FORECASTING THE DISTRIBUTION OF NORD POOL SPOT ELECTRICITY PRICES - A BOTTOM-UP APPROACH

Marius Fuglerud

Knut Erik Vedahl

Institute of Industrial Economy and Technology Management

Norwegian University of Science and Technology

Project assignment TIØ4550

December 10, 2010

ABSTRACT

We forecast the distribution of the Nord Pool system price during two years, based on a bottom up model. Empirical models for consumption, generation, exchange and marginal water value are established and combined using a market equilibrium. Separate scenarios for the most influential risk factors lead to different price scenarios. The model provides a picture of the uncertainty in future spot prices and consequently the risks related to price variation. Compared to stochastic optimization models, this approach is more intuitive and less computationally intensive. We conclude that the forecasts of the price distribution seem realistic, except under certain conditions. Price jumps are not modeled properly, and some interaction between underlying factors is not captured. Distribution forecasts can be used in decision making related to the future electricity price. In particular, the model can support midterm planning for hydro power companies and risk management in energy-intensive industry.

CONTENTS

List of Tables

List of Figures

1. INTRODUCTION

In this thesis, we develop a bottom-up model for the Nord Pool system price, building upon the work of Vehviläinen and Pyykkönen (2005). Separate empirical models for consumption, generation from different technologies, exchange and the marginal water value will be established, and combined in a relationship for the market equilibrium. Scenarios for the most influential risk factors will be used as inputs to the model. The output will be a forecasted distribution of the system price with weekly granularity over a period of two years.

Predictions of the distribution of future system prices are of importance for players in the deregulated Nordic power market. Our model is targeted for midterm planning, and represents a more intuitive and less computationally intensive methodology than the widely used stochastic optimization models. More elaborated models for area prices are also needed in the planning process, but our model may represent a complementary and valuable indication of the future spot price distribution.

The initial motivation behind our approach of building empirical models for each component in the power system stems from Tipping (2007). Tipping demonstrates that an empirical function for the marginal water value leads to well-performing forecasts for the spot price in New Zealand. Following the path of Tipping, two questions arise: Will a similar empirical model for the marginal water value perform well in the Nordic market? Does it make sense to build empirical functions in the same manner for other factors influencing the spot price?

Part of the answer is presented by Vehviläinen and Pyykkönen (2005). They estimated models for consumption, generation and the marginal water value from historical data for the Nord Pool area. Through a market balance and scenarios for temperature and precipitation, future spot price paths with monthly granularity were generated.

We find the approach presented by Vehviläinen and Pyykkönen highly interesting, but also see several possible ways of improvement. In this thesis, their model will be extended in multiple ways. First, using precipitation as a risk factor may be too inaccurate, especially regarding determination of whether the precipitation falls as rain or snow. Instead of precipitation, we will use reservoir inflow and snow reservoirs directly as risk factors. As additional risk factors, we introduce wind power generation and the spot price at the German electricity exchange, EEX. Wind power generation is included in order to avoid underestimation of the total generation, while the close connection between Nordic and continental power markets makes it necessary to include the impact of those markets. For the same reason, the net exchange between the Nordics and other markets will be modeled and included in a market balance. The marginal water value functions used by Tipping (2007) and Vehviläinen and Pyykkönen (2005) will be compared, and possible improvements will be investigated.

The empirical models in the paper by Vehviläinen and Pyykkönen are what we will denote fundamental models: Models where functional form and variables are selected based upon physical or economic fundamentals. Despite this approach provides intuitive and relative simple models, model evaluation is difficult. The forecasting abilities may be reduced due to undesired statistical properties as non-stationarity and autocorrelation in residuals. Therefore, we will introduce statistical models as SARIMA models, error correction models and linear transfer functions for some of the components in the system.

For example, the performance of an error correction model and a linear transfer function will be compared in modeling of electricity consumption. The error correction approach is inspired by Johnsen and Willumsen (2010), while the use of a linear transfer function builds on the work of Murray and Ringwood (1994).

A brief overview of the framework is given in Figure 1. First, the risk factors are modeled. As the figure indicates, all other variables are assumed to be dependent on one or more of these risk factors. Second, consumption and generation from baseload technologies are estimated. Regulated generation, hydro reservoirs and the system price will be simultaneously calculated using a market balance, a reservoir balance and an estimated function for the marginal water value. The system price is defined by the intersection of the aggregate supply and demand curves for the Nord Pool area, without taking account to transfer constraints (Nord Pool Spot, 2010). Therefore, transmission capacities are not included in the framework.

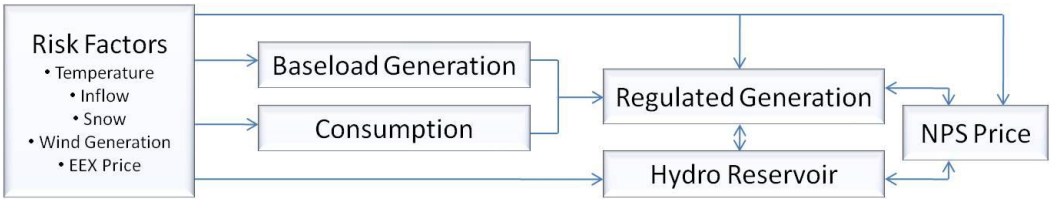

FIGURE 1. OVERVIEW OF THE FRAMEWORK

The next chapter outlines the framework in detail. Chapter 3 provides an overview of the data used in the estimation process. The estimated models, forecasts and simulations are presented in Chapter 4, while we in Chapter 5 discuss the results and draw conclusions.

2. FRAMEWORK

Summary of notations:

t -	Time period index	$G_{hydro,t}$ -	Hydro generation
I_t -	Inflow	$G_{hydroR,t}$ -	Regulated hydro generation
$R_{W,t}$ -	Water reservoir level	$G_{hydroU,t}$ -	Unregulated hydro generation
$R_{WCAP,t}$ -	Water reservoir capacity	$G_{wind,t}$ -	Wind generation
$R_{S,t}$ -	Snow reservoir level	$G_{condense,t}$ -	Condense generation
$R_{TOT,t}$ -	Total reservoir level	$G_{baseload,t}$ -	Baseload generation
MWV_t -	Marginal water value	$G_{CHPD,t}$ -	CHP district generation
Q_t -	Consumption	NI_t -	Net import
HDD_t -	HDD	$S_{NP,t}$ -	Nord Pool system price
DL_t -	Day length	$S_{EEX,t}$ -	EEX spot price
RTI_t -	Retail trade index	$S_{Coal,t}$ -	Coal spot price
		$S_{Gas,t}$ -	Natural gas spot price

2.1 RISK FACTORS

2.1.1 SARIMA MODELS FOR INFLOW, SNOW RESERVOIRS AND HDD

Reservoir inflow, snow reservoirs and HDD vary in a seasonal pattern through the year. In order to capture the seasonality, we use SARIMA (seasonal autoregressive integrated moving average) models for each process. SARIMA models have often been used to model seasonality in factors influencing electricity supply and demand, see e.g. Krogh Kristoffersen (2007) for an example of use of a SARIMA process to model inflow. A general SARIMA process can be written as (Pankratz, 1991):

$$\emptyset(B^S)\emptyset(B)\nabla_S^D\nabla^d Y_t = C + \theta(B^S)\theta(B)u_t \quad [1]$$

where

- Y_t is the time series to be modeled
- S is the length of one season
- $\nabla_S^D\nabla^d = (1-B^s)^D(1-B)^d$ denotes the differencing transformation, where D and d is the degrees of seasonal and non-seasonal differencing, respectively
- $\emptyset(B)$ is a non-seasonal autoregressive (AR) polynomial of order p
- $\emptyset(B^S)$ is a seasonal autoregressive (SAR) polynomial of order P
- $\theta(B)$ is a non-seasonal moving-average (MA) polynomial of order q
- $\theta(B^S)$ is a seasonal moving-average (SMA) polynomial of order Q
- C is the intercept
- u_t is the time series of residuals

In the case of weekly granularity and a period of one year, *S* equals 52. A single seasonal difference, *D = 1*, is applied to all three series due to the strong seasonal pattern. The lag structure *(p, q, P, Q)* will be optimized using an information criterion. However, we restrict $P + Q \leq 1$, due to the characteristics of the processes (Nau, 2010).

2.1.2 ARIMA MODEL FOR WIND POWER GENERATION

Wind power generation is simulated using an ARIMA(p, d, q) process. The advantage of an ARIMA model, for instance compared to a Brownian motion, is that the possible autocorrelation in wind speeds from one week to the next will be captured.

2.1.3 MEAN-REVERTING PROCESS FOR THE EEX SPOT PRICE

The EEX spot price is modeled as a mean-reverting process, motivated by the fact that electricity prices tend to move to back to an average equilibrium price after jumps created by events like plant outages and cold waves (Blanco and Soronow, 2001). The starting point in developing stochastic processes with mean reversion is the arithmetic Ornstein-Uhlenbeck process (Dixit and Pindyck, 1994):

$$dS(t) = \eta\big(\bar{S} - S(t)\big)dt + \sigma dZ \qquad [2]$$

where

- S is the spot price
- η is the speed of mean reversion
- $\bar{S}$ is the long-run equilibrium price
- σ is the volatility
- dZ is the increment to a Wiener process, which creates random moves in the price

In order to avoid negative prices, we model the natural logarithm of the price series:

$$d[\ln(S(t))] = \eta[ln(\bar{S}) - \ln(S(t))]dt + \sigma dZ \qquad [3]$$

The stochastic differential equation (3) is the limit of the following AR-process, which will be used to create price scenarios:

$$\ln(S_t) - \ln(S_{t-1}) = \ln(\bar{S})(1 - e^{-\eta}) + (e^{-\eta} - 1)\ln(S_{t-1}) + \epsilon_t$$
$$\epsilon_t \sim N(0, \sigma_\epsilon^2), \quad \sigma_\epsilon^2 = \frac{\sigma^2}{2\eta}(1 - e^{-2\eta}) \qquad [4]$$

Using a mean-reversion model to simulate the EEX price is governed by some weaknesses. First, future jumps will not be simulated. Second, the rate of mean reversion is assumed to be constant. In reality, the speed of reversion is dependent on the nature of the previous jump.

2.2 CONSUMPTION

2.2.1 FACTORS INFLUENCING CONSUMPTION

In their model for electricity consumption in Norway, Johnsen and Willumsen (2010) introduce HDD (heating degree days), day length, wind speed, electricity price, price of alternative fuels, economic activity level and holiday dummies as explanatory variables. The intuition of including these variables can be summarized as follows:

- *Heating degree days (HDD):* Electricity consumption in the Nordics is closely tied to the temperature, since electricity is one of the most important heating sources. The concept of HDD provides an intuitive link between temperature and heating demand, by assuming that the heating demand starts at a certain temperature, the critical temperature. The HDD for a given period is here defined as the number of degrees the average temperature of the period is below the critical temperature (Murray and Ringwood, 1994)
- *Day length:* A decrease in day length should increase the demand for lightning
- *Wind speed:* An increase in the average wind speed reduces the effective temperature, thus the heating demand should increase
- *Electricity price:* The electricity demand should decrease if the price increases
- *Price of alternative fuels:* If the price of substitutes to electricity increases, the electricity consumption should increase
- *Economic activity level indicator:* If the aggregated level of consumer spending increases, it is reasonable that also electricity consumption increases
- *Holiday dummies:* Electricity consumption may be reduced at holidays, when factories are temporarily shut down and office buildings closed

By including the spot price as a factor influencing consumption, the consumption cannot be simulated independently from the price model. In order to keep the model simple, the spot price will not be included in the demand equation. This choice is justified by studies of the price elasticity of electricity demand: The elasticity is found to be low, although significant. A study from Canada indicates short term price elasticity of -0.16 for households and -0.11 for industry (Wangensteen, 2007).

2.2.2 MODELS FOR CONSUMPTION

To incorporate the above factors we consider two classes of models, both frequently used to model electricity demand: Error correction models and dynamic regression models.

ERROR CORRECTION MODELS

The general error correction model with one explanatory variable can be written as (Brooks, 2008):

$$\Delta Y_t = \beta_1 \Delta X_t + \beta_2 (Y_{t-1} - \gamma X_{t-1}) + u_t \quad [5]$$

where

- Y_t is the time series of the dependent variable
- X_t is the time series of the independent variable
- β_1 and β_2 are parameters to be estimated
- γ is the cointegrating coefficient
- u_t is the time series of residuals

The model can be extended with more explanatory variables; in our case the whole armory of variables listed in 2.2.1 will be applied. [5] assumes that Y is integrated of order one and that Y and X are cointegrated, hence the linear combination $Y_t - \gamma X_t$ must be stationary. The model states that Y changes from *t-1* to *t* due to changes in X, but also due to eventual disequilibrium from the long-term relationship $Y_{t-1} - \gamma X_{t-1}$ (Brooks, 2008). Johnsen and Willumsen (2010) used an error correction model in their work. Albeit good fit with historical data (R^2 = *89%*), the model had two weaknesses:

- Residuals were autocorrelated
- The consumption data used in the model was not non-stationary, hence an error correction model should not be the appropriate choice

DYNAMIC REGRESSION MODELS

A dynamic regression model with one input can be written as (Pankratz, 1991):

$$Y_t = C + v(B)X_t + N_t \quad [6]$$

where

- Y_t is the output time series
- X_t is the input time series
- $v(B)$ is a linear transfer function
- N_t is the disturbance series, which may be autocorrelated

The model can easily be generalized by adding further inputs.

After the model [6] is estimated, the estimated disturbances $\widehat{N}_t$ are examined for non-stationarity. By rearranging [6]:

$$\widehat{N}_t = Y_t - C - v(B)X_t \quad [7]$$

If the residual autocorrelation function shows signs of non-stationary, the difference operator has to be applied to both sides of [6], which mean that the whole model has to be differentiated. To get rid of autocorrelation in the disturbances, a SARIMA model [1] has to be estimated for $\widehat{N}_t$. Substituting the necessary difference operators and the disturbance SARIMA model into [6], we end up with:

$$\nabla_S^D \nabla^d Y_t = \nabla_S^D \nabla^d v(B)X_t + \frac{\theta(B^S)\theta(B)}{\emptyset(B^S)\emptyset(B)} u_t \quad [8]$$

where

- u_t is the time series of residuals, which should not be autocorrelated

The formulation of the transfer function $v(B)$ can be selected by examining the impulse response function. Given the characteristics of electricity consumption, it is reasonable that the consumption responds immediately to changes in the explanatory variables (Pankratz, 1991). As our base case, we therefore use transfer functions without lags of the input.

2.3 GENERATION

In this thesis, the Nordic electricity generation is divided into five categories due to the technology: Hydro power, wind power, condensing power, CHP district and baseload. As different factors affect the generation level of these technologies, the generation level for each is modeled separately. Condensing power represents generation at a coal or gas fired steam power plant where the energy of the steam is used entirely for electricity generation. As the name suggests, the steam is condensed to water after the turbine. On the other hand, combined heat and power generation (CHP) utilizes the waste heat in district heating (CHP district) or as a process stream in the industry (CHP industry) (Nordel, 2009). CHP industry and nuclear power are here classified as baseload generation[1].

The generation from each category in 2008 is summarized in Table 1. Hydro power is the major generation technology, with a 57% share of the total generation.

