonal patterns in the behaviour of electricity prices. Escribano et al. (2011) present a general model that simultaneously considers seasonality, mean reversion, generalised autoregressive heteroscedasticity (GARCH) behaviour and time-dependent jumps. They conclude that electricity prices in deregulated markets are mean reverting, with high volatility clustering (GARCH effects) and with stochastic jump intensities, even after adjusting for seasonality. Kaaresen and Husby (2000) develop a state space model which incorporates latent factors, such as seasonality, trend and mean reversion. This model is applied to the electricity spot price of Nord Pool. A noteworthy finding is that a high number of underlying state factors can cause unidentifiability problems.

The Markov-switching model of Hamilton (1989), also known as the regime switching model, is frequently applied in the context of modelling electricity prices endogenously. This model permits switching between a base regime and higher/lower level regimes, thereby capturing complex dynamic patterns such as jumps. Higgs and Worthington (2008) employ three different models to capture frequent extreme price spikes in the Australian electricity spot market. These models include a basic stochastic model, a mean-reverting model, and a regime-switching model. The results show that the regime-switching model outperforms the other two in terms of capturing extreme price spikes. However, a shortcoming of the regime-switching model is the unrealistic assumption of constant transition probabilities (Hagfors et al., 2016). Moreover, Haldrup and Nielsen (2005) use regime-switching models to capture potential occasional price jumps caused by transmission congestions. They argue that the price behaviour in the Nordic electricity spot market is dependent upon whether the bilateral market is subject to congestion or non-congestion.

Whereas the switching mechanism in Markov-switching models is controlled by an unobservable state variable, quantile methods inherently associate a separate regime with each quantile. This is examined by Bunn et al. (2016), who use a multifactor, dynamic, quantile regression model to capture effects such as mean reversion, spikes and time-varying volatility. In addition, they assess a rich set of exogenous fundamental drivers, such as the prices of gas, coal and carbon, forecasts of demand and excess capacity. These fundamental drivers may have different effects across the quantiles. Next, we will provide a brief review of literature that considers such exogenous fundamental drivers when modelling the electricity spot price. There is strong evidence that exogenous factors have substantial impact on the electricity spot price (Chen and Bunn, 2010). However, defining these fundamental drivers is still a non-trivial issue. Koopman et al. (2007) use a periodic seasonal autoregressive fractionally integrated moving average (ARFIMA) model with GARCH disturbances, to explain the dynamics of daily electricity spot prices in four European electricity markets. Their modelling framework proves particularly successful for Nordic electricity prices, and they conclude that a significant part of the short-term price movements in Nordic electricity prices can be explained by weekly water reservoir levels and daily electricity consumption. Extending ARIMA and GARCH models to include weather forecasts, Huurman et al. (2007) investigate the relation between electricity prices and weather variables, such as air temperature, total precipitation and wind speed. Their empirical results suggest that such weather factors have explanatory power for the day-ahead power spot prices in Nord Pool. Lland and Dimakos (2010) investigate the relationship between daily values of the NO1 area price and the system price using general additive models with exogenous explanatory variables. They find that water reservoir level, Elspot capacity and Elspot net capacity utilization are the most important explanatory variables, and conclude that lower capacity and more flow coincide with higher price spreads.

Mount et al. (2006) introduce a regime-switching model where the transition probabilities are modelled as functions of the load and/or the implicit reserve margin, thereby avoiding the issue of constant transition probabilities mentioned earlier. Using price data from the Pennsylvania New Jersey Maryland (PJM) Power Pool, the results show accurate price spike predictions. However, the model requires accurate reserve margin measurements, which are difficult to obtain (Hagfors et al., 2016). Karakatsani and Bunn (2008) apply three fundamental price models to investigate the impact of exogenous fundamental drivers on the British electricity market. These models include a linear regression model, a Markov-switching regression model, and a state-space model with time-varying explanatory variables. Their empirical findings suggest that price models based on fundamental drivers and their time-varying effects are the most effective and potentially the most useful in practice. Chen and Bunn (2010) also recognise the need to specify a model in the context of a rich set of explanatory variables, which may have nonlinear influences. Thus, they use a special type of regime-switching model, the logistic smooth transition regression model, thereby extending the research on regime-switching models with time-varying transition probabilities performed by Mount et al. (2006). The conclusion is that electricity prices react nonlinearly to demand, margin, carbon, fuel prices and market concentration. Hagfors et al. (2016) use a standard logit model to investigate the relationship between fundamental drivers and extreme spike occurrences in the German day-ahead electricity market. Their results suggest that extreme price occurrence have clear drivers and that wind power is particularly important in relation to negative price occurrences.

This paper extends the research regarding how fundamental factors influence the electricity spot price formation. Our context is the price zone NO5 in the Nord Pool spot market, and the endogenous and exogenous factors in our analysis include seasonality, level, spikes, temperature, reservoir level, oil price, gas price and coal price.

Chapter 3

The Electricity Spot Price - Market and Data

3.1 The Nord Pool Spot Market

In the early 1990s, the Nordic countries decided to deregulate the market for trading of electrical energy, reshaping the landscape of a traditionally monopolistic and government-controlled power sector (Weron and Misiorek, 2008). Today, Nord Pool is Europes leading power market, offering trading, clearing and settlement in both day-ahead markets and intraday markets across nine European countries. The day-ahead market, also referred to as the Elspot market, is the primary market for power trading. Here, contracts are made between seller and buyer for the delivery of power, hour-by-hour the following day. The intraday market, also called the Elbas market, supplements the Elspot market and helps secure necessary balance between supply and demand. Here, hourly contracts are continuously traded in the period between clearance in the day-ahead market and up to one hour before the hour of operation . An important prerequisite for an effective power market is a connected and well-developed power grid with access for all players. Grid management and grid operations are considered natural monopolies, and there are no possibilities for competition within the sector (Aarrestad and Hatlen, 2015). Thus, to prevent grid companies from exploiting their positions, the authorities have put in place strictly regulations of their operations as monopolies. As of today, Statnett is the main grid operator and owner in Norway.

The power market distinguishes between wholesale and end-users. In the wholesale market, large volumes are bought and sold between power producers, power suppliers, brokers, energy companies and large scale consumers. In the Nordic countries, these players trade on Nord Pool Spot. In the end-user market, individual consumers enter agreements to purchase power from a supplier of their own choice. Norways end-user market consists of one third household customers, one third industry and one third medium sized consumers. By medium sized consumers, we consider hotels and chain stores (NordReg, 2014).

Price Formation

Electricity differs from other commodities, because it cannot be easily stored. This implies that a constant balance between supply and demand is required, in order to have power system stability (Kaminski, 2013). The market price of power is determined each day on the Nord Pool Spot exchange. Players who wish to buy or sell power on the Elspot power market send their orders to Nord Pool Spot at least at noon the day before the power is to be delivered to, or withdrawn from, the power grid. At Nord Pool spot, purchase orders are aggregated to a demand curve and sale offers are aggregated to a supply curve. The intersection of these two curves indicate the market price of power for a specific hour, as shown in *Figure 3.1*. This way of calculating the price is referred to as *double action*, because both buyers and the sellers submit orders. At most other auctions, only the bidder places orders.

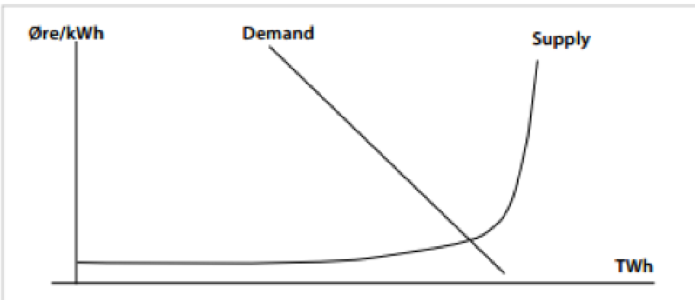

Figure 3.1: Supply and demand Curve

The rising supply curve indicates the amount of power that producers are willing to produce at different prices, thereby reflecting the marginal production cost of power in each type of plant. Hydropower, wind power and nuclear power have the lowest marginal costs in the Nordic region, and are offered at low prices. Gas, biopower and coal-based power, on the other hand, have higher marginal costs and are therefore located further to the right on the supply curve. The electricity spot price for each period is set by the most expensive generator required to satisfy demand (Hagfors et al., 2016).

Area Prices

Nord Pool Spot operates with one system price and several area prices. The system price is determined by the balance between total supply and demand in the Nordic market. This price does not consider the physical capacity constraints in the transmission networks, which is why Nord Pool Spot also operates with area prices. As of today, Norway is divided into five areas, where the prices are determined by area-specific market characteristics. Some areas have surplus of power, while others have deficit. Power system stability is determined by the balance between import and export across the different bidding areas. Insufficient capacity in the grid system for import or export results in different prices amongst areas, and are referred to as bottlenecks. In case of no restrictions in power flow, all bidding areas in the Nordic region have the same price, the system price.

Mirza and Bergland (2012) discusses the issue of power producers taking advantage of bottlenecks in the grid system. By deliberately choosing to produce less, these producers offer power at a higher price, thereby forcing their area to import power. If the amount of needed import exceeds the area-specific import capacity, the area will have to buy power at the high power price (Mirza and Bergland, 2012).

3.2 Descriptive Statistics

Our data consists of the Nord Pool electricity spot price (EUR/MWh) in Bergen, from 2002 until late October 2016. Due to several changes in the price zones since the beginning of our data period, we use the price of whichever zone Bergen has been located in. As of today, Bergen is located in NO5. During the rest of the paper we will therefore refer to Bergen as NO5. However, price data from before 15th of March 2010 is equivalent to NO1. We use data with a monthly resolution, because this enables us to capture seasonality effects. The development of the electricity prices is shown in *Figure 3.2*, together with its logarithm. In *Figure 3.3*, we present the differentiated electricity price and the differentiated log of the electricity price.

Table 3.1 displays relevant descriptive statistics of the electricity spot price, the logarithm of the price, the differentiated price and the differentiated log price. In addition, we have included several autocorrelation (ACF) tests, the Jarque-Bera normality test, Ljung-Box test for independence and the Dickey-Fuller/augmented Dickey-Fuller tests for stationarity. We observe a non-normal behaviour in all series, We note that stationarity of both El-spot price and lnEl-spot price is rejected, whereas stationarity of Δ El-spot price and Δ lnEl-spot price is confirmed. When testing for independence with the Ljung-box test, all the time-series where confirmed within a interval of 10%, but within a intervall of 5% ΔLnEl-Spot price NO5 was rejected, not independent. All the time series has confirmed normality regarding the Jarque-Bera test with a probability of 1%.