TABLE 1. ELECTRICITY GENERATION IN THE NORD POOL AREA IN 2008 (NORDEL, 2009)

Electricity generation 2008 [GWh]					
	Denmark	Finland	Norway	Sweden	Total
Hydro power	27	16889	140663	68429	226008
Wind power	6977	262	917	1995	10151
Condensing power	11732	8209	432	866	21239
CHP, district heating	14034	14659	119	7209	36021
Baseload:					
- CHP, industry	1879	12080	596	6256	20811
- Nuclear power	0	22038	0	61266	83304
Total	34648	74137	142727	146021	397533

The sum of total generation and net import[2] to the Nord Pool area should be equal to the electricity consumption:

$$Q_t = G_{hydro,t} + G_{wind,t} + G_{condense,t} + G_{baseload,t} + G_{CHPD,t} + NI_t \quad [9]$$

Generation for all technologies except hydro will be modeled in the following sections. Then, the hydro generation is estimated using the market balance [9].

2.3.1 BASELOAD

Baseload units have a high cost of adjusting the generation level and a low marginal cost (Vehviläinen and Pyykkönen, 2005). Therefore, baseload generation is run more or less independently of short term spot price variations, and has a lower variation in output than other technologies. Due to the relative constant patterns from year to year, Vehviläinen and Pyykkönen (2005) model the future generation in time period *t* as the average historic generation (assuming no additional generation capacity in the forecast period):

[1] Strictly speaking, CHP district is also a baseload technology. For convenience, CHP district is not included in the baseload term in this thesis

[2] Total import less total export

$$G_{baseload,t} = \bar{G}_{baseload,t} \quad [10]$$

Hence, the forecast will take account of both seasonal variations caused by consumption, and planned outages which typically occur at the same time each year (in the summer, when consumption is low). Using this approach, the risk for unexpected temporary outtakes is not taken into account in a systematic manner.

2.3.2 CHP DISTRICT

The level of CHP district generation (G_{CHPD}) depends on the temperature, as the waste heat is used for heating purposes. Vehviläinen and Pyykkönen (2005) introduce the following model:

$$G_{CHPD,t} = C_1 + C_2[(T_t - T_{min})^+ - (T_t - T_{max})^+] \quad [11]$$

The generation is constant (C_1) for temperatures above T_{max}, a linear function of temperature in the interval $\langle T_{max}, T_{min} \rangle$, and continues at the maximum generation capacity for temperatures lower than T_{min}.

In our model, CHP district generation is a function of HDD, and the generation is assumed to be constant when the temperature exceeds the critical temperature. Thus, the model transforms to:

$$G_{CHPD,t} = C_1 + C_2\left[HDD_t - \left(HDD_t - HDD_{max,t}\right)^+\right] \quad [12]$$

The impact of including the maximum generation capacity will be tested. Without this limit, the model simplifies to a linear regression model:

$$G_{CHPD,t} = C_1 + C_2 HDD_t \quad [13]$$

2.3.3 CONDENSE

Vehviläinen and Pyykkönen (2005) states that the asking price of condense power generation is a linear function of the generation level ($G_{condense}$). Substituting the system price (S_{NP}) for the asking price, their model can be written:

$$S_{NP,t} = (C_1 G_{Condense,t} + C_2)^+ \quad [14]$$

This attempt to replicate the supply curve will be compared with statistical models, where more explanatory variables are included. In addition to the electricity spot price, the generation may be affected by the consumption level (Q), as well as fuel prices (S_{Gas} and S_{Coal}). Including these parameters gives:

$$G_{Condense,t} = C_1 + C_2 S_{NP,t} + C_3 Q_t + C_4 S_{Coal,t} + C_5 S_{Gas,t} + u_t \quad [15]$$

An error correction model is estimated if the condense generation series turns out to be non-stationary:

$$\begin{aligned} d\left(G_{Condense,t}\right) = {} & C_1 d\left(S_{NP,t}\right) + C_2 d(Q_t) + C_3 d\left(S_{Coal,t}\right) + C_4 d\left(S_{Gas,t}\right) \\ & + C_5\left(G_{Condense,t-1} - \gamma_1 S_{NP,t-1} - \gamma_2 Q_{t-1} - \gamma_3 S_{Coal,t-1} - \gamma_4 S_{Gas,t-1}\right) + u_t \end{aligned} \quad [16]$$

2.4 NET IMPORT

The Nordic countries exchange electricity with the Netherlands, Germany, Poland, Estonia and Russia. Table 2 shows that the highest transmission capacity is shared with Germany. Electricity exchange between two markets is influenced by the price difference and restricted by transmission capacities (Wangensteen, 2007). For simplicity, we assume that all exchange occurs with the German market. This simplification is supported by the correlation between the spot prices in the German power market and the markets in the Netherlands (SKM Energy Consulting, 2003). On the other hand, Bobinaitė et. al. (2006) found that the correlation between the spot price in Poland and the EEX price was low, because of an illiquid market in Poland. Similar characteristics may be reasonable for the electricity price in Estonia and Russia.

TABLE 2. TRANSMISSION CAPACITY SHARED WITH OTHER MARKETS IN 2008 (NORDEL, 2009)

Transmission capacity 2008 [MW]		
	From Nord Pool	To Nord Pool
Denmark - Germany	2 250	1700
Finland - Russia	0	1560
Finland - Estonia	350	350
Norway - Russia	50	50
Norway - Netherlands	700	700
Sweden - Germany	600	600
Sweden - Poland	600	600

We will model the net import as a function of the price difference between the Nord Pool spot price (S_{NP}) and the German spot price (S_{EEX}), assuming that the transmission capacity restriction is not active:

$$NI_t = C_1 + C_2(S_{NP,t} - S_{EEX,t}) \quad [17]$$

The feedback loop between the price difference and the electricity exchange is not included in the model: A high price difference will lead to high electricity exchange, but likewise will high exchange reduce the price difference. Also, the weekly granularity may be disadvantageous: A close relationship between instantaneous electricity exchange and instantaneous price difference does not necessarily implicate that the difference of the weekly average prices can explain the weekly accumulated net import.

2.5 HYDRO GENERATION AND THE MARGINAL WATER VALUE

Hydro power generators continuously face the problem of maximization of the value of the water in their reservoirs. If an amount of water is released now, it has a value given by the spot price. However, this amount of water may have a higher value if it is stored in the reservoir until prices are higher and the availability of water is more constrained. The expected marginal water value (MWV) is the expected opportunity cost of producing a marginal unit today instead of storing that water to a later period. In equilibrium, the asking price for a given hydro generator is equal to the marginal water value of the reservoirs to that generator[3]. If a hydro plant is the marginal unit on the aggregated supply curve, the

[3] Neglecting other variable costs and assuming an efficient market

system price should be equal to the marginal water value of this generator. The expected marginal water value depends on reservoir level, expected future spot price, expected future inflow and expected future demand. Hydro generators typically use complex stochastic optimization methods to estimate the water value of their reservoirs, and use the estimation output as a decision tool in planning of generation. (Wangensteen 2007, Tipping 2007 and Doorman 2010)

Complex stochastic optimization methods are outside the scope of this study, and another approach to marginal water value calculation is needed. Tipping (2007) points out that the median historic storage trajectory of a water reservoir is a reasonable estimate of the optimal reservoir management strategy. The optimal strategy is to keep the expected marginal water value constant over time, which means that the marginal water value can be approximated as constant along the median trajectory. Due to the stochastic nature of inflow, it is not possible to keep the reservoir equal to the median trajectory in every period. If the reservoir level in a given week is less than the historic median for that week, there is an increased risk of running out of water in the next weeks, and the value of storing water will therefore increase. On the other hand, if the reservoir level is higher than the median, the risk for future spills will increase; hence the marginal water value will decrease. MWV curves constructed by optimization based reservoir management models (Figure 2) illustrate this relationship: The water value increases exponentially when the reservoir level decreases, and decreases exponentially when the reservoir level is increasing (Tipping 2007, Batstone 2003).

In addition to the deviation from the median storage trajectory, the water value may depend on the season. If the reservoir level in the winter season is lower than the average level for that time of the year, this should increase the water value more than the same deviation in the summer season. In the winter, inflow is low and demand is high. Thus, the risk of running out of water is much higher than in the summer, when the inflow is high relative to the demand. This intuitively relationship is confirmed by an empirical study by Johnsen (2001).

A major simplification is now introduced: We accumulate the hydro reservoirs for all generators in the market, and assume that the marginal water value of the marginal generator equals the marginal water value of the accumulated reservoir.

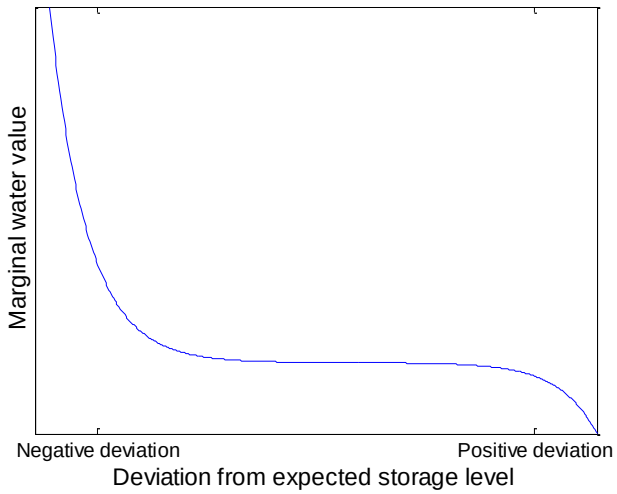

FIGURE 2. CONCEPTUAL ILLUSTRATION OF A MWV CURVE, AFTER BATSTONE (2003)

A mathematical formulation of the approximated water value has to include all the above mentioned factors. Tipping (2007) proposes the following function, motivated by empirical evidence from the market in New Zealand:

$$MWV_t = c_j + w_j e^{x_j (y_j + 0.01 RSL_{t-1})^{z_j}} \quad [18]$$

where

- RSL_t is the the relative storage level in period *t*, which is defined by Tipping as the difference between the actual storage level and the 45-day moving average of the tenth percentile trajectory
- *c*, *w*, *x*, *y* and *z* are constants to be estimated. Note that the *x* must have a negative sign
- *j* is an index denoting the season

Thus, the marginal water value increases exponentially when the relative storage level decreases.

Another function is suggested by Vehviläinen and Pyykkönen (2005):

$$MWV_t = c_{WVslope}\left(R_{TOT,t} - C_{TOT,t}\right)\left(1 + e^{-c_{ResPenalty}\left(R_{W\%,t} - c_{ResMin\%}\right)}\right) + c_{WVLevel} \quad [19]$$

where

- $C_{TOT,t}$ is the historical average total reservoir in period *t*
- $R_{W\%,t}$ is the filling degree of the hydro reservoir in period *t*, calculated as $\mathrm{R_{W,t}/R_{WCAP}}$
- $c_{WVslope}$, $c_{ResMin\%}$ and $c_{WVLevel}$ are constants to be estimated

[19] states that the water value is proportional to the deviation from the historical average total reservoir, but increases exponentially if the hydro reservoir decreases to an estimated limit, $c_{ResMin\%}$. As apparent, there are a few interesting differences between [18] and [19]. First, Vehviläinen and Pyykkönen incorporate the effect of the snow reservoir on the water value. Second, instead of proposing that the exponential increase is dependent on the difference between the reservoir level and a time-of-the-year dependent limit (the tenth percentile trajectory), they state that the exponential increase depends on the difference between the reservoir level and an absolute limit ($c_{ResMin\%}$).

Note that neither [18] nor [19] take account to the exponential decrease in the marginal water value due to spill risk for high reservoir levels (Figure 2). Vehviläinen and Pyykkönen model behavior near the maximum reservoir level through a division between regulated and unregulated inflow. The unregulated inflow is assumed to be of the form:

$$I_{U\%,t} = C_1 + e^{C_2(1 - R_{W\%,t})} \quad [20]$$

where

- $I_{U\%,t}$ is unregulated inflow in percent of total inflow

A constant share of the inflow, C_1, pass through run-of-river hydro plants without storage capacity. Of the (1-C_1)% of the inflow which arrives to plants with storage capacity, $e^{C_2(1-R_{W\%,t})}$% are sent directly through the turbines in order to reduce the risk for later spillages. This exponential increase in unregulated inflow with the reservoir level corresponds to the decrease in marginal water value with the filling degree.

Combining the aspects presented above, we will use the following representation for the marginal water value:

$$MWV_t = \delta_t f_1(I_t, R_{TOT,t}, Q_t)\left(1 + e^{-\beta_1\left(R_{W\%,t} - C_{W\%,t}\right)}\right) + (1-\delta_t) f_2(I_t, R_{TOT,t}, Q_t)\left(1 + e^{-\beta_2\left(R_{W\%,t} - C_{W\%,t}\right)}\right) \quad [21]$$

where

- δ_t is a binary variable denoting the season. Two seasons will be applied: Season 1 is the snow melting period, characterized by high inflow and increasing water reservoirs. Season 2 is the snow accumulation period, where inflow is low relative to demand and the water reservoirs are declining
- f_1 and f_2 are linear functions of inflow, the total reservoir level and consumption
- $R_{W\%,t} - C_{W\%,t}$ is the relative storage level of the hydro reservoirs. $C_{W\%,t}$ may either be the average reservoir level, the tenth percentile, a moving average of one of these or a time-of-the-year independent constant; the representation will be selected based upon fit with historical data
- β_1 and β_2 are constants determining the impact of negative relative storage levels on the MWV

By including the functions f_1 and f_2, the potential impact of other variables than the hydro reservoir is included. The optimal functional form of f_1 and $f_2$may be non-linear, but linear functions are used in order to keep the model simple. Thus, our model is a generalization of Vehviläinen and Pyykkönen's model for the marginal water value. Vehviläinen and Pyykkönen's function for unregulated inflow will be used in order to model the behavior for high reservoir levels. For simplicity, we assume that the unregulated generation is dependent on the reservoir level at the end of the previous period (*t-1*), and not the reservoir level at the end of the current period as stated in [20].

To keep track of the hydro reservoir, the following reservoir balance is used:

$$R_{W,t} = R_{W,t-1} + I_t - G_{hydro,t} - Spill_t \quad [22]$$

where

$$G_{hydro,t} = \mathrm{G}_{hydroR,t} + \mathrm{G}_{hydroU,t} \quad [23]$$

$R_{W,t}$ is the reservoir level at the end of period *t*. Spill arises when reservoirs are full or when power generators release water in order to prevent full reservoirs. Since $R_{W,t}$ is the reservoir level at the end of period *t* and the marginal water value [21] is assumed to be equal to the asking price of the marginal hydro generator in the system, we implicitly assume that hydro generators have access to short-term planning tools which give accurate forecasts of the reservoir level one week in advance.

2.6 THE SYSTEM PRICE

As stated in previous chapter, the system price should equal the marginal water value of the aggregated power system, given that a hydro plant is the marginal unit on the supply curve. However, there are two regulating power technologies in the system: Hydro and condense. Vehviläinen and Pyykkönen state that condense power is the marginal unit when there are not enough hydro generation capacity to meet the demand after condense power has been generated up to an asking price equal to the water value. Using this approach requires detailed data on asking prices for the hydro power generators in the system, in order to estimate equation [21]. Therefore, we assume that hydro power always is the marginal unit on the supply curve. This approximation is reasonable, as the installed capacity of condense power is only 11.8% of the installed capacity of hydro power in the Nord Pool area (Nordel, 2009). Substituting the system price for the marginal water value in equation [21] yields:

$$S_{NP,t} = \delta_t f_1(I_t, R_{TOT,t}, Q_t)\left(1 + e^{-\beta_1\left(R_{W\%,t} - C_{W\%,t}\right)}\right) + (1-\delta_t) f_2(I_t, R_{TOT,t}, Q_t)\left(1 + e^{-\beta_2\left(R_{W\%,t} - C_{W\%,t}\right)}\right) \quad [24]$$

The impact of the EEX spot price on the Nord Pool system price is included indirectly through the market balance [9]: Net import reduces the amount of hydro power generation necessary to meet demand, which leads to higher reservoir levels and lower prices the next period. With a period length of one week, it is clear that the impact of the EEX price must be included without lags of one period. Thus, the EEX spot price is added as an explanatory variable to the linear regressions f_1 and f_2.

Electricity price time series are characterized by price jumps, which occurs due to for instance unexpected outtakes of power plants, cold waves or low water reservoirs (Blanco and Soronow, 2001). A reasonable argument is that the fundamental model [24] should reflect all factors influencing the price, and thus be able to model jumps. However, few price shocks may be present in the data, as a limited number of years are used to model the price equation[4]. Thus, it will be difficult to obtain the correct parameter estimates which fundamentally determine the size of a shock. Therefore, a shock dummy variable ϑ_t is introduced, which equals one when the conditions for a shock are present. These conditions have to be related to demand factors (i.e. cold waves) or low water reservoirs, since the possibility of unexpected plant outtakes is not reflected in our models. Such conditions are most likely to occur in the winter season, and therefore shocks are allowed to appear in this period only. The shock is assumed to increase the price by a factor e^{β_3}. Clearly, this representation is simple and a subject for future improvements.

The resulting price equation is:

$$S_{NP,t} = \delta_t f_1(I_t, R_{TOT,t}, Q_t, S_{EEX,t})\left(1 + e^{-\beta_1\left(R_{W\%,t} - C_{W\%,t}\right)}\right) + (1-\delta_t)(1+e^{\beta_3\vartheta_t}) f_2(I_t, R_{TOT,t}, Q_t\, S_{EEX,t})\left(1 + e^{-\beta_2\left(R_{W\%,t} - C_{W\%,t}\right)}\right) \quad [25]$$

[4] Denmark was included to the Nord Pool area as the last country in 2000, so market data from 2000 and forward will be used to model the price equation

where

- ϑ_t = 1 if conditions for a jump is satisfied, zero otherwise
- β_3 is the factor which determines the size of the jump

The price equation [25] constitute together with the equations for condense generation [14]-[16], net import [17], market equilibrium [9] and hydro reservoir balance [22] an equation system which has to be solved simultaneously for each time period *t*.

3. DATA

Time series data are collected from Denmark, Finland, Norway and Sweden, and then aggregated to representative data for the whole Nord Pool area. Table 3 shows the data that may be included in the model. Those are converted into weekly basis and divided into an estimation period and a forecast period. In order to obtain out-of-sample forecasts of two years, the forecast period starts in week 27, 2008.

TABLE 3. OVERVIEW OF COLLECTED DATA

Description	Period	Scaling	Source
Consumption modeling			
Consumption	2000-2010	Daily	Nord Pool Spot
Day length	-	Daily	Norwegian Water Resources and Energy Directorate (NVE)
Temperature, Norway	1993-2010	Daily	Norwegian Meteorological Institute
Temperature, Finland	1993-2010	Daily	Finnish Climate Centre
Temperature, Sweden	1993-2010	Daily	Swedish Meteorological and Hydrological Intitute (SMHI)
Temperature, Denmark	1993-2010	Daily	SMHI
Heating oil prices	2000-2010	Monthly	Statistics Norway
Retail trade index, Norway	2000-2010	Monthly	Statistics Norway
Retail trade index, Sweden	2000-2010	Monthly	Statistics Sweden
Retail trade index, Finland	2000-2010	Monthly	Statistics Finland
Retail trade index, Denmark	2000-2010	Monthly	Statistics Denmark
Generation modeling			
Electricity generation per technology, Finland	2000-2010	Weekly	Finnish Energy Industry
Electricity generation per technology, Sweden	2000-2010	Weekly	Svensk Energi
Electricity generation per technology, Denmark	2000-2010	Hourly	Energinet
Electricity generation, Norway	2000-2010	Weekly	Nord Pool Spot
Import and Export, Finland	2000-2010	Weekly	Finnish Energy Industry
Import and Export, Sweden	2000-2010	Weekly	Svensk Energi
Import and Export, Denmark	2000-2010	Hourly	Energinet
Import and Export, Norway	2000-2010	Weekly	Nord Pool Spot
Coal prices	2000-2010	Weekly	McCloskey
Spot prices			
Nord Pool	2000-2010	Weekly	Nord Pool Spot
EEX (Phelix)	2000-2010	Daily	Reuter EcoWin
Hydrology			
Water reservoir, Norway	1993-2010	Weekly	NVE
Water reservoir, Sweden	1993-2010	Weekly	Svensk Energi
Water reservoir, Finland	1993-2010	Weekly	Finnish Environment inst.
Inflow	1995-2010	Weekly	Nord Pool Spot
Snow reservoir, Norway	1993-2010	Daily	NVE
Median culmination, Norway	1971-2000	-	NVE
Median culmination, Sweden	1981-2005	-	SMHI

3.1 CONSUMPTION MODELING

Daily mean temperatures from each country over 17 years are collected. For Norway, the temperatures in Oslo, Bergen, Værnes and Tromsø are weighted according to population in the region they represent[5]. From Sweden, the temperature in Stockholm is used, the temperature in Harmaja, near Helsinki, is used from Finland, whereas the temperature in Malmö is used for Denmark. The temperature in Malmö reflects the temperature in Copenhagen, in lack of available Danish data. Temperatures will be used in modeling of consumption, and therefore temperatures from the most populated sites are chosen. A possible weakness is that all these locations reflect a coast climate.

Instead of using temperature directly, HDD, heating degree-days, is applied in the model. As result, the temperature is only affecting the model when it is lower than the critical temperature, which is selected as 16 degrees according to Johnsen and Willumsen (2010). Daily HDD values for each country are summed up to weekly figures. After creating HDD series with weekly granularity for each country, a new series is created by weighting the four series based on consumption in each country. This gives weekly accumulated weighted HDD, illustrated in Figure 3.

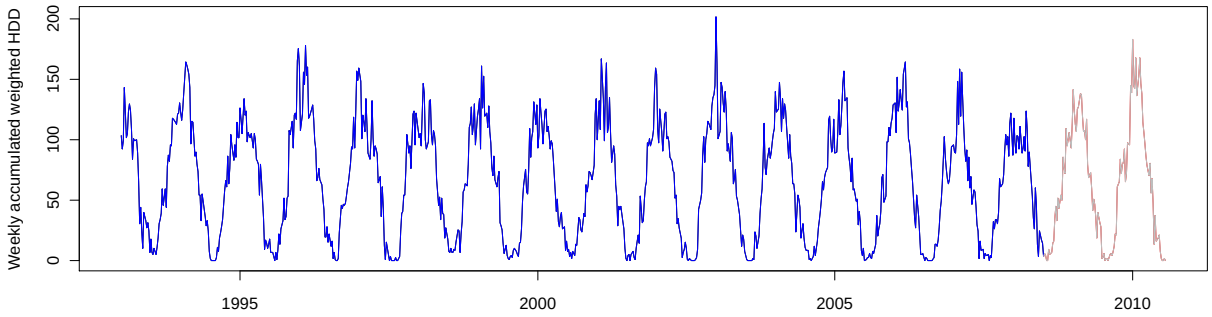

FIGURE 3. HDD TIME SERIES

Data for day length, time from sunrise to sunset, is collected for each day during the year. Retail trade indexes for each country are weighted due to consumption, and a 12 month moving average is used to obtain an indicator for the trend in consumer expenditure. Heating oil can be used as a substitute for electricity, and the price series is illustrated in Figure 4. There is a rapid growth in the end of the sample period, and it may be complicated to model further progression. Representative wind speed data for the whole area is hard to obtain, and will not be applied in the model.

[5] Blindern represents Oslo, Akershus, Hedmark, Oppland, Buskerud, Østfold, Vestfold, Telemark and Aust-Agder. Bergen represents Vest-Agder, Rogaland, Hordaland, Sogn og Fjordane and Møre og Romsdal. Værnes represents Sør-Trøndelag and Nord-Trøndelag. Tromsø represents Nordland, Troms and Finnmark

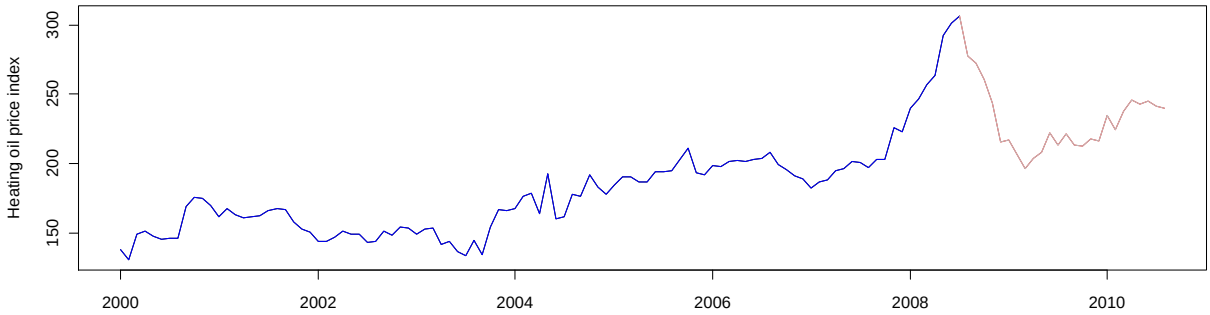

FIGURE 4. HEATING OIL PRICE TIME SERIES

A consumption model based on an appropriate selection of these data will be compared with the real consumption series in Figure 5. The consumption has a seasonality corresponding to that of the HDD, suggesting that HDD is the most important explanatory variable. Note that the consumption figures include TSO consumption, i.e. system losses.

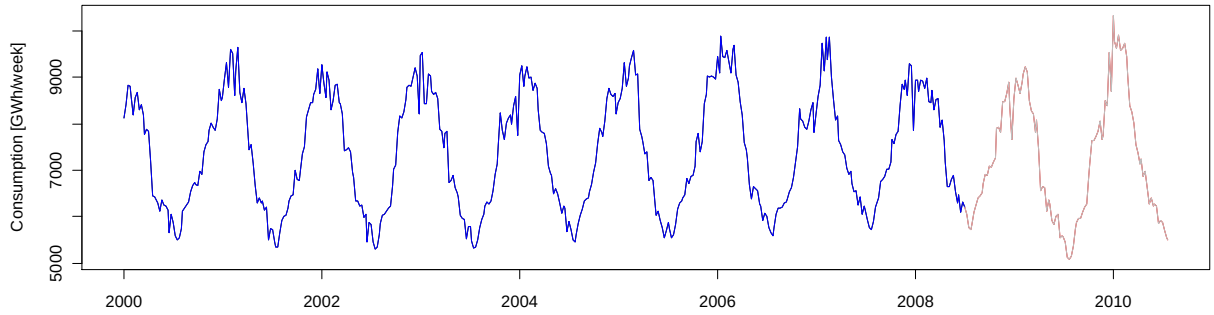

FIGURE 5. CONSUMPTION TIME SERIES

3.2 GENERATION MODELING

Figure 6 illustrates historic electricity generation from different technologies. Hydro, CHP district and baseload generation vary in a seasonal pattern with peaks in the winter. The trend for condense generation is somewhat different, whereas wind generation is completely random. Wind power generation is increasing during the period, but contributes to a small fraction of the total generation. The baseload curve has an uneven shape, as result of outages of plants due to maintenance or unforeseen problems. Data for the division of hydro power generation in regulated and unregulated generation is not found.

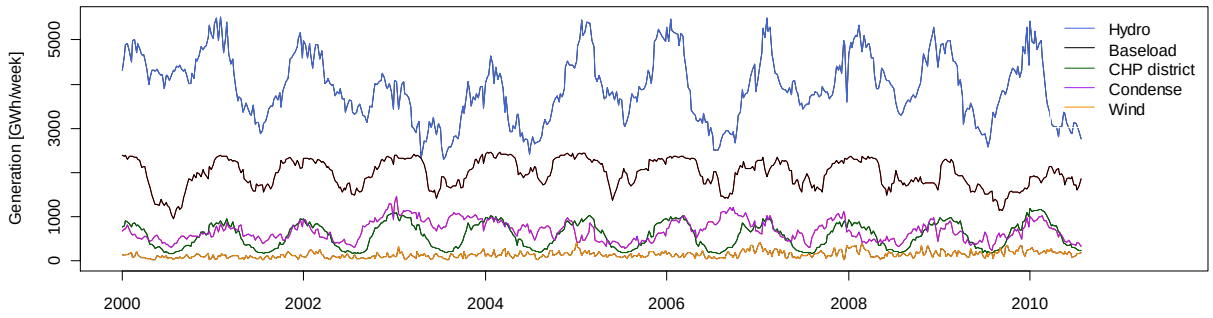

FIGURE 6. GENERATION PER TECHNOLOGY

In Norway almost all electricity is generated by hydro plants (Nordel, 2009), and thus the total Norwegian generation is set equal to the hydro generation. Generation data for hydro, condense, CHP district, CHP industry, nuclear power and wind are collected for Sweden and Finland. Here, baseload generation is calculated as the sum of CHP industry and nuclear power generation. Historical generation in Denmark is divided into central generation, decentral generation and wind generation (Energinet.dk, 2010). According to Rasch (2010), decentral generation stem mostly from municipal CHP units. Table 1 demonstrates that CHP industry generation is small, there is no nuclear power and hydro power is negligible. Therefore, condense generation is set equal to the central generation.

As coal and natural gas is used as fuel in condense units, these prices may affect the generation levels. Figure 7 shows a rapid growth in the coal price prior to the end of the sample period, complicating modeling the further price path. The natural gas price evolves in a similar manner.

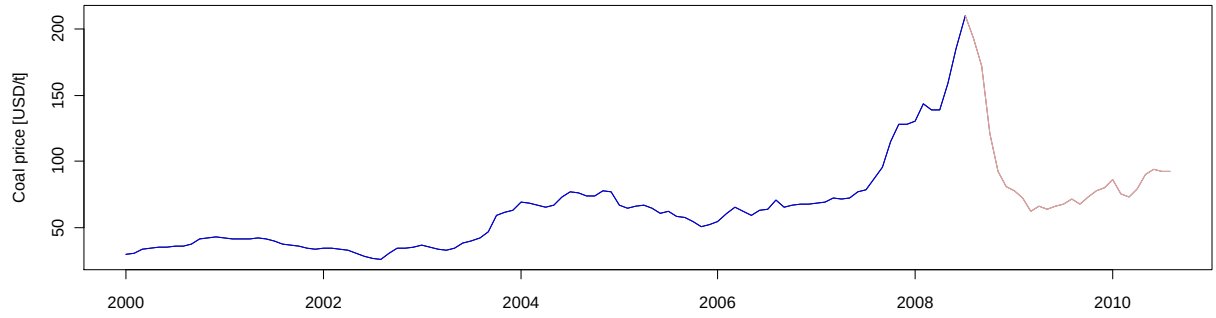

FIGURE 7. COAL PRICE TIME SERIES

The net electricity import to the Nord Pool area is illustrated in Figure 8, which clearly indicates that the Nordic countries together are a net importer of electricity.