In the last two decades, three events have significantly affected the price of electricity in NO5, thereby creating outliers in the price distribution. In the beginning of August 2007 there was a 30% higher inflow rate than average, causing all the Nordic reservoirs to be filled. Consequently, the power production increased drastically and the spot price index of electricity registered all-time low prices (Johnsen, 2007). Again, in 2008, the power price in NO5 decreased by 41% from the first to the second quarter, while the other price zones registered increasing prices. This can also be explained by unusual high inflow rates in this period, and by reduced export capacity from NO5 (Johnsen, 2008). Lastly, in July 2012 the price registered a sharp one month decline of 28% compared to previous years. This was due to snow melting occurring later than usual, thereby causing high inflow rates during the summer months and an increase in power production (Pettersen, 2012). In

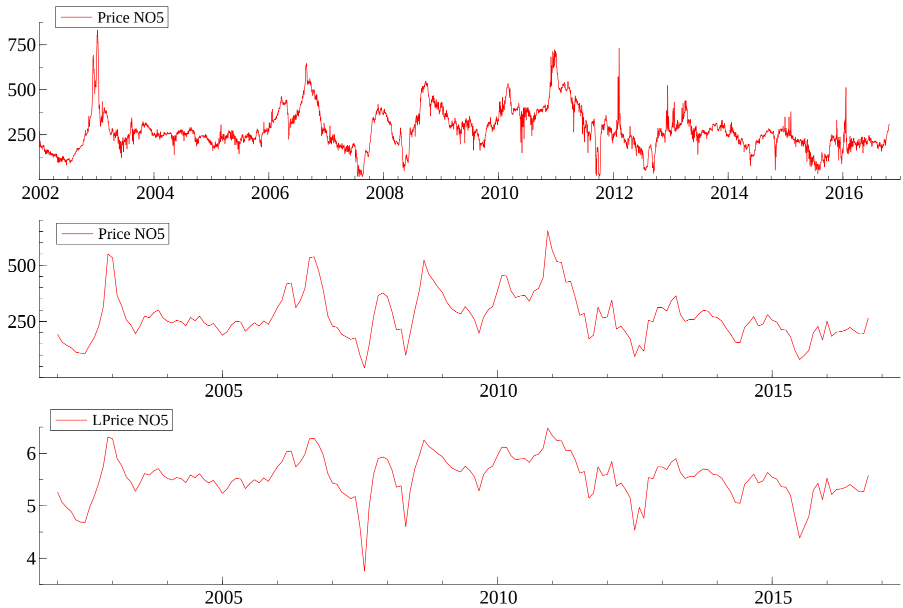

Figure 3.2: Data for the electricity price in NO5. a) Daily data, b) aggregated monthly data, and c) log of the aggregated monthly data.

Chapter 5 and Chapter 6, we will refer to these events as pulse effects.

Figure 3.3: Differentiated price and differentiated log price

Table 3.1: Descriptive statistics of the electricity spot price (monthly average)

	General information			**Descriptive statistics**					
	N.Obs	Min date	Max date	Mean	Std.dev	Min	Max	Skewness	Kurtosis
El-spot price, NO5	178	01.01.02	23.10.16	273.732	105.454	42.706	653.214	0.818	0.95
lnEl-spot price, NO5	178	01.01.02	23.10.16	5.535	0.409	3.754	6.482	-0.702	1.822
ΔEl-spot price, NO5	177	01.01.02	23.10.16	0.412	57.004	-167.264	233.951	0.661	2.272
ΔlnEl-spot price, NO5	177	01.01.02	23.10.16	0.002	0.248	-0.86	1.208	0.584	4.402

Tests

ACF(1)	ACF(2)	ACF(3)	ACF(4)	ACF(5)	ACF(6)	ACF(7)	ACF(8)	ACF(9)	ACF(10)
0.853	0.6545	0.4535	0.2682	0.1621	0.0963	0.0663	0.08	0.0804	0.0785
0.8162	0.6089	0.3996	0.2264	0.1156	0.0552	0.0372	0.0561	0.0858	0.0671
0.1713	0.0067	-0,0557	-0,2751	-0,1398	-0,1232	-0,1452	0.0501	0.0161	0.0394
0.0606	0.0029	-0,1012	-0,1752	-0,1394	-0,1161	-0,0937	-0,0240	0.1381	0.0271

ACF(11)	ACF(12)	ACF(24)	ACF(36)	ACF(48)	JB	Q(12)	DF	ADF
0.0684	0.0321	0.1835	0.0124	-0,0851	51.014	272.7141	-1.215256	-1.432498
0.042	0.0115	0.1161	0.0664	-0,0634	24.921	232.973	-0.200274	-0.161877
0.1246	-0,0945	0.1168	-0,0279	-0,0442	16.881	35.55005	-11.02186	-8.585854
0.0281	-0,0340	0.044	0.0388	-0,0352	24.704	20.10389	-12.32241	-8.94413

Critical values: JB:$\chi^2_{2,\alpha=10\%} > 4,61; \chi^2_{\alpha=5\%} > 5,99; \chi^2_{alpha=1\%} > 9,21$

Ljung Box: $\chi^2_{12,\alpha=10\%} > 18,6; \chi^2_{12,\alpha=5\%} > 21; \chi^2_{12,\alpha=1\%} > 26,2;$

DF,ADF:$\tau_{\alpha=10\%} < -1.62; \tau_{\alpha=5\%} < -1.95; \tau_{\alpha=1\%} < -2.58$

Chapter 4

Fundamental Price Drivers and Descriptive Statistics

4.1 Price Drivers

Norway has a cold climate compared to the average for International Energy Agency countries. The Scandinavian residential energy use is dominated by space heating due to high dwelling areas as well as the climate, the result is a higher energy consumption than most European countries (Unander et al., 2004).

The Nordic electricity prices have been historically low due to a large share of cost-effective hydropower and nuclear power. This has led to a development of an electricity-intensive industry structure as well as a large part of households use electricity for space heating. Therefore, the electricity consumption in Nordic region is relatively high compared with other European countries. The development of the GDP and average temperatures during the year, with lower electricity demand during the summer and increased consumption in the winter time (NordReg, 2014).

All of the explanatory variables included in the model are aggregated to monthly data, logarithmically transformed and shown in **Fig. 4.4**.

Figure 4.1 shows the electricity consumptions in Norway by sector. As seen by the graph, a large part of the overall electricity consumptions is from households and service industries. The single largest sector is energy-intensive industries.

Heating Degree Day

Although energy efficiency and improved heating systems have helped reduce the energy consumption in service sectors and households, energy use in these sectors remain highly

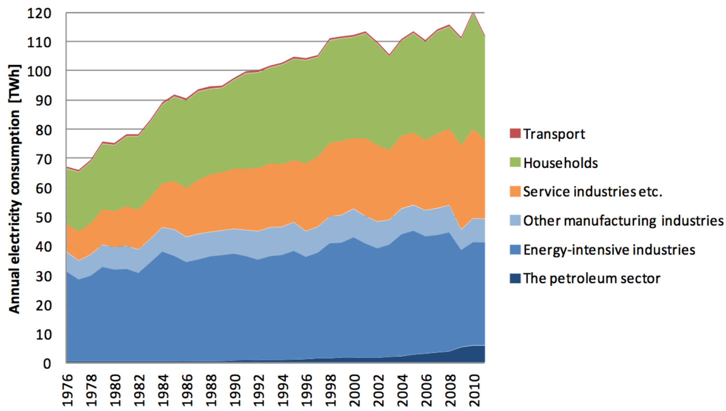

Figure 4.1: Electricity consumption in mainland Norway by sector

influenced by outdoors temperatures. Approximately 60% of a households energy consumption and 40% of energy consumption in commercial buildings is used for space heating (Bergersen et al., 2004).

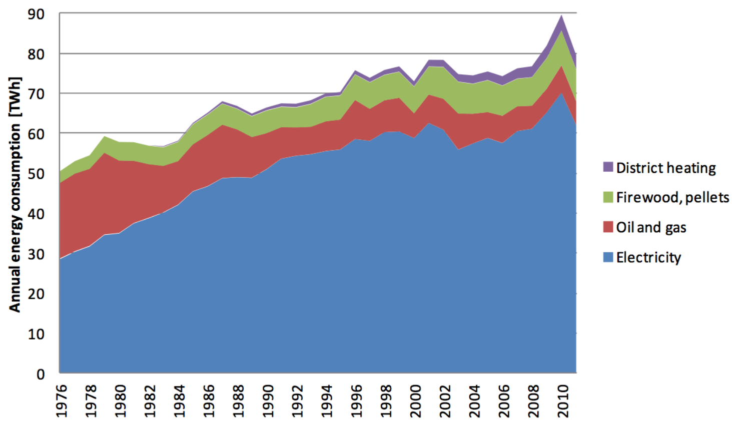

Figure 4.2: Energy consumption by energy product in households and service sectors

As shown in **Fig. 4.2**, a large part of households and service sectors use electricity for heating. The main cause for the large rise in energy consumption in 2010 and fall 2011 was outdoor temperature variations. In 2010 there were two very cold periods in the same year. The start of 2011 was also very cold, which explains the peak in energy consumption in the graph. (Bergersen et al., 2004).

The paper of Henley and Peirson has undertaken an empirical analysis of the relationship between temperature and electricity demand for space heating in the UK. The relationship is important, as it is the major determinant of the short-run demand for energy. They researched households who have electricity as their main source for space heating.

Henley and Peirson investigated the relationship between outside air temperature and the residential demand for space heating energy. They discovered empirically that there is a clear rise in demand for electric energy when temperatures fall. They also discovered the demand for energy when the outside temperature is 20°C and above is solely for non-heating purposes. **Fig. 4.3** is a graph,that explains the relationship between temperatures and demand for electricity. (Henley and Peirson, 1997)

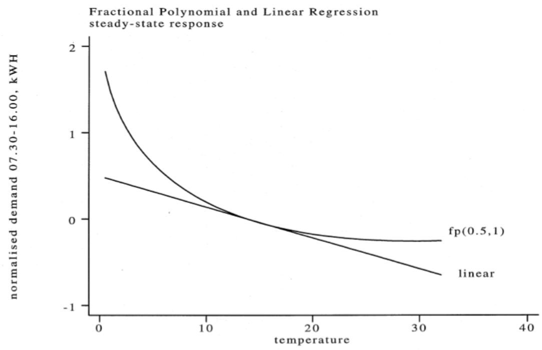

Figure 4.3: Demand as a function of temperature

Heating degree day is a measurement designed to measure the demand for energy needed to heat a building. The heating requirements for a given building at a specific location is considered to be directly proportional to the number HDD at that location. In this model we have used 17°C as a base temperature, also suggested by Seljon et al. (Seljom et al., 2011). We expect less correlation between energy consumption and temperature when temperatures exceed 17°C. We do not expect a large energy consumption difference between temperatures of e.g. 22°C and 26°C because neither of these temperatures require heating of homes. After this transformation, the temperatures were aggregated to a monthly level, added a constant of 1 000 and then log transformed. The reason for adding the constant is without it we would see large variations in the time series due to the numbers that approach zero. By adding this constant, the log of the temperature HDD is mellowed out.

$$HDD = \begin{cases} 17 - T, & T < 17 \\ 0, & T \geq 17 \end{cases} \tag{4.1}$$

The temperatures used in our model are monthly averaged temperatures measured from Bergen Airport, Flesland, from 1.1.2002 24.10.2016. Bergen is the largest city in price zone 5 and its temperatures will be the largest influence for the demand for electricity in price zone 5.

4.1.1 Power production in Nord Pool

Table. 4.1. below shows the average production split from 2002-2014. Norway has an average power generation of 52,9% of total power production in Nord Pool. To capture the supply side dynamics of the system price, we look towards what influences the marginal cost for these different productions. Nord Pool is a power exchange with a system price determined by demand and supply, we must include drivers for production and most important the ones influencing the total generation the most. Norwegian hydropower is the largest power generator in Nord Pool. The second largest producers are Hydropower from Sweden as well as Nuclear power production from Finland and Sweden (Spot, 2012).

Country Energy Source	Denmark	Finland	Norway	Sweden	Sum	Share of total generation
Hydropower	0.0	12.3	121.4	65.8	199.4	52.9
Nuclear power	0.0	22.3	0.0	58.0	80.3	21.3
Fossil fuels	21.8	24.2	4.8	5.4	56.1	14.9
Wind power	8.9	0.5	1.3	6.1	16.7	4.4
Other renewables	2.4	10.5	0.0	11.2	24.1	6.4
Non-identifiable	0.0	0.7	0.0	0.0	0.7	0.2
Total production	33.1	70.4	127.4	146.4	377.4	100.0

Table 4.1: Production Split 2004-2012

4.1.2 Reservoir level deviation from average

The water reservoir can be portrayed as the hydro power plants fuel reserve, but unlike fuel for the fossil power plants, water is free. Although water is free, a historical water scarcity caused by little precipitation will increase the value of the remaining water. Reservoir level is a fundamental driver for power prices in Norway, as stated by Kresen and Husby (Kaaresen and Husby, 2000), Haldrup and Nielsen (Haldrup and Nielsen, 2006) and Povh et al. (Povh et al., 2010).