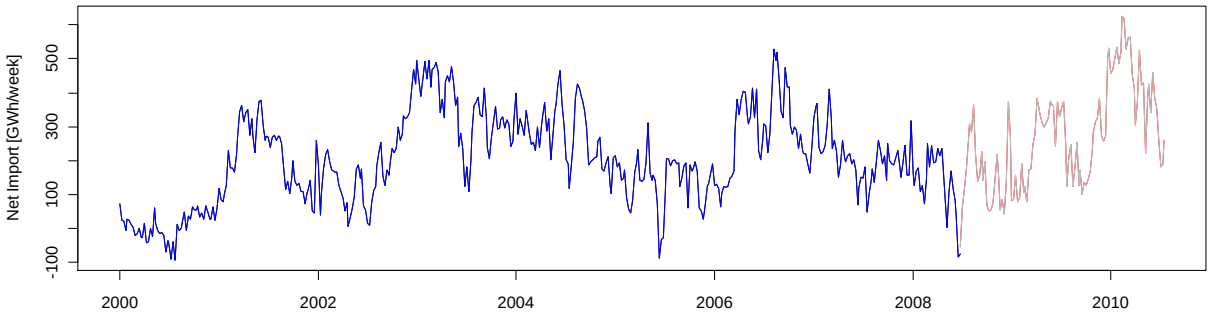

FIGURE 8. NET IMPORT TIME SERIES

3.3 SPOT PRICES

The German spot electricity price is the Phelix price at the European Electricity Exchange (EEX), here referred to as the EEX price. This price and the Nord Pool system price are plotted in Figure 9. The positive correlation between the two price series indicates that inclusion of the EEX price may improve a model for the Nord Pool price. The price forecast will be compared with the actual price displayed as the light blue line to the right in the figure.

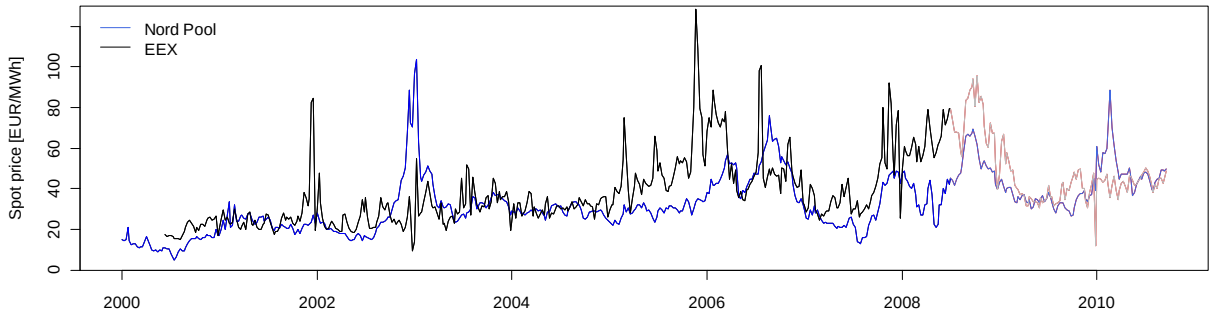

FIGURE 9. NORD POOL AND EEX SPOT PRICE SERIES

3.4 HYDROLOGY

The total weekly inflow into the hydro reservoirs in Norway, Sweden and Finland is plotted in Figure 10. There is clearly a seasonal behavior, with peak levels late in the spring and low levels during the winter.

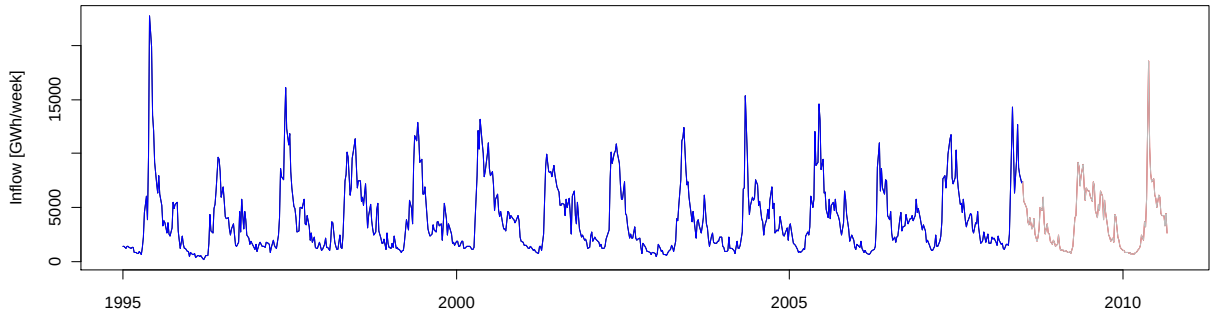

FIGURE 10. INFLOW TIME SERIES

In Figure 11, the energy content of the Norwegian snow reservoirs which drain to hydro power reservoirs is graphed as a percentage of the median culmination. The median culmination is the maximum value of the time series of median snow reservoirs, based on data from 1971 to 2000. Precipitation is generally overestimated during the year, and NVE has therefore set the energy content to zero after each season to avoid escalating levels in the data (Holmquist, 2010). Therefore, a rapid decrease in reported reservoir levels may appear at this time. In order to get the energy content in TWh, the values are multiplied with the median culmination. The median culmination in Norway is probably between 40 and 70 TWh (Holmquist, 2010), and not easy to decide due to lack of measurements. Consequently, this is a major source of error. Similar to what is done in Johnsen and Willumsen (2010), 60 TWh will be used in the model.

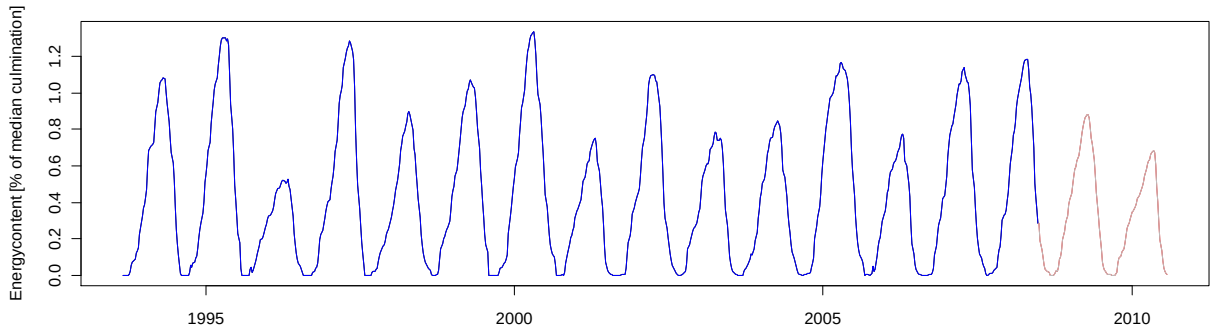

FIGURE 11. SNOW RESERVOIR TIME SERIES

Calculations by SMHI indicate that the median culmination is around 24 TWh in Sweden (Johansson, 2010). Historical time series for the energy content of snow reservoirs in Sweden was not available, and the Norwegian time series will therefore be assumed to be representative for Sweden.

There is also lack of time series for snow reservoirs in Finland. Due to different topography and lower levels of hydropower generation, neglecting the Finnish snow reservoirs should not be a significant source of error. Hence, the total snow reservoir is calculated by multiplying the Norwegian percentage of median culmination with 84 TWh.

The water reservoir levels measured at weekly basis for the three hydropower nations are graphed in Figure 12. Logically, the seasonality is exact opposite to that of the snow reservoirs. Data for hydro spillage is not found.

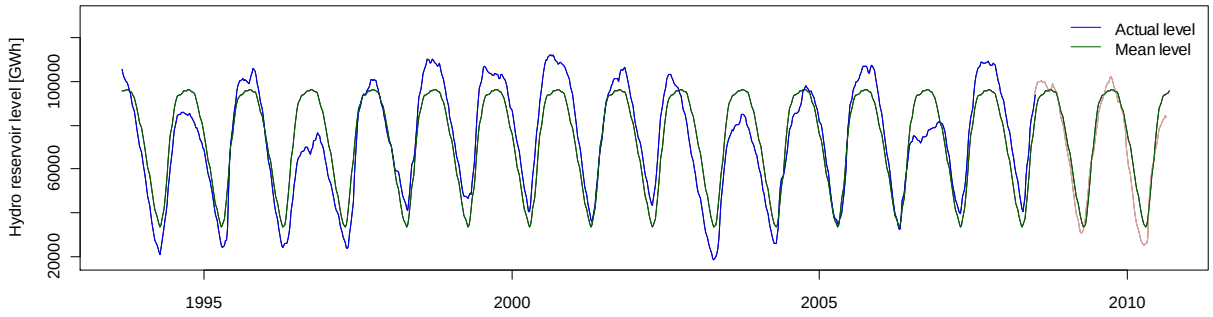

FIGURE 12. HISTORIC HYDRO RESERVOIR LEVELS

4. ESTIMATION

Statistical software are applied to find optimal coefficient estimates for the models. Models for risk factors are estimated in R, whereas EViews is used for other estimations. Simulations of risk factors and calculation of scenario paths for the system price, condense generation, hydro generation and hydro reservoirs are performed in MATLAB. Source code is attached in Appendix C. As the total number of scenarios grows exponentially with the number of simulations for each risk factor, the number of simulations for each factor is kept relative low (Table 4). The highest number of simulations is applied to inflow, which is assumed to be the most influential risk factor. In total, 18000 scenarios for the system price are generated.

TABLE 4. NUMBER OF SIMULATIONS FOR EACH RISK FACTOR

Number of Simulations	
Inflow	30
Snow	5
HDD	5
Wind	3
EEX	8

The statistical models are evaluated and selected based on tests for stationarity, parameter significance, residual autocorrelation and residual distribution. For the simple fundamental models, it is likely that not all relevant explanatory variables are included in the model. Thus, residuals will typically exhibit autocorrelation. The parameter estimates are still unbiased when residuals are autocorrelated, but standard error estimates are generally not correct (Brooks, 2008). Therefore, t-ratios cannot be used to evaluate parameter significance. The practical consequences of this are limited; since the fundamental models are selected according to the underlying theory all included parameters should be significant. Improving the statistical features of the fundamental models by extending them to e.g. SARIMAX models may be a extension of our work.

4.1 INFLOW, SNOW RESERVOIRS AND HDD

Stationary is tested using the Augmented Dickey-Fuller (ADF) test with 52 lags. The lag length is chosen equal to the period, according to Brooks (2008, p. 329). The null hypothesis, that the series is non-stationary, is rejected if the p-value is less than 10%. The test gives a p-value of 46% for inflow, which indicates non-stationarity. After non-seasonal differencing, the test indicates a stationary time series, and d is therefore set to one. For the snow series, the p-value is less than 1%, and the time series is clearly stationary. Hence, non-seasonal differencing is not necessary. For HDD, the p-value equals 11%, and the null hypothesis of non-stationary is not rejected. Since the test accept the null with small margin, models are compared for both $d = 0$ and $d = 1$. In the case of $d = 0$, the standard error estimates does not converge, and $d = 1$ is chosen.

As outlined in 2.1.1, SARIMA(p,d,q)(P,D,Q)$_S$ models with $D = 1$, $S = 52$ and $P + Q \leq 1$ are selected in order to capture the seasonality for inflow, snow and HDD. Models are estimated for all possible combinations of AR and MA lags, up to an upper limit of 9 lags, using maximum likelihood estimation. If there are autocorrelation in the residuals, the estimated model is rejected. From the remaining models, the

one with lowest AIC (Akaike's information criteria) is selected. The script for implementation of this procedure in R is attached in Appendix C.

The Ljung-Box test is used to detect autocorrelation in the residuals. The null hypothesis of no autocorrelation until lag *n* is rejected if the p-value is less than 0.10. The parameter *n* is set to 5% of the length of the series, as the test starts to deteriorate when number of lags exceeds this level (Burns, 2002). However, due to the seasonal pattern in the series, testing for at least 52 lags would be a reasonable choice. The drawback with this approach is that the number of AR and MA lags necessary to avoid failure of the Ljung-Box test will be boosted up.

4.1.1 INFLOW MODEL

A model with 3 AR-terms, 8 MA-terms and an SMA-term is selected in order to model the inflow. Due to the SMA term, the model can be interpreted as a seasonal exponential weighted moving average (EWMA) model. The weights applied to observations in previous seasons decline exponentially, and are given by $\theta_s^{k-1}(1+\theta_s)$ where *K* is the number of seasons back in time (Pankratz, 1991). The coefficients for these terms are given in Table 5.

TABLE 5. COEFFICIENTS FOR THE INFLOW MODEL

AR1	AR2	AR3	MA1	MA2	MA3
0.0156	-0.0174	-0.7096	-0.2790	-0.4341	0.6393
MA4	MA5	MA6	MA7	MA8	SMA1
-0.2390	-0.4405	-0.0465	-0.0177	-0.1219	-0.8333

Characteristics of the residuals are presented in Appendix A.1, including ACF plot, normal quantile plot and Ljung-Box p-values for different number of lags. The Ljung-Box test indicates no autocorrelation in the residuals for the selected lag length and significance level. A Shapiro-Wilk normality test yields a p-value approximately equal to zero, and the null hypothesis of normally distributed residuals is clearly rejected. Normally distributed residuals are not critical for obtaining good parameter estimates (Hipel et. al., 1977), but standard error estimates assume normally distributed residuals (Brooks, 2008). Hence t-ratios cannot be used to assess the significance of the parameters. Therefore, all estimated parameters are included the model.

The residual distribution will be used in simulation of inflow scenarios. The density plot (Figure 13) shows that the distribution has excess kurtosis and fat tails. Fat tails indicates that very high and low inflow levels are more likely to observe than in the case of normally distributed inflow residuals.

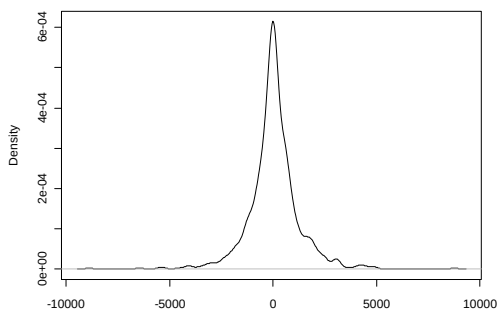

FIGURE 13. DENSITY OF INFLOW RESIDUALS

Since the residuals cannot be generated using a normally distributed random variable, we will use historical residuals to bootstrap an empirical distribution (McDonald, 2006). Residuals will be drawn by using the uniformly distributed pseudorandom numbers from a Sobol sequence as percentiles. A Sobol sequence is a low-discrepancy sequence of quasi-random numbers more evenly distributed than a similar series of random numbers (Glasserman, 2004); hence the variance will be reduced and the average of the simulations will converge more rapidly to the theoretical forecast. The relative high number of historical observations indicates that the bootstrapping procedure will perform well. Alternatively, a theoretical distribution could have been fitted to the residuals or a Box-Cox transformation could have been applied to the inflow series in search for normally distributed residuals (Hipel et. al., 1977).

Figure 14 illustrates how the actual inflow evolves compared to simulated scenarios and the forecasted level. The forecast tracks the realized inflow relatively well most of the time, except around week 20 in 2010 when there is a jump in the actual inflow. The mean of the simulations are near the forecasted level, indicating that a sufficient number of scenarios are simulated.

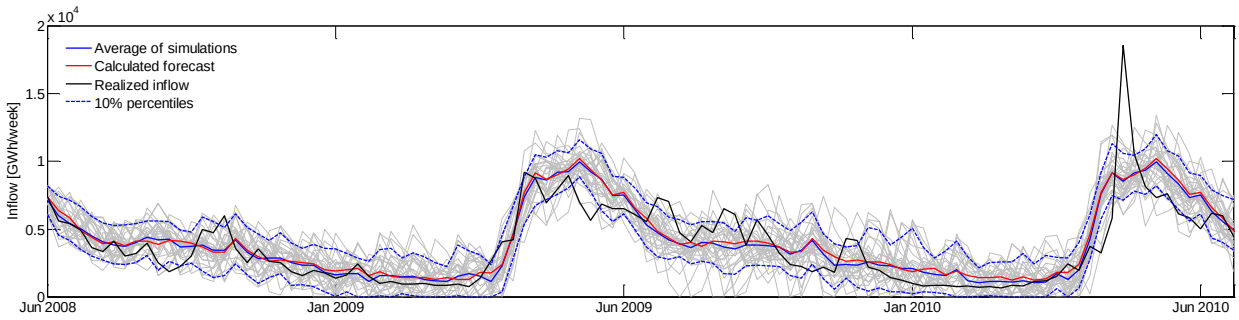

FIGURE 14. INFLOW FORECAST AND SCENARIOS

4.1.2 SNOW RESERVOIR MODEL

The selected model for the energy content of snow contains 7 AR-terms, 4 MA-terms and an SMA-term. The residuals are non-normally distributed; their distribution is unsymmetrical and leptokurtic. Appendix A.1 contains density plot and other residual characteristics. It is clear that the model pass the Ljung-Box test for the given specifications. Coefficient estimates are given in Table 6.

TABLE 6. COEFFICIENTS FOR THE SNOW RESERVOIR MODEL

AR1	AR2	AR3	AR4	AR5	AR6	AR7
-0.1603	0.1186	0.3115	0.3253	0.7920	-0.2738	-0.2802
MA1	MA2	MA3	MA4	MA5	MA6	SMA1
1.7016	1.7839	1.5719	1.1142	0.0435	-0.2477	-0.9177

Actual, forecasted and simulated levels are illustrated in Figure 15. Scenarios are simulated using bootstrapping of residuals. Due to the high variation in historical snow reservoirs (Figure 11) and the low number of simulations, percentiles are generated using random numbers instead of a Sobol sequence. The figure shows that the forecast fit reasonable well for the 2008-2009 season, while the low snow reservoirs in the next season cause the forecast to fail.

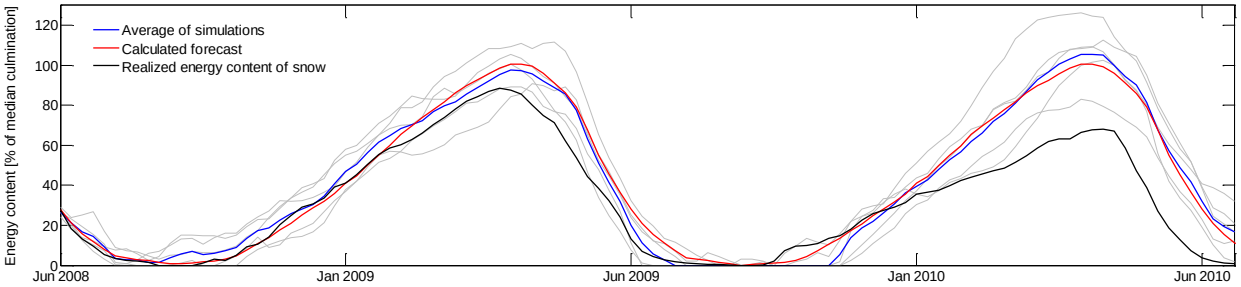

FIGURE 15. SNOW FORECAST AND SCENARIOS

4.1.3 HDD MODEL

For HDD, a model with 6 AR-terms, 6 MA-terms and an SMA-term is selected. Since the SMA coefficient equals -1, all past seasons are given equal weights in the forecast. The residual characteristics (Appendix A.1) indicate that the residuals are non-normally distributed. Ljung-Box p-values imply that the model pass requirements related to autocorrelation in residuals. Table 7 summarizes coefficient estimates.

TABLE 7. COEFFICIENTS FOR THE HDD MODEL

AR1	AR2	AR3	AR4	AR5	AR6	
0.5203	-0.0634	-0.299	-0.0866	0.6954	-0.1545	
MA1	MA2	MA3	MA4	MA5	MA6	SMA1
-1.0863	0.1355	0.3178	-0.0362	-0.9782	0.6473	-1

The actual HDD tends to vary around the forecast, except for the cold winter of 2010. In this period the actual level was even higher than the upper scenario. Scenarios are simulated using bootstrapping of residuals and a Sobol sequence.

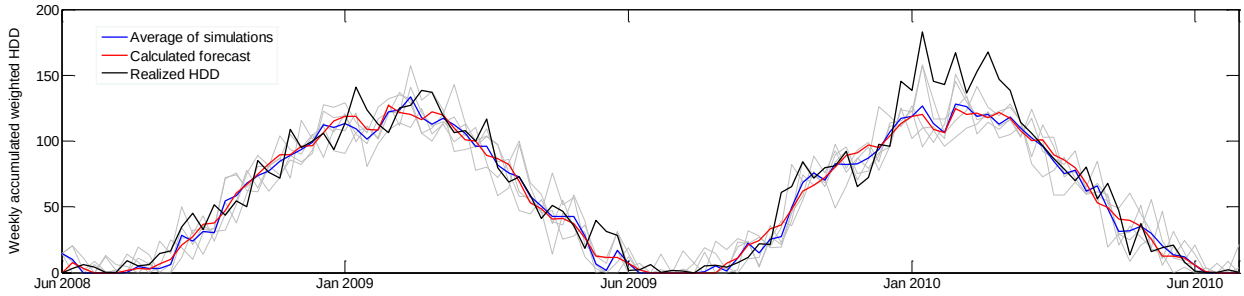

FIGURE 16. HDD FORECAST AND SCENARIOS

4.2 WIND POWER GENERATION

The ARIMA(p, d, q) process for wind power generation is estimated using maximum likelihood estimation. Models are compared for different numbers of AR and MA lags. From models without autocorrelation in the residuals, the alternative with lowest AIC is chosen.

Due to convergence problems, variance stabilization is applied to the time series; the logarithm of wind generation is modeled instead of absolute levels. The ADF test of the transformed series yields a p-value of 0.59 and *d* is set to one. A model with 5 AR-terms, 7 MA-terms and a constant term is selected. The constant term indicates an increasing trend in wind generation, reflecting the growth in installed capacity. Coefficients and standard errors are summarized in table 8.

TABLE 8. COEFFICIENTS FOR THE WIND POWER GENERATION MODEL

AR1	AR2	AR3	AR4	AR5		Constant
-0.2796	0.215	-0.5375	0.0226	0.7752		0.0016
MA1	MA2	MA3	MA4	MA5	MA6	MA7
-0.3477	-0.5824	0.6080	-0.3118	-0.9671	0.5070	0.0941

The residual characteristics (Appendix A.1) show non-normally distributed residuals, characterized by excess kurtosis and skewness. Results from Ljung-Box tests demonstrate that requirements concerning residual autocorrelation are satisfied.

Figure 17 shows simulated scenarios of wind generation during the forecast period. Both simulated and actual generations oscillate randomly. See the MATLAB script in Appendix C for implementation details.

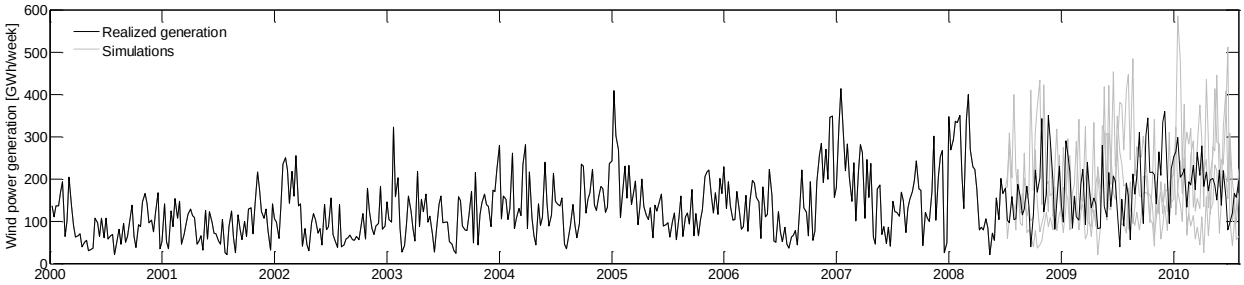

FIGURE 17. WIND POWER SIMULATIONS

4.3 EEX PRICE SIMULATIONS

To estimate the parameters of the AR-process [4], the following regression is specified:

$$ln(S_t) - ln(S_{t-1}) = a + b\, ln\,(S_{t-1}) + \epsilon_t \qquad [26]$$

The regression is estimated in EViews using data from 2004 to 2008. The parameters (Table 9) are calculated according to Dixit and Pindyck (1994).

TABLE 9. PARAMETERS OF MEAN-REVERTING PROCESS FOR THE EEX SPOT PRICE

Estimator	Value
$\bar{S} = -\frac{\hat{a}}{\hat{b}}$	3.80 ln(EUR/MWh)
$\eta = -\ln(1+\hat{b})$	16.67 %
$\hat{\sigma} = \hat{\sigma}_{\epsilon}\sqrt{\frac{\ln(1+\hat{b})}{(1+\hat{b})^2 - 1}}$	14.08 %

Figure 18 shows the resulting price paths created by running 8 simulations. The average of the simulations is close to the actual level, except for week 52 in 2009 and during the autumn of 2008. None of the scenarios manage to capture the high price period that autumn.

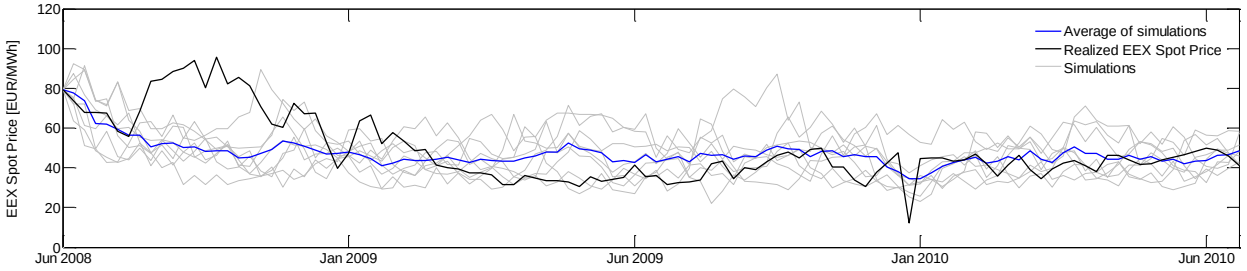

FIGURE 18. EEX SPOT PRICE SIMULATIONS

4.4 CONSUMPTION

4.4.1 DESCRIPTIVE STATISTICS

Table 10 presents the correlations between consumption and relevant explanatory variables introduced in 2.2.1. Heating oil prices will not be included in the model, because the rapid growth in the end of the estimation period makes future predictions difficult.

TABLE 10. CORRELATION BETWEEN CONSUMPTION AND EXPLANATORY VARIABLES

	Consumption
HDD	95.6 %
Day length	-83.0%
RTI	70.7 %

As apparent, the consumption is closely tied to the HDD. The correlation with day length is also high. However, HDD and daylength are correlated by a coefficient of -74%, which can cause problems with collinearity in the estimations. The retail trade index (RTI) is here compared with a 52-week moving average of consumption. The correlation of 70.7% demonstrates that the retail trade index is well suited to explain the growth pattern in consumption over time.

The scatter plot in Figure 19 adds further insight to the relation between consumption and HDD. The Nordic electricity consumption seems to be around 6000 GWh per week when no heating is required, and increases thereafter linearly with the HDD.

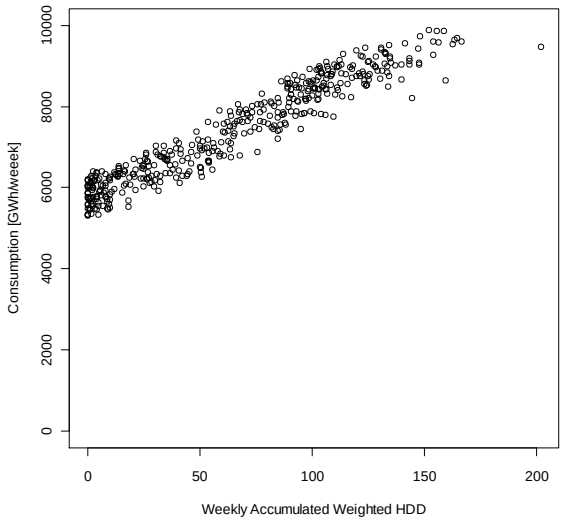

FIGURE 19. CORRELATION BETWEEN CONSUMPTION AND HDD

4.4.2 ERROR CORRECTION MODEL

The Augmented Dickey-Fuller test yields a p-value of 31%, hence the null hypothesis of a unit root cannot be rejected. While Johnsen and Willumsen (2010) found that the Norwegian consumption series in their study was stationary, we conclude that the Nordic consumption series is non-stationary. Consequently, an error correction model should be appropriate.

The following cointegration relationship is modeled:

$$Q_t = C + \beta_1 HDD_t + \beta_2 DL_t + \beta_3 ln(RTI_t) + \beta_4 D_{1,t} + \beta_5 D_{2,t} + u_t \quad [27]$$

where

- RTI_t is the 12-month moving average retail trade index in period *t*
- $D_{i,t}$, $i \in \{1, 2\}$ are holiday dummies for the summer holiday (3 weeks) and the Christmas holiday, respectively
- u_t is the time series of residuals

After estimating [27], the Engle-Granger test is applied to confirm that the estimated equation is cointegrated, i.e. that the residual series is stationary. Using a lag length of 31 based on the AIC criterion, the Engle-Granger test gives a p-value of 48%. Therefore, the null hypothesis of no cointegration cannot be rejected. If instead the SBIC criterion is used to select the lag length, a length of 3 is suggested, yielding an Engle-Granger p-value close to zero. This ambiguous result continues when applying the *ln*-transformation to consumption and day length.

Assuming that the variables are cointegrated, we estimate the whole error correction model [5]. The residuals show clear signs of autocorrelation, even when multiple lags of the difference terms in [5] are applied. As a consequence of the autocorrelation problems and doubt whether the variables are cointegrated, we conclude that a dynamic regression model may be a better choice for our data.

4.4.3 DYNAMIC REGRESSION MODEL

First, a regression similar to the cointegration expression in [27] is estimated by OLS:

$$Q_t = C + \beta_1 HDD_t + \beta_2 DL_t + \beta_3 ln(RTI_t) + \beta_4 D_{1,t} + \beta_5 D_{2,t} + u_t \quad [28]$$

A R^2 of 97.3 % indicates a good fit. A model with only HDD as input yields a fit of 92.1 %, illustrating that HDD is the by far most important variable to describe consumption.

The Augmented Dickey-Fuller test, with 52 lags, indicates that the residuals are non-stationary. Due to the seasonal pattern, both seasonal and non-seasonal differencing is applied to the model (Pankratz, 1991). Not surprisingly, the differentiated day length and HDD series are collinear, and day length is excluded as an input to the model. Furthermore, the retail trade index and both holiday dummies are not significant in the new model. Therefore, the model has boiled down to a single-input model with HDD as the only explanatory variable. With R^2 at 75 % for the differentiated model, the fit is still good[6].

A SARIMA model is fitted for the residual series using the procedure described in chapter 4.1, which yields a $(6,1,8)(0,1,1)_{52}$ model. The final model is therefore:

$$\nabla_{52}^{D} \nabla^{d} Q_t = \beta \nabla_{52}^{D} \nabla^{d} HDD_t + \frac{(1 - \theta_1 B^1 - \cdots - \theta_8 B^8)(1 - \theta_{52} B^{52})}{1 - \emptyset_1 B^1 - \cdots - \emptyset_6 B^6} a_t \quad [29]$$

The estimation results for the final model are summarized in Table 11. Appendix A.1 contains the residual characteristic.