We have included reservoir level deviation from historical average for our pricing zone, NO5, in our model. Our data is weekly from 2002-10.24.2016.

4.1.3 Production from fossil fuels

As seen in **Tab. 4.1**, Denmark generates 66% of its energy form fossile fuels. Combined in Nord Pool, energy from fossile fuels aggregate to 14,9% of the total power generation. The most common fossil fuels in Danish power stations are coal and natural gas. From 1990-2015 the coal share of the fuel consumption in Danish electricity consumption went from 92% to 41%. (DK, 2016)

Coal, Black liquor and Natural gas are the most used fossil fuels in Finland. This production constitutes 34% of their average power generation (Finland, 2016).

We have included the price of coal in our model as it is an important component in calculating the marginal cost of producing power from coal, as in return affects the amount of power produced by coal. Although their is a relationship between the price of Elspot and the price of coal, as stated by Chen and bunn, it is not a simple linear one (Chen and Bunn, 2010).

Specifically, we have used the price for Central Appalachian Coal (USA), traded at NYMEX. Finland imports most of its coal from Russia and USA. Dong Energy, the largest fossil power producer in Denmark, imports from a large range of different countries without revealing their composition, however they have revealed that they do not import coal from Columbia any more.

In our analysis we have also included the price for crude oil. Crude oil is not considered a direct substitute for electricity in heating of households in Norway, but is considered as an important explanatory variable for energy. Crude oil is also analysed as a driver for the supply side dynamics by Povh, Golob and Fleten in the thesis Modelling the structure of long-term electricity forward prices at Nord Pool (Povh et al., 2010). We have included the crude oil price traded at NYMEX in our model with daily data from 2002 - 10.24.2016.

We have also included the price for natural gas in our model as this is also a driver for the marginal cost of the supply side of power generation in Nord Pool. We have used the UK NBP (Natural Balancing Point) forward prices for 1 day traded at ICE, as the gas price. The UK NBP is considered as an indicator of Europes wholesale gas market. We have used daily data from 2002 - 10.24.2016

4.1.4 Nuclear Power Productions

Nuclear power plants use energy from a fission reaction to boil water and operates with the same principle as condensing power plants. In Finland and Sweden, nuclear power is used for base-load production. The operating rate for a nuclear plant is typically high, although price of fuel, uranium, is low compared to gas and coal. Since nuclear power stand for the second largest share of total power generation in Nord Pool, unexpected failures may have large price effects.

We have not included the fluctuation in prices for Uranium due to the low impact these prices have on production. We have not included any explanatory variables connected to nuclear power production in Finland or Sweden in our model at this point.

4.1.5 Production in Zone 5

Norwegian electricity production totaled 134 TWh in 2013, approximately 129 TWh of this was produced by hydropower, 1,9 TWh by windpower and 3,3 TWh by gas-fired power plants and other thermal plants. The average electricity production has been approximately 135 TWh per year for the last 15 years. Calculation from The Norwegian Water Resources and Energy Dicotorate shows that there is only currently remaining a hydropower potential of 33,8 TWh that can be developed (Aarrestad and Hatlen, 2015).

Hydropower production from power plants connected to regulation reservoirs is flexible and the potential energy is stored in the reservoirs established in lakes or by basins made by building a dam across a river. In periods of low consumption and high inflow, surplus water is collected in reservoirs. This water is then drawn from reservoirs to produce electricity. 96,2% of Norways power production was produced by hydropower in 2013 (NordReg, 2014).

The power production in Zone 5 has a yearly average of 29,3 TWh for 2013,2014 and 2015. The yearly average for the same years regarding consumption was 17,3 TWh, giving Zone 5 an average export of 12 TWh (Aarrestad and Hatlen, 2015).

Production from zone 5 is not included in our model due to the fact that the zones have changed in the time interval we are using data from. Zone 5 was established in 2010 and the production in the predeceasing zone for Bergen, zone 1, is much larger which would show as a large change in the power production in our estimations. These changes are due to change in zones and not changes in the market dynamics and that is why we have chosen to leave production out of the model at this point.

4.1.6 EU ETS Carbon Prices

The European Emissions Trading System (EU ETS) works on a principle called Cap and trade. A cap is the total amount of certain green house gases (GHG) that can be emitted by a installation covered by the system. To ensure that the total emissions fall, the cap is reduced over time. Within the cap, companies receive or buy emissions allowances which they can trade with other if they se fit. The limited number of allowances available ensures that they have a value. After each year, a company must surrender enough allowances to cover its emissions or else, higher fines are imposed.

A higher price on carbon allowances would raise the cost of electricity produced from fossil fuels relative to the cost of electricity from renewable sources, such as hydropower. We have not included the price of carbon allowances in our model as the carbon prices is very volatile and have had a massive jumps. In January 2009, the price crashed 47% in a short intervall of time. Grll and Kiesel (Grull and kiesel, 2012) concluded that market participants and regulators should not be surprised by several severe price drops and characterized the sensitivity of the Carbon Price as ”huge”.

4.2 Data Processing and Descriptive Statistics

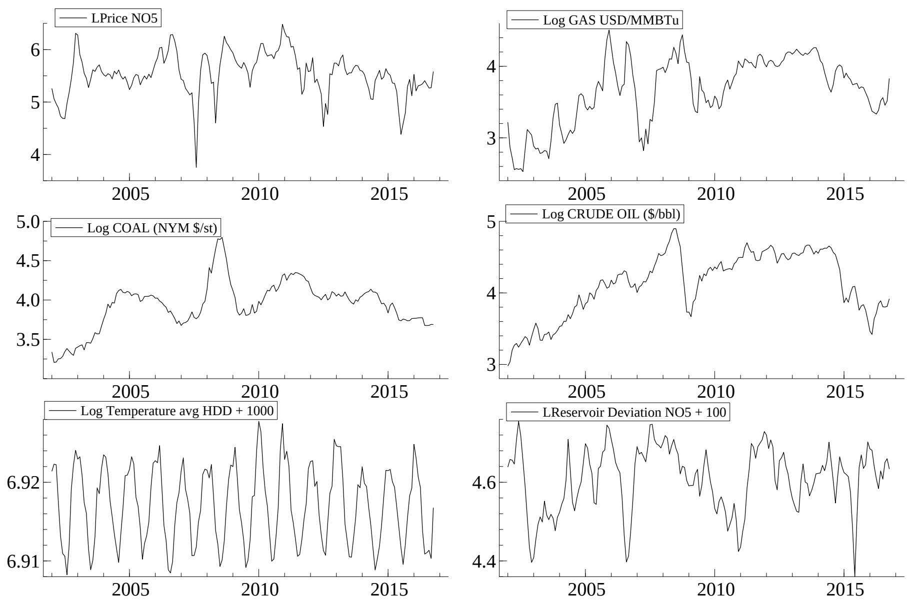

Figure 4.4: Explanatory variables: Gas, Coal, Crude Oil, Reservoir deviation NO5, Temperature

Chapter 5

Methodology

In order to both capture the endogenous factors of the electricity spot price, as well as identify the impact of exogenous fundamental drivers on the formation of the spot price, the electricity spot price is decomposed into unobserved stochastic processes. *Structural time series models* (STMs) allow for this decomposition in their formulation. In the literature, these models are also referred to as unobservable component models (UCMs). Basic STMs are formulated directly in terms of level, slope and seasonals, which have a natural interpretation and highlight the salient features of the electricity spot price (Proietti, 2002). To capture specific dynamic variations in the spot price, the basic models can easily be extended to include fundamental price drivers, referred to as explanatory variables, and outlier effects. STMs allow for time-varying sensitivities to such explanatory variables, thereby providing a sensible way of weighting the exogenous effects from different price drivers.

Structural time series models have a natural linear Gaussian state space representation, and are thus based on three assumptions concerning the residuals of the models. In decreasing order of importance, the residuals should satisfy independence, homoscedasticity and normality. The following sub-section (4.1) brings the reader through a presentation of selected structural time series models, in accordance with Commandeur and Koopman (2007). In these models, each unobserved component is formulated as a stochastically evolving process over time. A statistical treatment of the models is considered in sub-section (4.2). Please refer to Appendix A for a systematic application of the linear Gaussian state space approach to structural time series models and Appendix B for further elaboration on model assumptions.

5.1 Specification of Structural Time Series Models

5.1.1 The Local Level Model

The local level model (LLM) allows the level component to vary in time. The statistical specification of a local level model consists of a random walk component, to capture the underlying level, plus a random, white noise, disturbance term,

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, \qquad \epsilon_t \sim NID(0, \sigma_\epsilon^2), \qquad t = 1, ..., T \\ \mu_{t+1} &= \mu_t + \eta_t, \qquad \eta_t \sim NID(o, \sigma_\nu^2) \end{aligned} \tag{5.1}$$

where μ_t is the underlying level, ϵ_t is the observation disturbance and η_t is the white noise disturbance driving the level. Both disturbances are normally and independently distributed, with zero mean and variances σ_ϵ^2 and σ_η^2, respectively. In the literature on state space models, the noise, ϵ_t , is also referred to as the irregular component. In case $\sigma_\eta^2 = 0$, the model is simply a constant level model:$y_t = \mu_1 + \epsilon_t$, whereas when $\sigma_\epsilon^2 = 0$, the model reduces to a pure random walk and the trend coincides with the observations.

5.1.2 The Local Linear Trend Model

The local linear trend model generalizes the local level model by introducing a stochastic slope, ν_t, which itself follows a random walk. Thus,

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, \qquad \epsilon_t \sim NID(0, \sigma_\epsilon^2) \\ \mu_{t+1} &= \mu_t + \nu_t + \eta_t, \qquad \eta_t \sim NID(0, \sigma_t^2) \\ \eta_{t+1} &= \eta_t + \zeta_t, \qquad \zeta_t \sim NID(0, \sigma_\zeta^2) \end{aligned} \tag{5.2}$$

where the irregular, level and slope disturbances, ϵ_t, η_t and ζ_t, respectively, are mutually independent. This model has the same measurement equation as the local level model, but with a time-varying slope in the dynamics for μ_t. In case $\sigma_\zeta^2 = 0$, the trend reduces to a random walk with constant drift:$\mu_{t+1} = \mu_t + \nu_t + \eta_t$, whereas for $\sigma_\eta^2 = 0$, the trend is an integrated random walk: $\Delta^2 \mu_{t+1} = \zeta_{t-1}$. The latter is referred to as *smoothness prior* specification, as the resulting trend varies smoothly over time. Ultimately, when $\sigma_\zeta^2 = 0$ and $\sigma_\eta^2 = 0$, we obtain a deterministic linear trend model: $\mu_{t+1} = \mu_1 + v_1 t$.

5.1.3 Trigonometric Seasonal Models

Seasonality is considered one of the most important features of electricity prices, and refers to periodic fluctuations that occur regularly based on a particular season. According to Koopman et al. (2007), seasonality in electricity spot prices occur due to the strong dependence of electricity demand on weather conditions as well as social and economic activities, leading to different holiday and seasonal effects. Moreover, they argue that one should always be on the alert for such recurring patterns whenever a time series consists of hourly, daily, monthly or quarterly observations. To account for seasonal variation in a time series, the component γ_t may be added to the model. More specifically, γ_t represents the seasonal effect at time t that is associated with season $s = s(t)$ for $s = 1, , S$, where S is the seasonal length. Thus, $S = 12$ for monthly data, $S = 4$ for quarterly data and $S = 7$ when modelling the weekly pattern for daily data. The time-varying seasonal can be established in different ways.