TABLE 11. COEFFICIENTS FOR THE CONSUMPTION MODEL

AR1	AR2	AR3	AR4	AR5	AR6		β	
-1.7676	-1.3180	0.0099	1.1595	1.1227	0.2294		15.5028	
MA1	MA2	MA3	MA4	MA5	MA6	MA7	MA8	SMA1
1.3333	0.2569	-1.0487	-1.4544	-0.6339	0.4766	0.2330	-0.0335	-0.9963

Figure 20 shows the forecasted consumption. Note that forecasted, not actual, HDD is used as input. The forecast fit the realized consumption well, but the consumption is generally somewhat overestimated. The consumption is underestimated in the period from December 2009 to February 2010. This is due to the high HDD in this period, which is not captured by the HDD forecast.

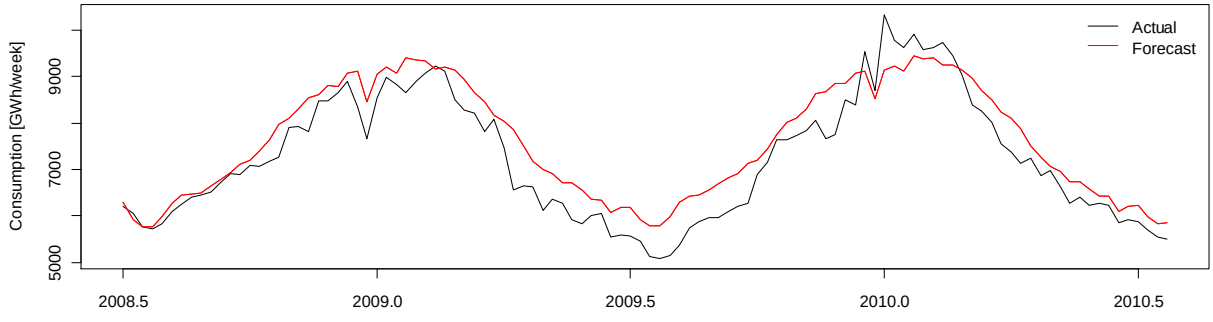

FIGURE 20. CONSUMPTION FORECAST

[6] Remark that R^2 for a differentiated and an undifferentiated model are not comparable

4.5 THERMAL GENERATION

4.5.1 BASELOAD

Baseload generation is forecasted using [10], with data from 2001 to 2008. The deviation from the actual generation is illustrated in Figure 21. The actual generation is considerably lower than forecasted in the end of 2008 and during the last year of the forecast period. The low generation in the end of 2008 was caused by outages of the Swedish nuclear power plants Oskarshamn 3, Forsmark 3 and Ringhals 1 (NVE, 2008[1]). During the autumn of 2009 and the winter of 2010, the generation was affected by frequent outages and delayed restart after maintenance of nuclear plants in Sweden (Johnsen, 2010). The impact of outages indicates that baseload generation should be treated as a risk factor, for instance through a jump process.

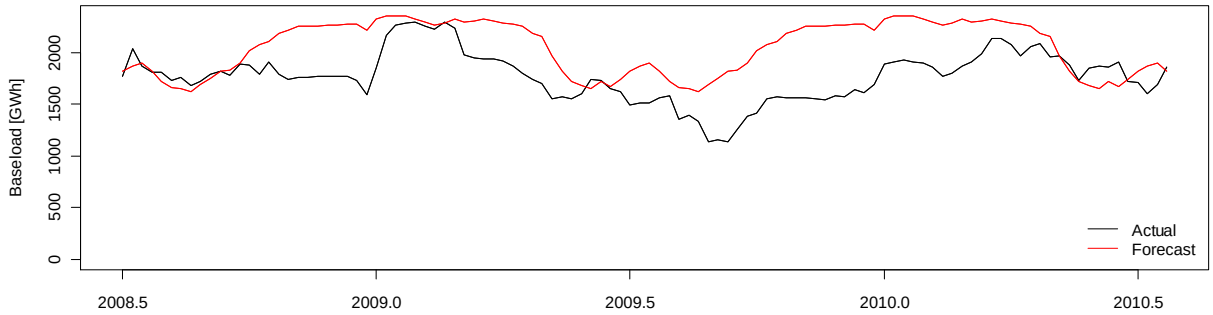

FIGURE 21. BASELOAD GENERATION FORECAST

4.5.2 CHP DISTRICT

The models [12] and [13] are estimated using ordinary least squares (OLS). Figure 22 illustrates the strong correlation between weekly CHP district generation (G_{CHPD}) and weekly HDD, and hereby back up the functional form of these models. Comparison of the estimated models shows that inclusion of maximum generation capacity does not improve the fit. Therefore model [13] is selected (R^2= 96%), which is estimated to:

$$G_{CHPD,t} = 189.3 + 5.84HDD_t \left[\frac{GWh}{week}\right] \qquad [30]$$

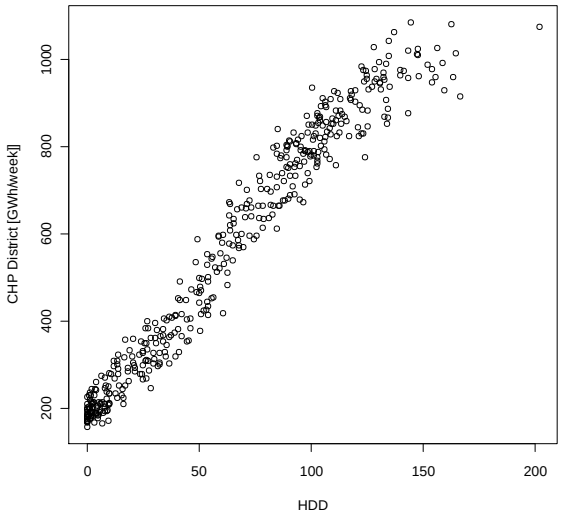

FIGURE 22. CHP DISTRICT GENERATION VERSUS HDD

Forecasted CHP district generation is calculated by inserting the HDD forecast into the above model. Figure 23 illustrates the forecasted generation compared to the actual. The forecast fits the actual values well the first year. The generation is underestimated during most of the 2009-2010 season, as result of a HDD forecast below the actual levels.

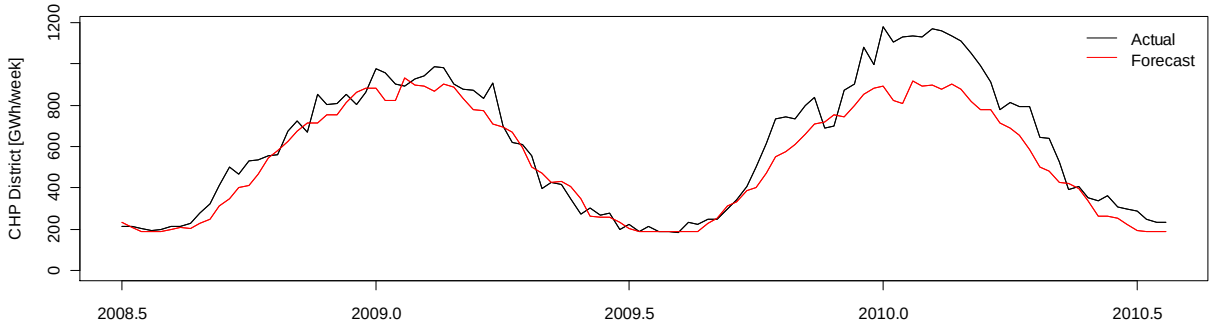

FIGURE 23. CHP DISTRICT GENERATION FORECAST

4.5.3 CONDENSE

Estimation of model [14] results in bad fit with historical data (R^2 = 45%), and statistical models [15,16] are considered. A positive and clear correlation between condense generation and the electricity price (67%) is illustrated in Figure 24. Figure 25 demonstrates somewhat weaker positive correlation between condense and electricity consumption (41%). The correlations between condense generation and the coal and gas prices are only -5.7% and -10%, respectively. In addition, modeling these prices is challenging as they seem to rise into the heaven in the end of the estimation period (Figure 7). Therefore, coal and gas prices are excluded as explanatory variables.

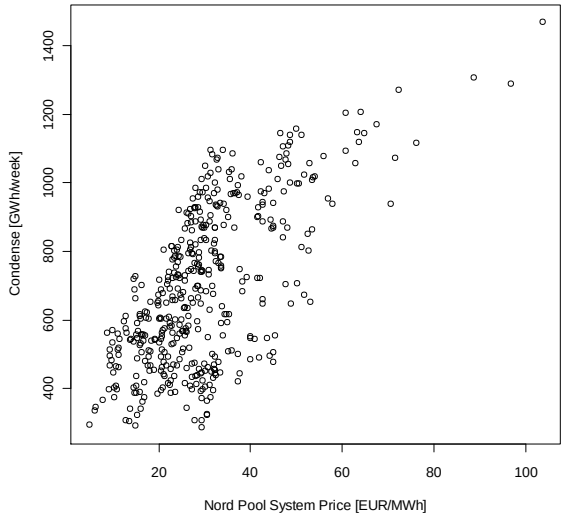

FIGURE 24. CONDENSE GENERATION VERSUS SPOT PRICE

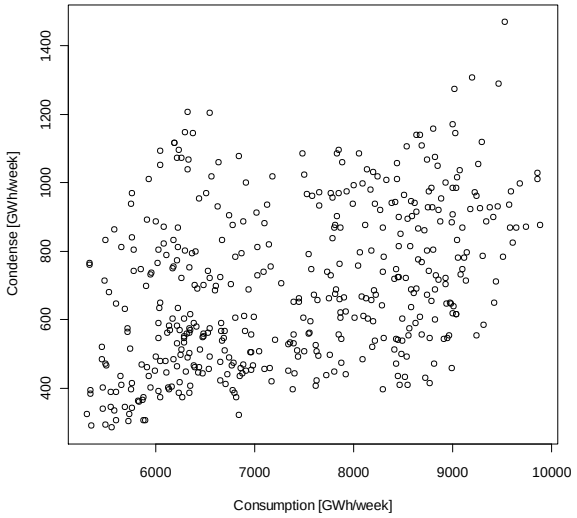

FIGURE 25. CONDENSE GENERATION VERSUS CONSUMPTION

The ADF test applied to the original condense series yields a p-value of 15.4%, indicating non-stationarity. After differentiation, the p-value becomes 1.7%, confirming that the differentiated series is stationary. OLS estimation gives the following model:

$$d\left(G_{condense,t}\right) = 7.88d\left(S_{NP,t}\right) + 0.11d(Q_t) \quad \left[\frac{GWh}{week}\right] \qquad [31]$$

The error correction term appeared to be insignificant, and is not included. The other coefficient estimates are significant at the 5% level (Appendix A.3), and their positive signs are consistent with the positive correlation illustrated above. There is no sign of autocorrelation in the residuals, as the Ljung-Box p-value

is 35.2% when testing 23 lags (5% of the time series length) and 11.6% when testing 52 lags. The fit is reasonable, with R^2 = 47%.

The out-of-sample performance of the model is illustrated in Figure 26, where condense generation is calculated using realized price and consumption. In other words, this graph does not show the forecast of condense generation, just the precision of the estimated model. The modeled values fit the actual generation quite well.

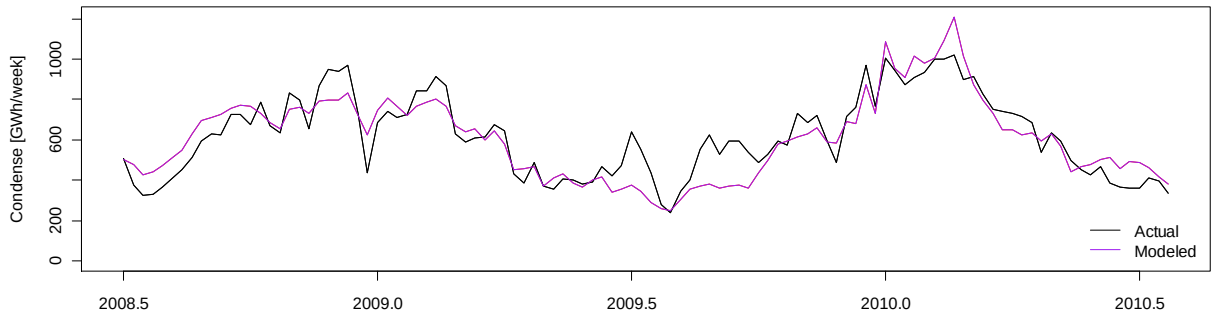

FIGURE 26. ACTUAL AND MODELED CONDENSE GENERATION

4.6 NET IMPORT

Net import is estimated as a function of the difference between the Nord Pool and EEX spot prices, according to [17]. Figure 8 shows that the net import has increased during the last 10 years. A shorter estimation period (from 2004) is therefore selected in order to avoid underestimation. OLS estimation yields:

$$NI_t = 283.5 + 5.44\left(S_{NP,t} - S_{EEX,t}\right) \quad \left[\frac{GWh}{week}\right] \qquad [32]$$

Given the limitations of the model discussed in Section 2.5, a fit of 52.8% can be considered as satisfactory. Figure 27 illustrates the out-of-sample performance, where the actual net import is compared with net import calculated using the realized price difference in the forecast period.

FIGURE 27. ACTUAL AND MODELED NET IMPORT

4.7 THE PRICE EQUATION

Figures 28 and 29 show plots of spot price against relative storage level[7] for the Nord Pool area and New Zealand. The plot for New Zealand, taken from Tipping (2007), displays why Tipping's marginal water value function performs that well: There is clearly an exponential relationship between the spot price and the relative storage level. For the Nord Pool area, the picture is more ambiguous: Albeit a strong negative correlation is present, it seems that the variation in spot price cannot be credibly explained by water reservoir levels alone. Thus, including more explanatory variables as in the proposed price equation [25] seems necessary.

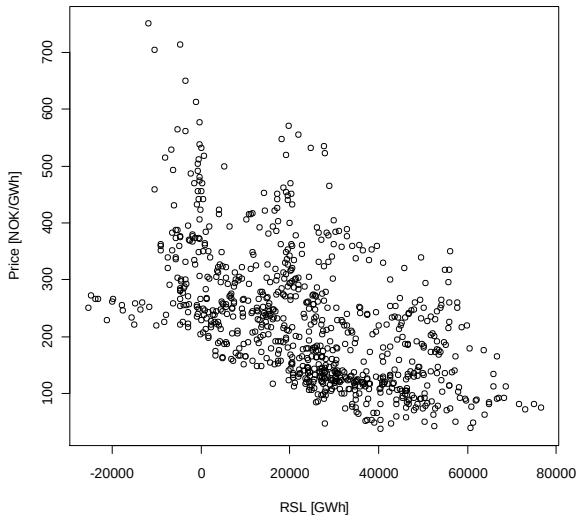

FIGURE 28. PRICE VERSUS RSL IN NORD POOL

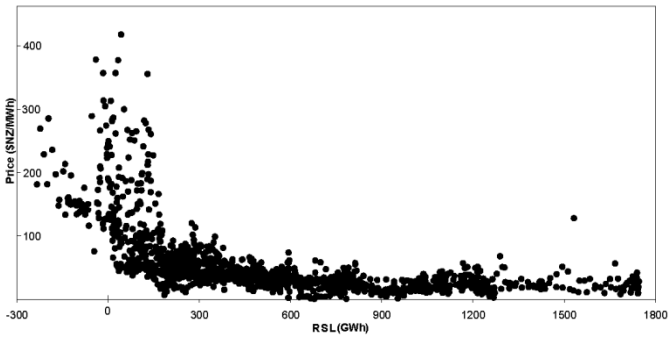

FIGURE 29. PRICE VERSUS RSL IN NEW ZEALAND

The form of the price equation [25] is selected by estimation of the different alternatives presented in 2.6 by maximum likelihood. Then, the model with highest log-likelihood is selected[8]. The equation turns out to be:

$$S_{NP,t} = \delta_t \left(C_{1,1} + C_{1,2}Q_t + C_{1,3}\left(R_{TOT,t} - R_{TOTMEAN,t}\right) + C_{1,4}S_{EEX,t}\right)\left(1 + e^{-\beta_1\left(R_{W\%,t} - R_{WMEAN\%,t}\right)}\right)$$

$$+ (1-\delta_t)(1+e^{\beta_3\vartheta_t})\left(C_{2,1}Q_t + C_{2,2}S_{EEX,t}\right)\left(1 + e^{-\beta_2\left(R_{W\%,t} - R_{WMEAN\%,t}\right)}\right) \qquad [33]$$

The coefficient estimates are given in Table 12. Inclusion of inflow does not improve the model in any season. Not surprisingly, inclusion of snow reservoirs (through the total reservoir deviation $R_{TOT,t} - R_{TOTMEAN,t}$) is improving the model only in the snow-melting period. All coefficient estimates have the expected sign. Note that β_2 is larger than β_1: The impact of a negative deviation from the mean hydro reservoir is larger in the winter season than in the summer season, as expected. Based on historical data, the snow melting season ("summer season") is set to start in week 17, while the snow accumulation season ("winter season") starts in week 35.