A seasonal pattern may be captured by a set of trigonometric terms at the seasonal frequencies, $\lambda_j = \frac{2\pi j}{s}$ for $j = 1, ..., [\frac{s}{2}]$.

$$\gamma_t = \sum_{j=1}^{[\frac{s}{2}]} \Big(\tilde{\gamma}_j \cos \lambda_j t + \tilde{\gamma}_j^* \sin \lambda_j t \Big), \qquad t = t, ..., T \tag{5.3}$$

where $\tilde{\gamma}_j$ and $\tilde{\gamma}_j^*$ are given constants. For a time-varying seasonal this can be made seasonal by replacing $\tilde{\gamma}_j$ and $\tilde{\gamma}_j^{**}$ by the random walks

$$\begin{aligned} \tilde{\gamma}_{t+1} &= \tilde{\gamma}_{j,t} + \omega_{j,t}, \qquad \tilde{\omega}_{j,t} \sim NID(o, \sigma_\omega^2), \qquad t = 1, ..., T \\ \tilde{\gamma}_{t+1}^* &= \tilde{\gamma}_{j,t}^* + \omega_{j,t}^*, \qquad \tilde{\omega}_{j,t} \sim NID(o, \sigma_\omega^2), \qquad t = 1, ..., T \end{aligned} \tag{5.4}$$

Each of the two seasonal models can be combined with either of the trend models to give a structural times series model. For example, adding a time-varying seasonal to the local linear trend model of monthly data, $S = 12$, and $j = 1, .., 6$, gives

$$\begin{aligned} y_t &= \mu_t + \gamma_t + \epsilon_t, \qquad \epsilon_t \sim NID(0, \sigma_\epsilon^2) \\ \mu_{t+1} &= \mu_t + \nu_t + \eta_t, \qquad \eta_t \sim NID(0, \sigma_\eta^2) \\ \nu_{t+1} &= \nu_t + \zeta_t, \qquad \zeta_t \sim NID(0, \sigma_\zeta^2) \\ \tilde{\gamma}_{1,t+1} &= \tilde{\gamma}_{1,t} + \tilde{\omega}_{1,t}, \qquad \tilde{\omega}_{1,t} \sim NID(0,_\omega^2) \\ \tilde{\gamma}_{1,t+1}^* &= \tilde{\gamma}_{1,t}^* + \tilde{\omega}_{1,t}^*, \qquad \tilde{\omega}_{1,t}^* \sim NID(0,_\omega^2) \\ \tilde{\gamma}_{2,t+1} &= \tilde{\gamma}_{2,t} + \tilde{\omega}_{2,t}, \qquad \tilde{\omega}_{2,t} \sim NID(0,_\omega^2) \\ \tilde{\gamma}_{2,t+1}^* &= \tilde{\gamma}_{2,t}^* + \tilde{\omega}_{2,t}^*, \qquad \tilde{\omega}_{2,t}^* \sim NID(0,_\omega^2) \\ \tilde{\gamma}_{3,t+1} &= \tilde{\gamma}_{3,t} + \tilde{\omega}_{3,t}, \qquad \tilde{\omega}_{3,t} \sim NID(0,_\omega^2) \\ \tilde{\gamma}_{3,t+1}^* &= \tilde{\gamma}_{3,t}^* + \tilde{\omega}_{3,t}^*, \qquad \tilde{\omega}_{3,t}^* \sim NID(0,_\omega^2) \\ \tilde{\gamma}_{4,t+1} &= \tilde{\gamma}_{4,t} + \tilde{\omega}_{4,t}, \qquad \tilde{\omega}_{4,t} \sim NID(0,_\omega^2) \\ \tilde{\gamma}_{4,t+1}^* &= \tilde{\gamma}_{4,t}^* + \tilde{\omega}_{4,t}^*, \qquad \tilde{\omega}_{4,t}^* \sim NID(0,_\omega^2) \\ \tilde{\gamma}_{5,t+1} &= \tilde{\gamma}_{5,t} + \tilde{\omega}_{5,t}, \qquad \tilde{\omega}_{5,t} \sim NID(0,_\omega^2) \\ \tilde{\gamma}_{5,t+1}^* &= \tilde{\gamma}_{5,t}^* + \tilde{\omega}_{5,t}^*, \qquad \tilde{\omega}_{5,t}^* \sim NID(0,_\omega^2) \\ \tilde{\gamma}_{6,t+1} &= \tilde{\gamma}_{6,t} + \tilde{\omega}_{6,t}, \qquad \tilde{\omega}_{6,t} \sim NID(0,_\omega^2) \\ \tilde{\gamma}_{6,t+1}^* &= \tilde{\gamma}_{6,t}^* + \tilde{\omega}_{6,t}^*, \qquad \tilde{\omega}_{6,t}^* \sim NID(0,_\omega^2) \end{aligned} \tag{5.5}$$

where ν_t is the stochastic slop of the trend μ_t, and γ_t is the seasonal effect component

$$\begin{aligned} \gamma_t = \tilde{\gamma}_1 \cos\left(\frac{2\pi \cdot 1}{12} t\right) + \tilde{\gamma}_1^* \sin\left(\frac{2\pi \cdot 1}{12} t\right) + \tilde{\gamma}_2 \cos\left(\frac{2\pi \cdot 1}{12} t\right) + \tilde{\gamma}_2^* \sin\left(\frac{2\pi \cdot 1}{12} t\right) \\ + \tilde{\gamma}_3 \cos\left(\frac{2\pi \cdot 1}{12} t\right) + \tilde{\gamma}_3^* \sin\left(\frac{2\pi \cdot 1}{12} t\right) + \tilde{\gamma}_4 \cos\left(\frac{2\pi \cdot 1}{12} t\right) + \tilde{\gamma}_4^* \sin\left(\frac{2\pi \cdot 1}{12} t\right) \\ + \tilde{\gamma}_5 \cos\left(\frac{2\pi \cdot 1}{12} t\right) + \tilde{\gamma}_5^* \sin\left(\frac{2\pi \cdot 1}{12} t\right) + \tilde{\gamma}_6 \cos\left(\frac{2\pi \cdot 1}{12} t\right) + \tilde{\gamma}_6^* \sin\left(\frac{2\pi \cdot 1}{12} t\right) \end{aligned} \tag{5.6}$$

5.1.4 The Local Level Model with Explanatory Variables

To investigate the exogenous effects of fundamental price drivers on the development of the electricity spot price, explanatory variables can be brought into the model. In the literature, explanatory variables are also referred to as regression variables. If regression variables are added to the local level model, the measurement equation becomes

$$y_t = \mu_t + \sum_{j=1}^{s-1} \beta_{j,t} x_{j,t} + \epsilon_t, \qquad \epsilon_t \sim NID(0, \sigma_\epsilon^2) \qquad t = 1, ..., T \tag{5.7}$$

where $x_{j,t}$ is a continuous predictor variable and $\beta_{j,t}$ is an unknown regression weight, for $j = 1, ..., k$. Since all components are allowed to change over time, the regression weights are also allowed to change over time. In case of one predictor variable with regression weight $\beta_{1,t}$ the local level model takes the form

$$\begin{aligned} y_t &= \mu_t + \beta_{1,t} x_{1,t} + \epsilon_t, \qquad && \epsilon_t \sim NID(0, \sigma_\epsilon^2) \qquad t = 1, ..., T \\ \mu_t &= \mu_{t-1} + \eta_t, && \eta_t \sim NID(0, \sigma_\eta^2) \\ & \beta_{t+1} + \tau_t, && \tau_t \sim NID(0, \sigma_\tau^2) \end{aligned} \tag{5.8}$$

The modelling of k explanatory variables requires k state equations, one for each explanatory variable.

5.1.5 The Local Level Model with Intervention Variables

To capture and assess the exogenous effect of outliers, also called temporary pulse effects, in the spot price distribution, intervention variables can be added to the models. Suppose we wish to measure the change in level due to an intervention at time t. An intervention variable i_t can be added to the local level model, as follows:

$$\begin{aligned} y_t &= \mu_t + \lambda i_t + \epsilon_t, \qquad && \epsilon_t \sim NID(0, \sigma_\epsilon^2), \qquad t = 1, ..., T \\ \mu_{t+1} &= \mu_t + \eta_t, && \eta_t \sim NID(0, \sigma_\eta^2) \end{aligned} \tag{5.9}$$

where λ measures the change in the level of the series at a known time t due to an intervention at time t. In case of level shift, the intervention variable is defined by

$$i_t = \begin{cases} 1, & t < \tau \\ 0, & t \geq \tau \end{cases} \tag{5.10}$$

whereas in case of a pulse, the intervention variable is defined by

$$i_t = \begin{cases} 1, & t < \tau, \quad t < \tau \\ 0, & t = \tau \end{cases} \tag{5.11}$$

Since y_t is in logarithms, while the intervention variable i_t is not, the value of regression weight λ can not be interpreted as an elasticity. The percent change due to an intervention equals

$$100\left(\frac{e^{\tilde{y},post} - e^{\tilde{y},pre}}{e^{\tilde{y},pre}}\right)$$

5.2 Statistical Treatment of Structural Time Series Models

The linear Gaussian state space representation of STMs, where the state vector is a vector of unobserved components, implies that the statistical treatment of STMs can be based on the Kalman filter and its related methods (Harvey, 2006). We will leave the detailed explanation of these matters to (Wei, 2006). In short, the Kalman filter is a recursive algorithm which estimates the state vector while the mean of the squared error is minimized (Welch and Bishop, 2006). The algorithm allows for previous estimates to be updated each time a new observation is brought in, and is thus able to capture the time-varying property of endognous and exogenous factors.
STAMP

The econometric software OxMetrics 7 and the module STAMP 8 was used in the implementation of the models in this paper. STAMP is an abbreviation for Structural Time Series Analyses, Models and Predictors, and runs within the interface of OxMetrics. In order to fit unobserved component time series models, the software uses the Kalman filter in an interactive menu-driven way (Koopman et. Al. 2007). The graphics included and the diagnostic outputs are produced by OxMetrics and Stamp.

Chapter 6

Application of Structural Time Series Models to the Electricity Spot Price in NO5

In this chapter, we apply the structural time series models introduced in Chapter 5 to the Nord Pool electricity spot price in NO5. We allow for both endogenous factors and the weighting of exogenous factors to vary with time, with the goal of capturing the complex price dynamics of our price distribution. The endogenous and exogenous components are combined in several ways, in terms of stochastic features and deterministic features, thereby creating a wide range of possible models. The process of selecting optimal regression variables and fitting a regression model, is based on assessment of the three model assumptions concerning the residuals: independence, normality, and homoscedasticity, together with a comparison of model performances in terms of the AIC evaluation criteria. Some economists have raised concerns regarding how selection rules and data mining may affect the resulting models. However, Hoover and Perez (Hoover and Perez, 2002) argue that intelligent mining of data can uncover complex relationships between economic variables.

6.1 Basic Structural Time Series Models with Intervention Variables

6.1.1 Intervention Variables

In order to capture the effects of the three extreme events mentioned in Chapter 3, we start the modelling process by including intervention variables. Simulation of intervention effects in STAMP detects three outliers in the price distribution, at the exact same time as the extreme events, thereby supporting earlier discussion. thus, these intervention variables are included in every model we present. The intervention coefficients are presented in *Table 6.1*,

are included in the rest of the modelling process.