[7] While the relative storage levels for New Zealand are based on daily data and the 45-day moving average of the tenth percentile, the relative storage level for the Nord Pool area are based on weekly data and the 5-week moving average of the tenth percentile.

[8] Since the model is fundamental, residual autocorrelation and stationarity issues are not considered. Thus, p-values cannot be used to assess significance

TABLE 12. ESTIMATED PARAMETERS FOR THE PRICE EQUATION

$C_{1,1}$	-10.41
$C_{1,2}$	2.70E-3
$C_{1,3}$	-9.00E-5
$C_{1,4}$	0.192
$C_{2,1}$	4.87E-4
$C_{2,2}$	9.55E-2
β_1	2.89
β_2	4.83
β_3	0.56
R^2	*72%*
Log-likelihood	*-1409*

The conditions leading to a price shock is found to be:

$$\vartheta_t = 1, if\ R_{W\%,t} \leq R_{W\%P,t} and\ Q_t \geq Q_{mean,t}$$

$$\vartheta_t = 0, otherwise$$

where

- $R_{W\%P,t}$ is the historical tenth percentile of the hydro reservoir filling degree in period *t*
- $Q_{mean,t}$ is the historical mean consumption in period *t*

As stated earlier, the representation of shocks in the price equation is simple and should be improved. In addition, historical data before the estimation period may be investigated in order to gain deeper insight in the factors which cause price shocks.

Figure 30 shows the actual price along with the fitted price in the estimation period and the out-of-sample calculated price in the forecast period. The estimated price series tracks the realized price in an overall convincing way. Considering the forecast period (Figure 31), it is apparent that the periods of high prices, in August-September 2008 and the winter of 2010, are not properly captured by the model.

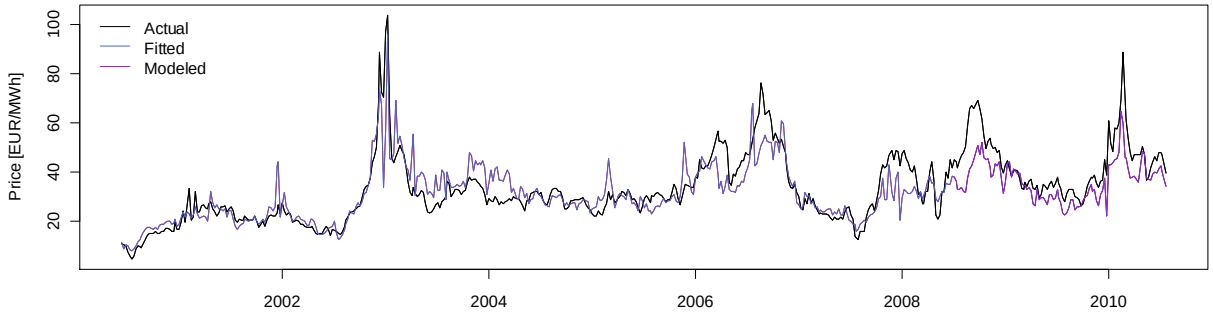

FIGURE 30. FITTED SYSTEM PRICES

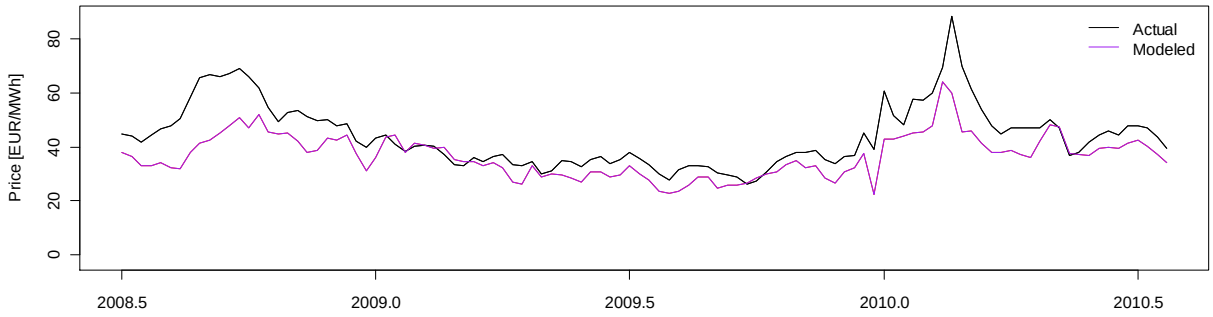

FIGURE 31. MODELED SYSTEM PRICE IN THE FORECAST PERIOD

Baseload supply was not lower than the historical average during the peak period in 2008, as indicated in Figure 21. Therefore, underestimation of the price cannot be explained by outtakes of nuclear plants. Coal and gas prices were high during these weeks, illustrated by the historical view of the coal price in Figure 7. The figure shows that the prices had been high for some time, and the increase in the electricity price cannot be explained by high fuel prices alone. Instead the price peak is likely to be a result of interaction between different factors: The EEX price increased rapidly in the period, with a 58% increase in the average weekly price from week 33 to week 35. In a period of increasing EEX price and high fuel prices, the hydrological situation was normalized. Until now, the filling ratio had been above its median level. Due to low inflow, the filling ratio decreased with 1.2 percentage points during week 34 and 35 (NVE, 2008[2]). A possible conclusion is that the model does not capture the increase in price due to a normalization of the hydrological situation in a period with high fuel prices and increasing EEX price.

The price peak in the winter of 2010 was caused by a combination of low temperatures and low hydro reservoir levels. A price peak is modeled in the period, but not of the correct size. This is not unexpected, due to the simple representation for price shocks.

The modeled price does not have the same starting point as the realized price (Figure 31). Thus, all relevant available information is not used to model the price. Including the price history in [33] will probably improve the model, but conflicts with the fundamental approach used to develop the price equation. An extension which maintains the fundamental nature of the model is inclusion of forward prices. Forward prices are the best predictors of future spot price variations (Vehviläinen and Pyykkönen, 2005), and could be included in the MWV function using the approach to Elverhøi et. al. (2010, formula 1).

4.8 PRICE SCENARIOS

The simultaneous calculation of equations [9], [22], [31], [32] and [33] for each time step in each scenario is performed in MATLAB (Appendix C). Concerning the reservoir balance [22], two simplifications have been introduced. First, hydro spillage is neglected. Separate time series data for spillage were not found, and calculation of implied historical spillage from equation [22] is not appropriate due to lack of precision in the data. Second, since data for the division of hydro power generation in regulated and unregulated generation were not found, Vehviläinen and Pyykkönen's empirical results are applied to [20]. Thus, the constant share of unregulated inflow is set to 20% (C_1 in [20]). The coefficient which determines the growth in unregulated inflow due to increasing reservoir levels (C_2) is set to -13.3.

Calculated scenarios for hydro reservoirs are shown in Figure 32, along with the realized reservoir series, the historical mean reservoir series and the 10% percentiles of the simulations. The expected reservoir (the average of the scenarios) is close to the realized reservoir in the first weeks, and approaches the historical average with time. This convergence is expected: All risk factors are mean-reverting in the long run, thus the expected hydro reservoir conditioned on information from week 26 in 2008 and backwards converges to the unconditional expectation with time. A complementary interpretation can be raised from the idea behind the water value function (Chapter 2.5): Since the optimal reservoir management policy is approximated to be the strategy which keeps the hydro reservoir at the long-term mean, the price formation should drive the reservoir level towards the mean. As apparent, the convergence is not perfect, due to inaccurate input data: The expected hydro reservoir at the end of the forecast period is slightly underestimated compared to the historical average.

The figure indicates that the number of simulations for each risk factor is not optimal. For example, the narrowing in the percentiles around January 2009 has no fundamental explanation. Increasing the total number of simulations or finding a better division of the number of simulations for each risk factor can solve this problem.

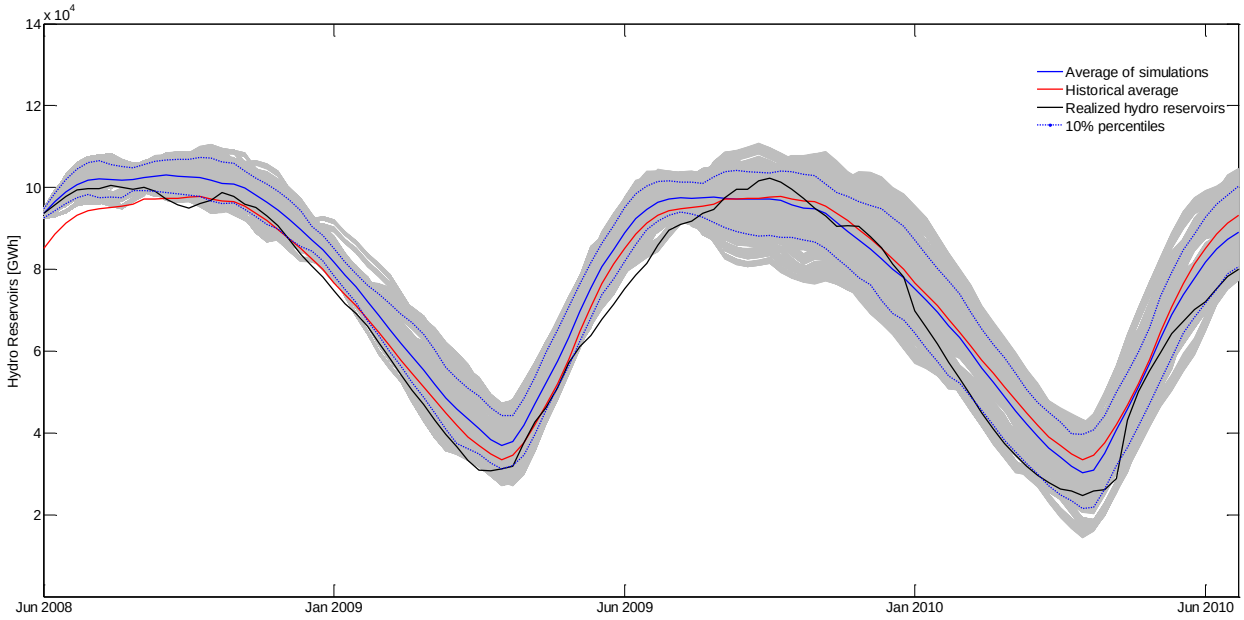

FIGURE 32. HYDRO RESERVOIR SCENARIOS

Figure 33 presents the scenarios for the system price. From the start of the forecast period and the next eight months, the forecasted price distribution does not match the realized price. The misfit has three main causes: The price equation does not manage to model the price peak in the autumn of 2008, due to the reasons discussed in Chapter 4.7. All paths for the EEX price are below the actual price in the same period. The hydrological conditions were favorable at the start of the forecast period, but the sudden normalization contributed to higher prices. Such normalization is not captured by the forecasts, where the expected hydro reservoir is higher than the historical average through the next year. From the end of February 2009, the modeled distribution performs better. The origin of high-price scenarios from August 2009 is tied to higher spread in the simulated reservoirs, where more scenarios are below the historic

average reservoir level. Scenarios containing price jumps appear in the 2009-2010 winter season, although not with a high probability. The likelihood for jumps depends on the conditions that are assumed to lead to price jumps, and should be interpreted with caution.

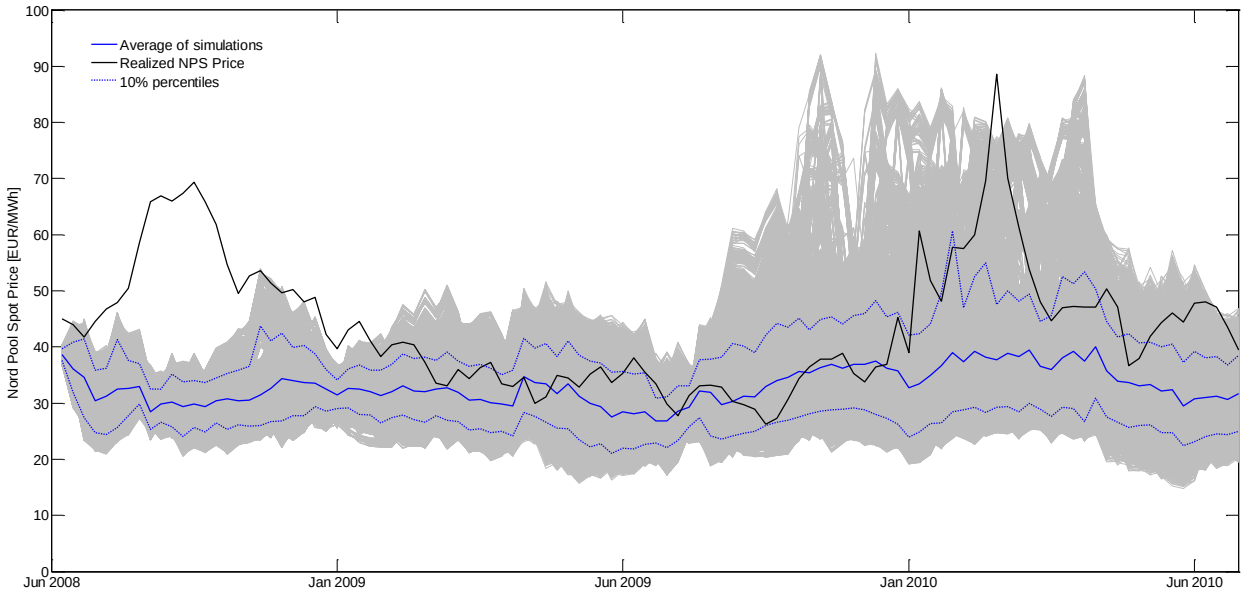

FIGURE 33. SYSTEM PRICE SCENARIOS

5. DISCUSSION

As the model considers the uncertainty in future spot prices, it is well suited for assessing risks related to price variation. It can be useful in planning with up to two years horizon, typically planning of generation and risk management in energy-intensive industry. The major strength of the model is inclusion of fundamental factors and relations between them. However, it is clear that not all interactions between the factors are captured. This was demonstrated during the high-price period in the autumn of 2008, when the hydrological situation was normalized in a period of high fuel prices and increasing EEX price. The episode indicates that inclusion of gas and coal prices can improve the model, despite the low correlation between these prices and thermal generation.