	Coefficient
Outlier 2007(8)	-1.03415
Outlier 2008(5)	-0.73816
Outlier 2012(7)	-0.52598

Table 6.1: Detected outliers in the spot price distribution of NO5

We see that the Elspot prices in NO5 in August 2007, May 2008 and July 2012 were detected as outliers.

6.1.2 Local level model

The local level model with a stochastic level provides the simplest structural time series model in our analysis:

$$Y = Level + Irregular + Interventions$$

The resulting model and its components is showed in *Figure. 6.1*.

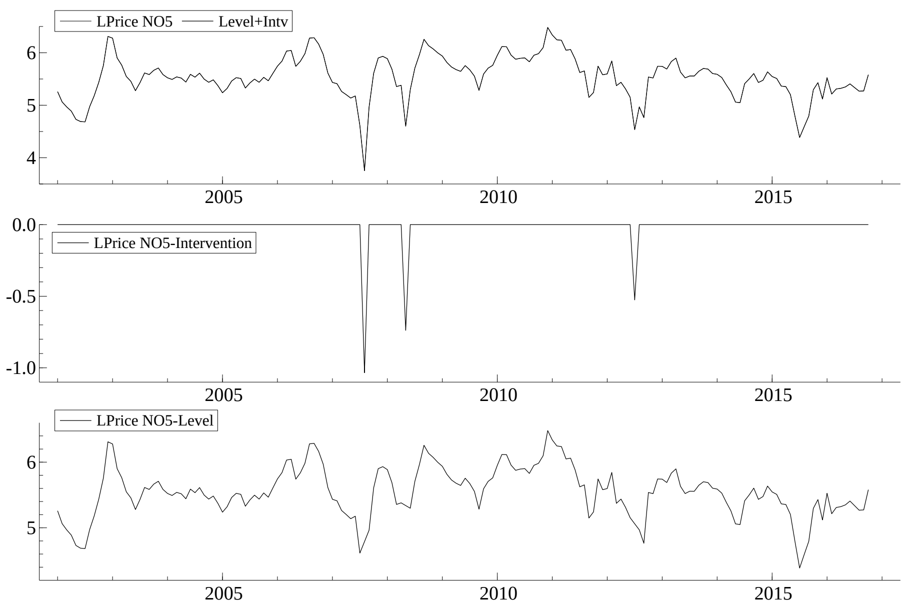

Figure 6.1: Local level model

The upper graph in *Figure 6.1* shows the logarithm of the price, together with a model with stochastic level and interventions. The middle graph shows the interventions with the three outliers. We subtract the interventions from LPrice NO5 such that the resulting level component corresponds to the one showed in the lower graph.

	Value	Prob.
Level	5.58087	[0.00000]

Table 6.2: State vector analysis at 2016(10)

	Statistic	Value	Critical value	Assumption satisfied
Independence	Q(24)	46.071	35.172	-
	r(1)	0.15631	± 0.1499063378	-
	r(24)	0.08384	± 0.1499063378	+
Homoscedasticity	H(58)	1.7382	1.68	-
Normality	N	14.532	5.99	-

Table 6.3: Residuals diagnostics test for local level model

The diagnostics test for the assumption of independence, homoscedasticity and normality of the residuals of the analysis are presented in *Table 6.3* and the value for level is presented in *Table 6.2*. The value of autocorrelation at lag 1 far exceed the confidence interval of a 95% confidence limit for the time series. However, for lag 24, the value of autocorrelation did not exceed the 95%confidence limit. The high amount of dependency between the residuals is confirmed by the value of the Q-test. This value exceeds the $\chi^2_{(23;0.05)} = 35.172$, evaluated as a whole the first 24 autocorrelations deviate from zero. The Q-statistic is a general omnibus test that is used to check whether the combined first, in this case 24, autocorrelations deviate from zero. With both the autocorrelation at lag 1, r(1) and the general Q-test indicating significant serial correlation in the residuals, the null hypothesis of H_0 = Data are random, is failed. Failing this statistic shows that the model is not independent.

The H-statistic in *Table 6.3* test whether the variances of two consecutive and equal parts of the residuals are equal to one another. The result of the test show that the first 58 elements of the residuals is unequal to the variance of the last 58 elements of the residuals. This is because H(58)=1.7382 is larger than the critical value of $F_{(58,58,0.0025)} = 1.68$. This means that the assumption of homosedasticity of the residuals is also not satisfied in this analysis.

The N-statistic in *Table 6.3* tests whether the skewness and kurtosis of the distribution of the residuals comply is Gaussian or normally distributed. The null hypothesis of normally distributed residuals is rejected.

The AIC for this test equals -3.1775, with smaller values shows of AIC showing a better fit for the model.

6.1.3 Local Linear Trend Model

The local linear trend model generalize the local level model by introducing a stochastic slope, which itself follows a random walk. The model is:

$$Y = Level + Slope + Irregular + Interventions$$

The slope determines the angle of the trend line with the x-axis and is able to change over time. Often in literature on time series, the slope is referred to as the drift. Trend consists of both the slope and the level, and both are able to vary over time.

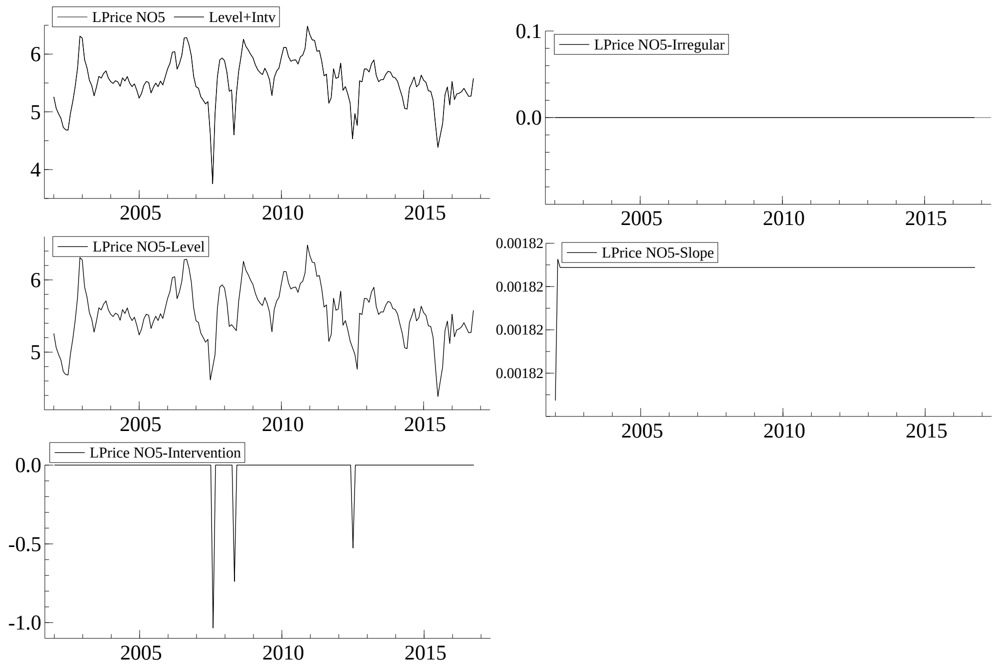

Figure 6.2: Local linear trend model

The LPrice NO5-Slope graph in *Figure 6.2* shows that there is very little variance in the slope component. In fact, both the slope components and the irregular components variances show values of zero, so the slope component could just as well be set to fixed without altering the model. At the last state, their coefficients are shown if *Table 6.4*.

	Value	Prob.
Level	5.58087	[0.00000]
Slope	0.00182	[0.90459]

Table 6.4: State vector analysis at 2016(10)

	Statistic	Value	Critical value	Assumption satisfied
Independence	Q(24)	43.626	33.924	-
	r(1)	0.16602	±0.1499063378	-
	r(24)	0.082197	±0.1499063378	+
Homoscedasticity	H(57)	1.7307	1.69	-
Normality	N	14.055	5.99	-

Table 6.5: Residuals diagnostics test for local linear trend model

The residuals diagnostics test for the local linear model is presented in *Table 6.5*. As in our previous model, the autocorrelation for lag 1, r(1), does not satisfy and for lag 24, r(24), it does satisfy the 95% confidence interval.

The overall Q-test for the first 24 autocorrelations confirms that the assumption of Independence is still not satisfied, as Q(24) is larger than the critical value, $\chi^2_{(22;0.05)} = 33.924$.

As in our previous model, both the assumption of homoscedasticity and normality is clearly violated.

The AIC for the local linear trend model is -3.1661 which is a little larger than the AIC for our previous model, making this model a slightly worse fit for the LPrice NO5. This suggests that inclusion of a stochastic slope has not helped the analysis.

6.1.4 Local Linear Trend Mode with Seasonal

In this part of the model we add a seasonal component, checking if it has a recurring pattern. To improve the model further a seasonal component is added to the local linear model. Both the level and the seasonal is able to vary over time.

$$Y = Level + slope + Seasonal + Irregular + Interventions$$

Figure 6.3 shows all of the different components in the local model with seasonal. As shown by the LPrice NO5-seasonal graph in the same figure, the variance of the seasonal disturbances is small. This indicates that the seasonal pattern does not change over the years.

	Value	Prob.
Level	5.51553	[0.00000]
Slope	0.00241	[0.86735]
Seasonal chi2 test	29.70526	[0.00176]

Table 6.6: State vector analysis for the linear trend model with seasonal at period 2016(10)

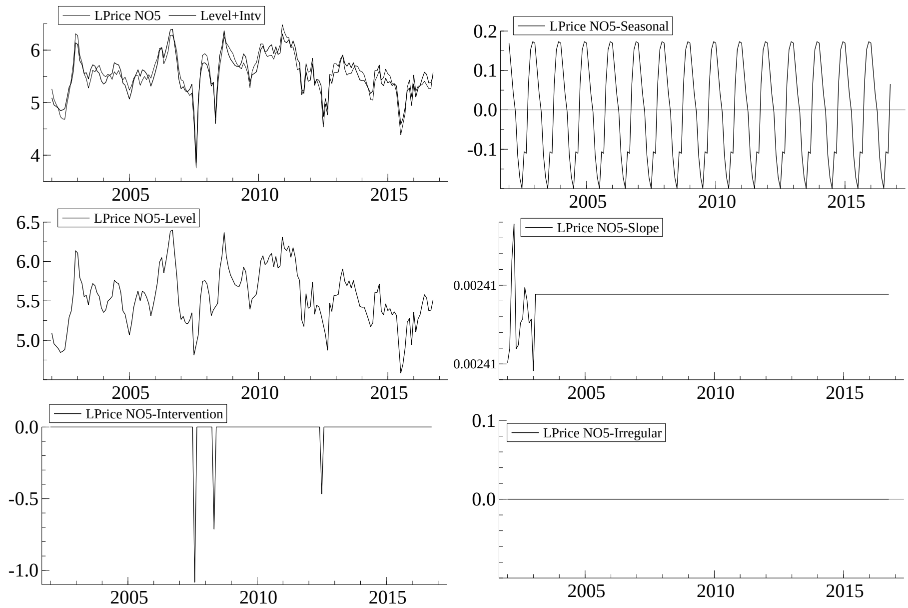

Figure 6.3: Local linear trend model with seasonal

	Statistic	Value	Critical value	Assumption satisfied
Independence	Q(24)	38.07	32.671	-
	r(1)	0.08100	0.1499063378	+
	r(24)	0.05885	0.1499063378	+
Homoscedasticity	H(54)	1.1459	1.71	+
Normality	N	7.5107	5.99	-

Table 6.7: Residuals diagnostics test for linear trend model with seasonal

The autocorrelation for lag 1 and 24, respectively r(1) and r(24), is within the 95% confidence interval. In the previous models, the autocorrelation for lag 1 was found to be unacceptably large.

The residuals of this model satisfies the required criteria for homoscedasticity. The previous two models did not satisfy this criteria. This is the first model with variances of the residuals of the third part of the series equal to the variance of the residuals corresponding to the last third part of the series.

This model has the lowest value for Normality yet, but it is still not below the critical value and the null hypothesis of normality has to be rejected.

The value of AIC is -3.2084 for this model, which is lower than for the previous models, making this the best fit for our time series.

6.2 Models with Explanatory Variables

There are many combinations of state space models we can apply to the time series, especially when including explanatory variables. In order to narrow down the possibilities, we will select one combination of the components level, trend and seasonal by setting these to either fixed or stochastic. We have compared the AIC values of the different possible combinations and the lowest AIC value found was -3.2084. The results for all of the combinations are shown in **Fig. 6.8**. We will continue with a stochastic level, fixed trend and fixed seasonal. This is also the configuration that is recommended by STAMP after running the model with every component set to stochastic.

Level	Trend	Seasonal	AIC
S	S	S	-3.2084
S	S	F	-3.2084
S	F	S	-3.2084
S	F	F	-3.2084
F	S	S	-2.8936
F	S	F	-2.8936
F	F	S	-1.9137
F	F	F	-1.9137

Table 6.8: Combinations of level, trend and seasonal with AIC

The three models described in this section are models A, B and C. Model A has a stochastic level and stochastic explanatory variables, model B has a fixed level and stochastic explanatory variables, and model C has a fixed level and fixed explanatory variables.

6.2.1 Model A

Model A contains a stochastic level and stochastic explanatory variables, while the slope and seasonal components are treated as fixed. The variances of the disturbance components shows that there was little to no variance in most of the components, indicating that they could be treated as fixed. Figure 6.4 shows the components of model A. The only component that shows a slight variation is crude oil.

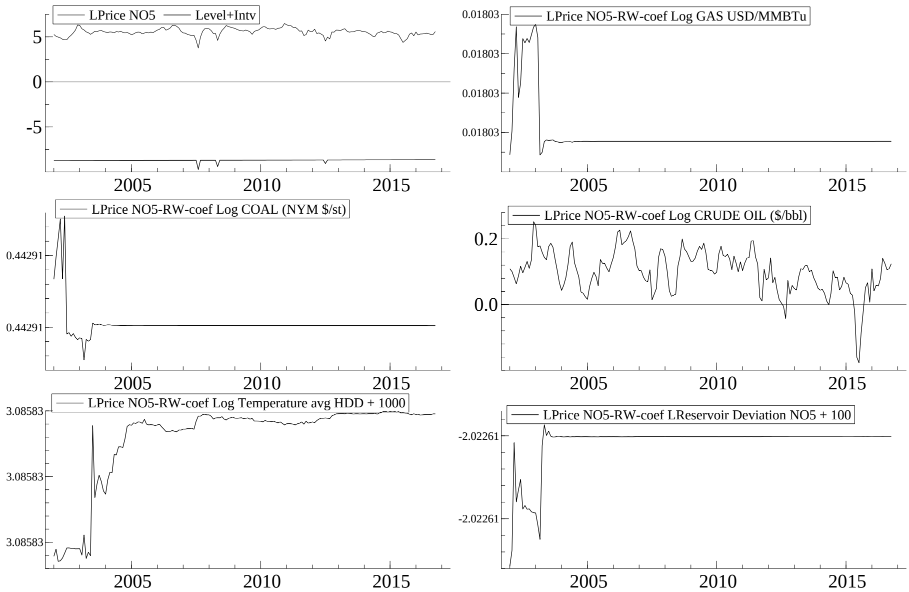

Figure 6.4: Model A - Component graphics

	Statistic	Value	Critical value	Assumption satisfied
Independence	Q(24)	43,701	32,671	-
	r(1)	0,058249	± 0.1499063378	+
	r(24)	-0,056627	± 0.1499063378	+
Homoscedasticity	H(52)	1,374	1,73	+
Normality	N	14,571	5.99	-

Table 6.9: Residual diagnostics - Model A

As seen in **Tab. 6.9**, the residuals in model A are not independently or normally distributed. Although the two values of autocorrelation at lags 1 and 24 are within the critical values, the ACF graph in **Fig. 6.5** shows that several lags are outside the critical value of ± 0.1499063378.

Model A has a AIC of -6.3085, making it the model with the best fit.

MANGLER: Her skal det st om vektingen til forklaringsvariablene og trekkes paralleller til markedet.

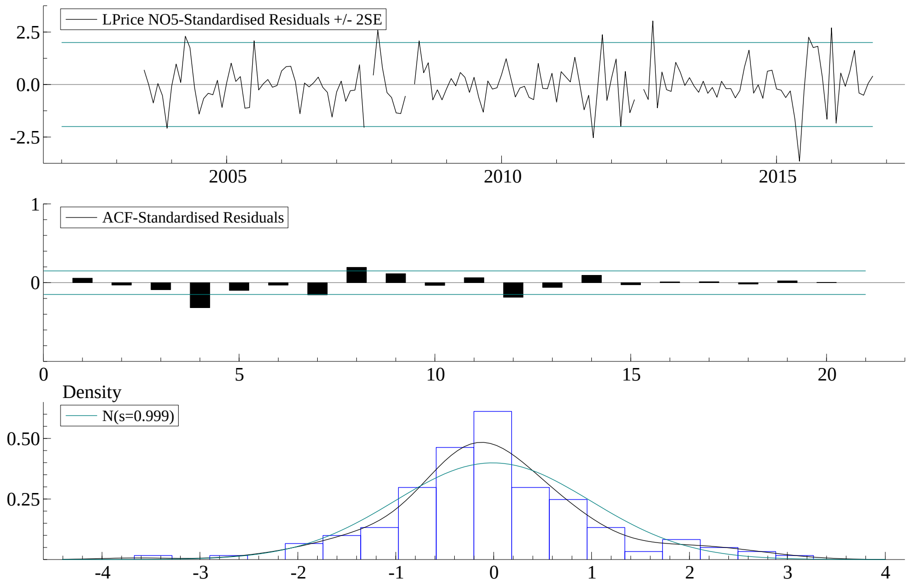

Figure 6.5: Model A - Residuals

6.2.2 Model B

Considering the lack of variance in the level component of model A, model B was made with a fixed level, trend and seasonal, together with stochastic explanatory variables.

From the y-axis on each of the time-varying coefficients in this model, we can see that the gas price, coal price, temperature heating degree days and reservoir deviation are all found to have little to no variance. Therefore, we will set these as fixed in the next model. The oil price, as in model A, seems to have certain time-varying specifications. The oil price coefficient shows no clear trend or seasonal effects, however, and it is, therefore, difficult to draw conclusions as to why it is time-varying and how it affects the electricity price. What we can clearly see is that the oil price usually will have a positive effect on the electricity price. That is, when the oil price goes up, the electricity price will also rise. One interesting observation from the plot of the oil price coefficient is that it appears, for some unknown reason, to have had a negative impact on the oil price in the middle of 2015. In other words, according to our model, if oil prices rose in the middle of 2015, electricity prices would fall, and vice versa. The coefficient both starts and stops at around 0.06, and we cannot conclude that the oil price has become neither more nor less important today than in 2002.

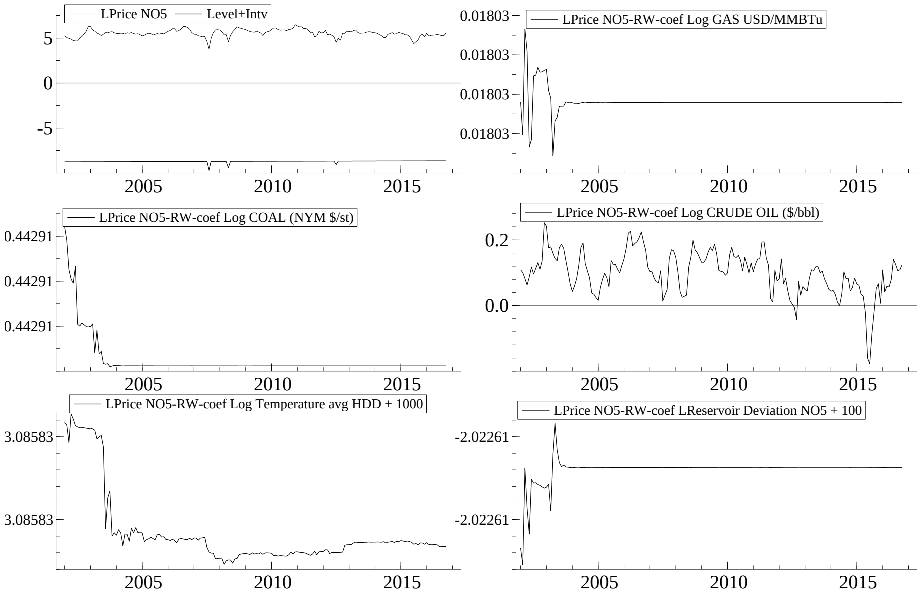

Figure 6.6: Model B - Component graphics

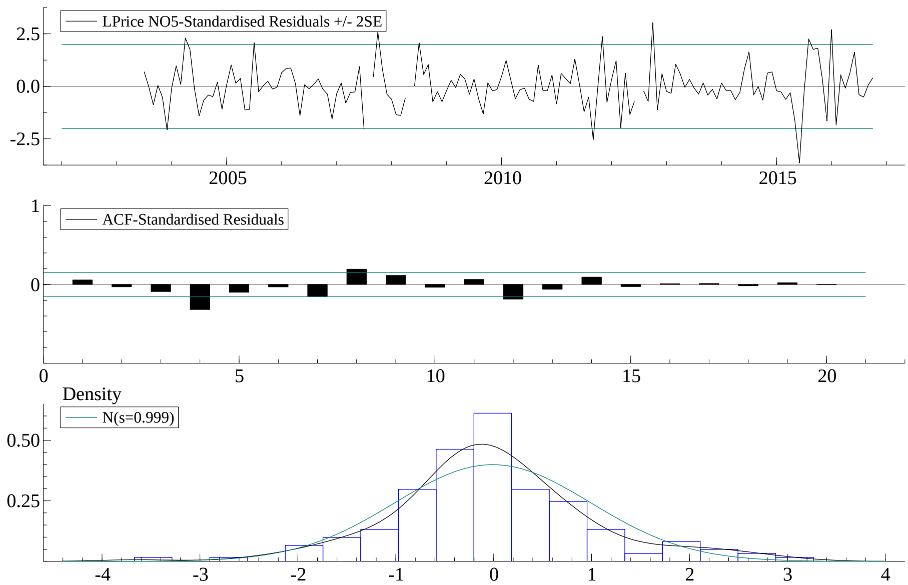

Figure 6.7: Model B - Residuals

Table 6.11: Coefficients

Residuals diagnostic tests	Coefficient	RMSE	t-value	Prob
Log GAS USD/MMBTu	0.01803	0.09875	0.18259	[0.85535]
Log COAL (NYM $/st)	0.44291	0.21832	2.02874	[0.04417]
LReservoir Deviation NO5 + 100	-2.02261	0.29681	-6.81443	[0.00000]
Log Temperature avg HDD + 1000	3.08544	7.71232	0.40007	[0.68965]

	Statistic	Value	Critical value	Assumption satisfied
Independence	Q(24)	43,701	32,671	-
	r(1)	0,058249	±0.1499063378	+
	r(24)	-0,056627	±0.1499063378	+
Homoscedasticity	H(52)	1,374	1,73	+
Normality	N	14,571	5.99	-

Table 6.10: Residual diagnostics - Model B

6.2.3 Model C

As mentioned, all of the coefficients that were allowed to vary with time, but were found to have no variance, are set as fixed in this model. This leaves only the coefficient of crude oil as time-varying.

The fixed coefficients of the explanatory variables show logical relationships with the electricity price and meet many of our assumptions about the electricity market. That is, we would expect increased electricity prices as a result of increased gas prices, which is confirmed. The same is true for the coal price and for the time-varying crude oil price (except for the previously mentioned short period in 2015, where the opposite was true). Another observation we would expect is that the reservoir deviation negatively impacts the oil price because of the way we have calculated it, which is the present value minus historical average value. In other words, a positive reservoir deviation value here would imply that there is more water in the reservoirs than usual. Therefore, when the reservoir deviation value is positive, there is more water in the reservoirs than usual, and the electricity price decreases (as shown by its negative coefficient). The opposite is also true; water scarcity is shown to increase the electricity price in zone NO5.

AIC: -6.3085

Table 6.12: My caption

Residuals diagnostic tests				
	Statistic	Value	Critical value	Assumption satisfied
Independence	Q(24)	38,813	32,671	-
	r(1)	0,063929	0,149906338	+
	r(24)	0,049348	0,149906338	+
Homoscedasticity	H(52)	1,3105	1,73	+
Normality	N	17,416	5,99	-

modell C - graphics/components.pdf

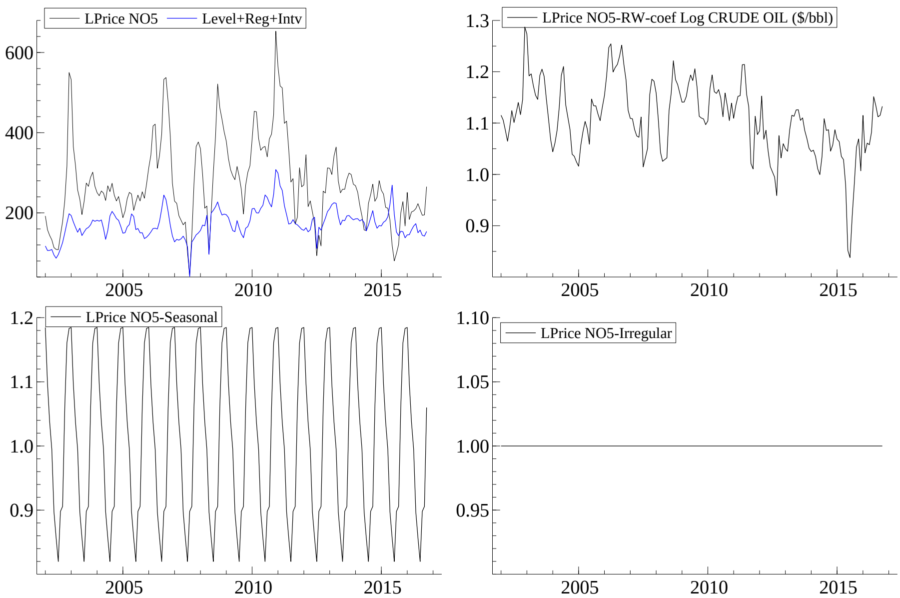

Figure 6.8: Model C - Component Graphics

modell C - graphics/Residuals.pdf

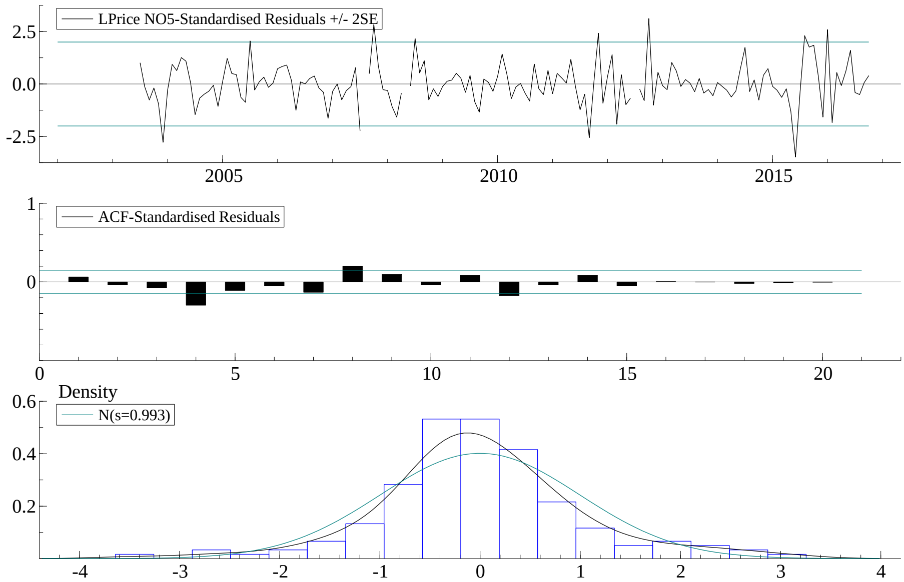

Figure 6.9: Model C - Residuals

The AIC-value of the model is -6.3085, which is equal to the AIC-values of models A and B. This is because the only difference is that we have changed all parameters that were found to have no variance from stochastic to fixed. Since this model has the best AIC value that was found in all models and has only one stochastic parameter, we conclude that model C is the best model we have found.

Chapter 7

Concluding remarks

In this paper we try to achieve further understanding as to how fundamental factors influence the Nord Pool spot electricity price. This analysis includes both endogenous factors, such as trend and seasonal components and exogenous factors, like temperature, reservoir deviation, and prices of oil, coal and gas. The results are not surprising and confirm many of our assumptions about the electricity price: The highest prices are usually in December and January, while the lowest prices are in June and July. The seasonal effects are positive in one half of the year, and negative in the other. Further, a slight upwards facing slope was detected, but the slope component was too small to make any conclusions about whether or not the electricity price is rising. Lastly, the coefficients of the explanatory variables did not show time-varying effects, except for the crude oil price, which itself did not show any clear trends. This gives reason to suspect that the market as a whole, at least in zone NO5, has not changed much since the beginning of our modelling period2002. The weighting of the coefficients is similar, there was no clear trend, and the seasonal component appears to be fixed as well.

Further work: - Forecasting/Out of sample testing - Lower frequency data (quarterly, yearly. Long-term forecasting - Higher frequency data (weekly, daily, intra-daily). Short-term forecasting

Bibliography

Aarrestad, K., Hatlen, L. M., 2015. Facts 2015, Energy and Water Resources in Norway. Norwegian Ministry of Petroleum and Energy.

Alexander, C., 2008b. Market Risk Analysis, Volume II, Practical Financial Econometrics. John Wiley.

Bergersen, B., Groth, L. H., Langseth, B., Magnussen, Ingrid H.and Spilde, D., Toutain, J. E. W., 2004. Energy consumption 2012 - household energy consumption. Norwegian Water Resources and Energy Directorate (16).

Bunn, D., Andresen, A., Chen, D., Westgaard, S., 2016. Analysis and forecasting of electricity price risks with quantile factor models. The Energy Journal 37 (2), 169–190.

Chen, D., Bunn, D., 2010. Analysis of the nonlinear respons of electricity prices to fundamental and strategic factors. IEE Transactions on power systems 25 (2), 595–606.

DK, E. N., 2016. Fuels.
URL http://www.energinet.dk/EN/KLIMA-OG-MILJOE/Miljoerapportering/Sider/Braendsler-forbrug-og-sammensaetning.aspx

Escribano A.and Pea, J.I.and Villaplana, P., 2011. Modelling electricity prices: International evidence. Oxford Bulletin of Economic Statistics 25 (2), 622–650.

ETS, E., 2016a. The eu emissions trading system.
URL https://ec.europa.eu/clima/policies/ets/index_en.htm

ETS, E., 2016b. Eu ets kernel description.
URL https://ec.europa.eu/clima/policies/ets/index_en.htm

Finland, S., 2016. Electricity and heat production by production mode and fuel in 2015.
URL http://www.stat.fi/til/salatuo/2015/salatuo_2015_2016-11-02_tau_001_en.html

Grull, G., kiesel, R., 2012. Quanitfying the co2 permit price sensitivity. Energiewirtsch 36, 101–111.

Hagfors, L., Kamperud, H., Paraschiv, F., Prokopczuk, M., Sator, A., Westgaard, S., 2016. Prediction of extreme price occurrences in the german day-ahead electricity market. Quantitative Finance 16 (12), 1929–1948.

Haldrup, N., Nielsen, M., 2005. A regime switching long memory model for electricity prices. Journal of Econometrics 135, 349–376.

Haldrup, N., Nielsen, M., 2006. Directional congestion and regime switching in a long memory model for electricity prices. Nonlinear analysis of Electricity Prices.

Hamilton, J. D., 1989. A new approach to the economic analysis of non-stationary time series and the business cycle. Econometrica 57 (2), 357–384.

Harvey, A., 2006. Forecasting with unobserved components time series models. Handbook of Economic forecasting.

Harvey, A. C., 1989. Forecasting, structural times series models and the kalman filter. Cambridge University Press.

Henley, A., Peirson, J., 1997. Non-linearitites in electricity demand and temperature: Parametric versus non-parametric methods. Oxford Bulletin of Economics and Statistics 59 (1), 149–162.

Higgs, H., Worthington, A., 2008. Stochastic price modelling of high volatility, mean-reverting, spike-prone commodities: The australian wholesale spot electricity market. Energy Economics 30 (6), 3172–3185.

Hoover, K. D., Perez, S., 2002. Another look at measures of forecast accuracy. Journal of Economic Methodology 7 (2), 195–210.

Huurman, C., Ravazzolo, F., Zhou, C., 2007. The power of weather: Some empirical evidence on predicting day-ahead power prices through weather forecasts. Tinbergen Institute Discussion Papers 36 (4).

Johnsen, T. A., 2007. Kvartalsrapport for kraftmarkaden. NVE.

Johnsen, T. A., 2008. Kvartalsrapport for kraftmarkaden. NVE.

Kaaresen, F., Husby, E., 2000. A joint state space model for electricity spot price and futures prices. Norsk Regnesentral.

Karakatsani, N. V., Bunn, D. W., 2008. Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. International Journal of Forecasting 24, 764–785.

Knittel, C., Roberts, M., 2005. An empirical examination of restructured electricity prices. Energy Economics 27 (27), 791–817.

Koopman, S. J., Ooms, M., Carnero, M., 2007a. Periodic seasonal reg arfima-garch models for daily electricity spot prices. Journal of the American Statistical Association 102 (447), 16–27.

Koopman, S. J., Ooms, M., Carnero, M., 2007b. Periodic seasonal reg arfima-garch models for daily electricity spot prices. Journal of the American Statistical Association 102 (447), 16–27.

Lucia, J., Schwartz, E. S., 2002. Electricity prices and power derivatives: Evidence from the nordic power exchange. Review of Derivatives Research 5, 5–50.

Lland, A., Dimakos, X., 2006. Modelling nord pools no1 area price. The Journal of Energy Markets 3 (1), 73–92.

Mirza, F. M., Bergland, O., 2012. Transmission congestion and market power: the case of the norwegian electricity market. The Journal of Energy Markets 5 (2), 59–88.

NordReg, 2014. Nordic Market Report. Nordic Energy Regulators.

Pettersen, F. E. L., 2012. Kvartalsrapport for kraftmarkaden. NVE.

Povh, M., Golob, R., Fleten, S.-E., 2010. Modelling the structure of long-term electricity forward prices at Nord Pool. Springer Berlin Heidelberg.

Proietti, T., 2002. Forecasting with structural time series models. A Companion to Economic Forecasting, 105–132.

Seljom, P., Rosenberg, E., Fidje, A., Haugen, J.-E., Meir, M., Rekstad, J., jarlset, T., 2011. Modelling the effects of climate change on the energy systema case study of norway. Energy Policy 39 (1), 7310–7321.

Spot, N. P., 2012. the Nordic Production Split 2004-2012. Nord Pool Spot.

Spot, N. P., 2014. the Nordic electricity exchange and the nordic model for a liberalized electricity market. Nord Pool Spot.

Unander, F., Ettestoel, I., ting, M., Schipper, L., 2004. Residential energy use: an international perspective on long-term trends in denmark, norway and sweden. Energy Policy 32 (32), 1395–1404.

Wei, W., 2006. Time Series Analysis - Univariate and Multivariate Methods. Pearson Education.

Welch, G., Bishop, G., 2006. An introduction to the kalman filter. University of North Carolina at Chapel Hill.

Weron, R., Misiorek, A., 2008. Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. International Journal of Forecasting 24, 744–763.

Appendix A

The linear Gaussian state space formulation

The general linear Gaussian state space model expresses all univariate state space models in one unified formulation, as follows:

$$\begin{aligned} y_t &= Z_t\alpha_t + \epsilon_t, \qquad \epsilon_t \sim NID(0, \sigma^2_\epsilon) \qquad t = 1, ..., T \\ \alpha_{t+1} &= T_t\alpha_t + R_t\eta_t, \qquad \eta_t \sim NID(0, Q^2_t) \end{aligned} \tag{A.1}$$

where y_t and ϵ_t are scalars, while the remaining terms denote vectors and matrices. We have that Z_t is an m x 1 observation vector, T_t is an m x m transition matrix and α_t is a m x 1 state vector of m unobservable components. R_t is a selection matrix that selects the rows of the state equation which have non-zero disturbance terms. The last term, η_t, is an r x 1 vector containing the r state disturbance with zero means, and unknown variances collected in an r x r diagonal matrix Q_t. In this general formulation, the top equation in **Eq. A.1** is called the observation equation, while the equation on the bottom is called the transition or state equation.

A.1 Structural Time Series Models in State Space form

A.1.1 The Local Level Model

The local level model allows the unobservable level component to vary in time. Hence, the state vector is consisting of only one element, making the local level model the simplest case of the general linear Gaussian state space model. If we define

$$\alpha_t = \mu_t, \qquad \eta_t = \eta_t, \qquad , Z_t = T_t = R_t = 1, \qquad Q^2_t = \sigma^2_\eta, \tag{A.2}$$

where all variables are of order 1 x 1 for t = 1,..,T, it is easily verified that **Eq. A.1** simplifies into the local level model described in **Eq. 5.1**.

A.1.2 The local linear Trend Model The local linear trend model generalizes the local level model by introducing a stochastic slope. The state vector of the local linear trend model is thus a 2 x 1 vector, because the model requires one element for the level and one element for the slope. If we define

$$\alpha_t = \begin{pmatrix} \mu_t \\ \nu_t \end{pmatrix}, \qquad \eta_t = \begin{pmatrix} \eta_t \\ \zeta_t \end{pmatrix}, \qquad Z_t = (1, 0), \qquad T_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \\ Q_t = \begin{bmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\zeta^2 \end{bmatrix}, \qquad and \qquad R_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{A.3}$$

It is easily verified that the scalar notation of **Eq. A.3** leads to the local linear trend model described in **Eq. 5.2**.

A.1.3 The local linear trend model with seasonal effects For the local linear trend model with trigonometric seasonal effects for monthly data, we define

$$\alpha_t = \begin{pmatrix} \mu_t \\ \eta_t \\ \tilde{\gamma}_{1,t} \\ \tilde{\gamma}^*_{1,t} \\ \tilde{\gamma}_{2,t} \\ \tilde{\gamma}^*_{2,t} \\ \tilde{\gamma}_{3,t} \\ \tilde{\gamma}^*_{3,t} \\ \tilde{\gamma}_{4,t} \\ \tilde{\gamma}^*_{4,t} \\ \tilde{\gamma}_{5,t} \\ \tilde{\gamma}^*_{5,t} \\ \tilde{\gamma}_{6,t} \\ \tilde{\gamma}^*_{6,t} \end{pmatrix}, \qquad \eta_t = \begin{pmatrix} \eta_t \\ \zeta_t \\ \tilde{\omega}_{1,t} \\ \tilde{\omega}^*_{1,t} \\ \tilde{\omega}_{2,t} \\ \tilde{\omega}^*_{2,t} \\ \tilde{\omega}_{3,t} \\ \tilde{\omega}^*_{3,t} \\ \tilde{\omega}_{4,t} \\ \tilde{\omega}^*_{4,t} \\ \tilde{\omega}_{5,t} \\ \tilde{\omega}^*_{5,t} \\ \tilde{\omega}_{6,t} \\ \tilde{\omega}^*_{6,t} \end{pmatrix},$$

$$
\mathrm{T}_t = \begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix}, \tag{A.4}
$$

A.1.4 The local level with explanatory variables

$$
\alpha_t = \begin{pmatrix} \mu_t \\ v_t \\ \beta_t \end{pmatrix}, \eta_t = \begin{pmatrix} \xi_t \\ \zeta_t \end{pmatrix}, T_t = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},
$$

$$
\mathrm{z}_t = \begin{pmatrix} 1 \\ 0 \\ x_t \end{pmatrix}, Q_t = \begin{bmatrix} \sigma_\xi^2 & 0 \\ 0 & \sigma_\zeta^2 \end{bmatrix}, R_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}
$$

The Local Level Model with Intervention Variables

The local level model with an intervention variable has the matrix representation

$$\alpha_t = \begin{pmatrix} \mu_t \\ \lambda_t \end{pmatrix}, \qquad \eta_t = \eta_t, \qquad Z_t = (1, i_t), \qquad T - t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{A.5}$$
$$Q_t = \sigma_\eta^2, \qquad and \qquad R_t = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

Appendix B

Test Statistics and Diagnostics Tests

B1: UNIT ROOT TESTS

Statistical tests of the null hypothesis that a time series is non-stationary against the alternative hypothesis that it is stationary, are called *unit root tests:*

$$H_0\colon X_t \sim I(1)$$
$$H_1\colon X_t \sim I(0).$$

The interpretation of this is that an autoregressive process is stationary if and only if the roots of its characteristic polynomial lie strictly inside the unit circle (Alexander, 2008b). The *Dicky-Fuller (DF) test* is the most basic unit root test. The test is based on the *Dicky-Fuller regression*, which is a regression of the form

$$\Delta X_t = \alpha + \beta X_{t-1} + \varepsilon_t.$$

The test statistic is the *t* ratio on $\hat{\beta}$, and it is a one-sided test for

$$H_0\colon \beta = 0$$
$$H_1\colon \beta =< 0.$$

If a test statistic falls into the critical region, we conclude that the process is stationary at the confidence level prescribed by the critical region. However, a problem with the Dickey-Fuller tests is that their critical values are biased if there is autocorrelation in the residuals of the Dickey-Fuller regression. The *augmented Dicky-Fuller (ADF) test* solves this issue by including as many lagged dependent variables as necessary to remove any autocorrelation in the residuals. The ADF-test of order *q* is based on the regression

$$\Delta X_t = \alpha + \beta X_{t-1} + \gamma_1 \Delta X_{t-1} + ... + \gamma_q \Delta X_{t-1} + \varepsilon_t.$$

The test proceeds as in the ordinary DF-test above, except that the critical values depend on the number of lags, q, that has been included. In this paper, we run all ADF-test with one lag,i.e. q = 1.

B2: IN-SAMPLE PROPERTIES OF THE RESIDUALS

Linear Gaussian state space models are based on three assumptions concerning the residuals of the analysis. In decreasing order of importance, the residuals should satisfy independence, homoscedasticity and normality. After selecting the regression variables and fitting a regression model, we therefore perform formal tests to reveal whether the assumptions of the model have been satisfied. In accordance with the literature on structural time series models (Harvey, 1989 and Durbin and Koopman, 2001), these tests are applied to the standardized one-step ahead prediction errors, which are defined as

$$e_t = \frac{\eta_t}{\sqrt{F_t}}, \qquad t = 1, ..., T$$

where η_t is the prediction errors and F_t is their variances.

Tests for Independence

For independence, we first consider whether individual autocorrelations of the residuals for lag 1 up to lag 12 are significant. If the residuals are significantly independent, they should not be greater than $\pm 1.96/\sqrt{n}$ according to a confidence interval of 95% (Commandeur and Koopman, 2007). Secondly, we apply the ***Ljung-Box Q-test***, which is a standard test for autocorrelation in a time series. That is, it tests whether or not autocorrelation is present across a number of textit k lags. Letting r_k denote the autocorrelation of the residuals of lag k,

$$r_k = \frac{\sum_{t=1}^{n-k}(e_t - \bar{e})(e_t - \bar{e})}{\sum_{t=1}^{n}(e_t - \bar{e})^2},$$

where $\bar{e}$ is the mean of the n residuals, the Ljung-Box test statistic is expressed as

$$Q(k) = n(n+2)\sum_{l=1}^{k}\frac{r_l^2}{n-l}$$

where n is the number of observations and k refers to the number of lags we have included in our H_0 hypothesis. Under the null hypothesis, the Q-statistic follows a χ^k-distribution.Thus, we rejects the null hypothesis if

$$Q > \chi^2_{1-\alpha,h},$$

where h is the number of lags being tested.

Testing for homoscedasticity

Wei (2008) defines homoscedasticity as the phenomenon of constant variance in a sequence of random variables. We assess homoscedasticity of the standardized prediction errors with the following test statistic

$$H(h) = \frac{\sum_{t=n-h+1}^{n} e_t^2}{\sum_{t=d+1}^{d+h} e_t^2},$$

where n is the number of observations, d is the number of diffuse initial elements and h is the nearest integer to $(n-d)/3$. Here, $H(h)$ tests whether the variance of the residuals in the first third part of the series is equal to the variance of the residuals corresponding to the last third part of the series. The test statistic is F-distribution with (h,h) degrees of freedom under the null hypothesis of equal variances. This suggests the use of a two-tailed test, locating the critical values for the upper and lower 2.5% in the F-distribution.

$$\begin{array}{ll} if \;\; H(h) > 1, & test \;\; H(h) < F(h, h, 0.025) \\ if \;\; H(h) < 1, & test \;\; 1/H(h) < F(h, h, 0.025) \end{array}$$

Jarque-Bera (JB) Test for Normality

To justify the assumptions of normality in the standardized prediction errors, we apply the *Jarque-Bera test*. This test is based on estimates of the skewness, S, and kurtosis, K, in a sample. The test statistic is given by

$$JB = n\left(\frac{S^2}{6} + \frac{(K-3)^2}{24}\right),$$

where n is determined by the number of observations and the number of covariates used in the regression. Under the null hypothesis that the distribution is normal, JB is tested against a χ^2-distribution with two degrees of freedom. In accordance with the critical values of the χ^2-distribution, the null hypothesis is rejected at 5% significance level whenever *JB* $>$ 5.99.

B3: EVALUATION CRITERIA OF MODEL COMPARISON

Akaike information criterion (AIC)

The *Akaike information criterion* is a measure of the relative quality of statistical models for a given set of data. Hence, it is commonly used for model selection. By penalizing models for using additional parameters, this evaluation criteria discourages overfitting.In accordance with Commandeur and Koopman (2007), we use the following AIC definition:

$$AIC = \frac{1}{n}[-2nlog Ł_d + 2(q + w)],$$

where n is the number of observations in the time series, log $Ł_d$ is the value of the diffuse log-likelihood function which is maximised in state space modelling, *q* is the number of diffuse initial values in the state, and *w* is the total number of disturbance variances estimated in the analysis. The model that yields the lowest value of the AIC is considered to be the best fit (Alexander, 2008b).