The model of this thesis can be improved in different ways. Simply increasing the number of simulations would improve the performance. Concerning data, longer time series can be collected for hydrological data as inflow and hydro reservoir levels. More reservoir data will lead to better estimates for the mean reservoir level. In addition, time series for hydro reservoirs spillage can be considered. Market data are collected during the period for which Nord Pool included the same areas as today. The model can possibly be improved by using data for periods when the market structure was different. In particular, time series for Norway could be applied to increase the knowledge about jumps in the spot price. Minor improvements can be achieved for snow and temperature data. The location of temperature measurements can be reconsidered. If more accurate estimates for median culmination of snow become available, this would improve the accuracy of the snow model. Also, time series for snow in Sweden and Finland would increase the performance of this model.

Each of the estimated models can be evaluated independently and the entire structure can be improved by component wise development of existing models. For the risk factors, implementation of a correlation structure should be considered. The distribution of the scenarios will be more accurate if the obvious correlation between inflow, snow reservoirs and HDD is implemented. In simulation of the EEX price, jump processes can be included. Also, the baseload generation can be treated as a risk factor, by including jump processes for outages. As discussed, inclusion of fuel prices in models for baseload and condense can be constructive. The net import model can be improved by differing between exchange to different markets, as the electricity spot price movements in Netherlands and Germany are unlike those in Poland, Estonia and Russia. Better representation of jumps in the system price equation is another improvement opportunity. The price equation does not include price history and forward prices, available information which may be beneficial to exploit. For the fundamental models in general, extension to statistical models (e.g. ARIMAX models) may improve forecast abilities.

REFERENCES

Batstone, S.R.J. (2003) *Aspects of risk management in deregulated electricity markets: Storage, market power and long-term contracts*. Ph.D. Thesis, Department of Management, University of Canterbury, New Zealand.

Bobinaitė, V., Juozapavičienė, A. and Snieška, V. (2006) Correlation of Electricity Prices in European Wholesale Power Markets. *Engineering Economics*, 4(49), 7-14.

Blanco, C. and Soronow, D. (2001) Mean Reverting Processes - Energy Price Processes Used For Derivatives Pricing & Risk Management. *Commodities Now* [online] June, Available from: http://www.fea.com/resources/pdf/a_mean_reverting_processes.pdf [Downloaded October 25, 2010].

Brooks, C. (2008) *Introductory Econometrics for Finance* (2. edition). New York: Cambridge University Press.

Burkardt, J. (2009) *Generate Sobol Datasets.*[online]. Available from: http://people.sc.fsu.edu/~jburkardt/m_src/sobol_dataset/sobol_dataset.html [Downloaded November 18, 2010].

Burns, P. (2002) *Robustness of the Ljung-Box Test and its Rank Equivalent*. Working paper [online] Available from: http://www.burns-stat.com/pages/Working/ljungbox.pdf [Downloaded September 29, 2010].

Dixit, A.K. and Pindyck, R.S. (1994) *Investment under Uncertainty* (1. edition). New Jersey: Princeton University Press.

Doorman, G.L. (2010) *Hydro Power Scheduling*. Compendium in the course ELK15, Department of Electric Power Engineering, Norwegian University of Science and Technology.

Elverhøi, M., Fleten, S-E., Fuss, S., Heggedal, A. M., Szolgayova, J. and Troland, O. C. (2010) *Evaluation of hydropower upgrade projects – a real potions approach*. [Online] MPRA Paper No. 23005. Available from: http://mpra.ub.uni-muenchen.de/23005/ [Downloaded December 6, 2010].

Energinet.dk (2010) *Udtræk af markedsdata*.[online]. Available from: http://www.energinet.dk/DA/El/Engrosmarked/Udtraek-af-markedsdata/Sider/default.aspx. [Downloaded September 27, 2010].

Glasserman, P. (2004) *Monte Carlo Methods in Financial Engineering*. New York: Springer

Holmqvist, E. (2010) Questions concerning snow data for Norway. [email]. (Personal communication, September 21, 2010).

Hipel, K.W., McLeod, A.I. and Lennox, W.C. (1977) Advances in Box-Jenkins Modeling. *Water Resources Research*, 13 (3), 567-575.

Johansson, B. (2010) Question regarding median culmination in Sweden. [email]. (Personal communication, September 29, 2010).

Johnsen, T.A. and Willumsen, M.Ø. (2010) *Electricity demand and temperature - an empirical methodology for temperature adjustment in Norway*. [online]. Attached the report for the Norwegian power market 4. quarter 2009. Available from: http://www.nve.no/Global/Publikasjoner/Publikasjoner%202010/Rapport%202010/rapport2010_01.pdf [Downloaded September 9, 2010].

Johnsen, T. A. (2001) Demand, generation and price in the Norwegian market for electric power. *Energy Economics*, 23(3), 227-251.

Johnsen, T. A. (2010) *Kvartalsrapport for kraftmarkedet 2. kvartal 2010*. [online]. Available from: http://www.nve.no/Global/Energi/Analyser/Kvartalsrapporter/Rapport%2014-10.pdf?epslanguage=no [Downloaded November 10, 2010].

Krogh Kristoffersen, T. (2007) *Stochastic Programming with Application to Power Systems.* Ph. D. Thesis, Department of Operations Research, University of Aarhus, Denmark

McDonald, R. L. (2006) *Derivatives Markets* (2. edition). Boston: Pearson Education.

Murray, F.T. and Ringwood J.V. (1994) Improvement of electricity demand forecasts using temperature inputs. *Simulation Practice and Theory*, 2(3), 121-139.

Nau, R.F. (2010) *Seasonal ARIMA models.* [online]. Available from: http://www.duke.edu/~rnau/seasarim.htm. [Downloaded October 13, 2010].

Nordel (2009) *Annual Report 2008.* [online]. Available from: https://www.entsoe.eu/fileadmin/user_upload/_library/publications/nordic/annualreport/Annual%20report%202008.pdf. [Downloaded September 13, 2010].

Nord Pool Spot (2010) *System price.* [online]. Available from: http://www.nordpoolspot.com/trading/The_Elspot_market/Price-calculation/System_price/. [Downloaded December 7, 2010].

NVE (2008[1]) *Kraftsituasjonen pr. 10. desember.*[online]. Available from: http://www.nve.no/PageFiles/1326/Monitor-uke-49%202008.pdf. [Downloaded November 10, 2010].

NVE (2008[2]) *Kraftsituasjonen pr. 3. september*.[online]. Available from: http://www.nve.no/PageFiles/1326/Monitor-uke-35.pdf. [Downloaded December 5, 2010].

Pankratz, A. (1991) *Forecasting with Dynamic Regression Models* [1. edition] New York: John Wiley & Sons.

Rasch, C. N. (2010) Questions regarding generation per technology in Denmark. [email]. (Personal communication, October 8, 2010).

Shumway, R. H. and Stoffer, D. S. (2010) *Time Series Analysis and Its Applications.* [online]. Available from: http://www.stat.pitt.edu/stoffer/tsa2/index.html. [Downloaded September 13, 2010].

SKM Energy Consulting (2003) *Dutch-Norwegian Interconnector - Feasibility study on the socioeconomic benfits of a cable between Norway and the Netherlands.* [online]. Consultant report

submitted to Statnet and TenneT. Available from: http://www.tennet.org/images/03-bijlage_A-SKM-haalbaarheidsstudie_tcm41-12096.pdf. [Downloaded December 3, 2010].

Tipping, J.P. (2007) *The Analysis of Spot Price Stochasticity in Deregulated Wholesale Electricity Markets.* Ph. D. Thesis, Department of Management, University of Canterbury, New Zealand.

Vehviläinen and Pyykkönen (2005) Stochastic factor model for electricity spot price—the case of the Nordic market. *Energy Economics, 27(2),* 351-367.

Wangensteen, I. (2007) *Power System Economics – the Nordic Electricity Market* (1. edition). Trondheim: Tapir academic press.

APPENDIX A. FURTHER ESTIMATION DOCUMENTATION

A.1. RESIDUAL CHARACTERISTICS

Figures A.1 – A.8 display the characteristics of the residuals for the risk factors and consumption models. The degrees of freedom in the Ljung-Box test is set equal to the particular lag length less the number of estimated coefficients: *n-(p+q+P+Q)*, according to Shumway and Stoffer (2010). Normal quantile plots (with points deviating from a straight line), density plots and Shapiro-Wilk p-values close to zero confirm that all models have non-normally distributed residuals.

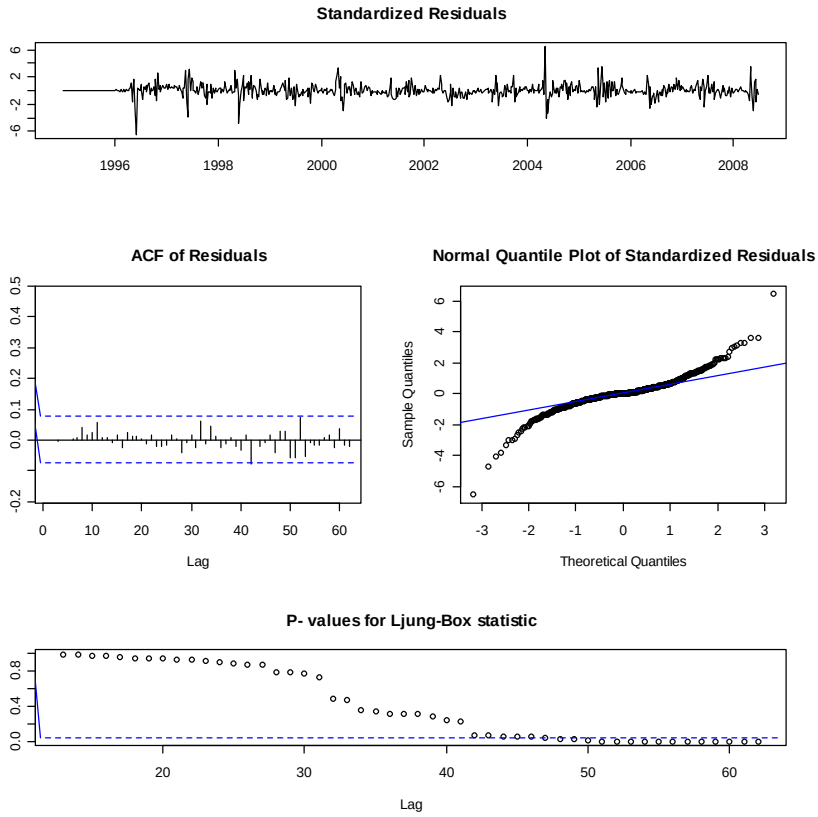

FIGURE A.1. CHARACTERISTICS OF INFLOW RESIDUALS

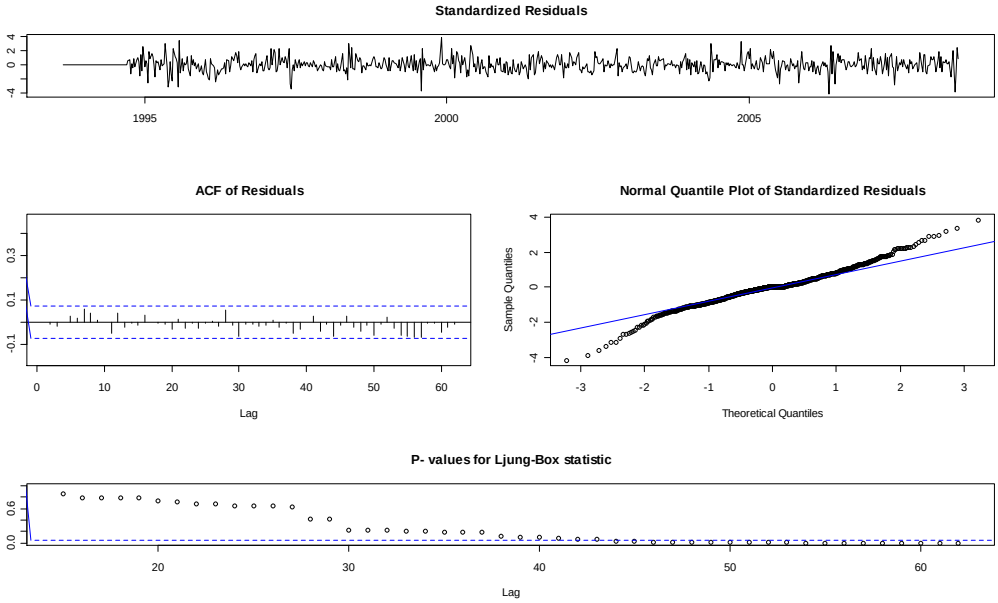

FIGURE A.2. CHARACTERISTICS OF SNOW RESERVOIR RESIDUALS

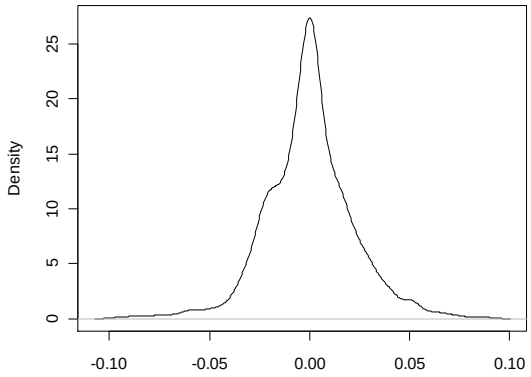

FIGURE A.3. DENSITY OF SNOW RESERVOIR RESIDUALS

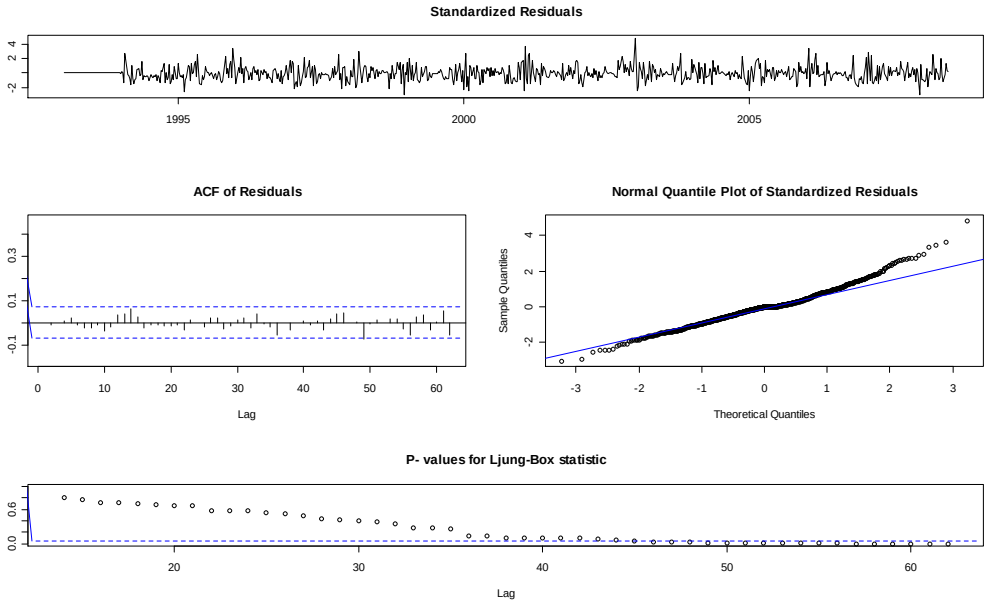

FIGURE A.4. CHARACTERISTICS OF HDD RESIDUALS

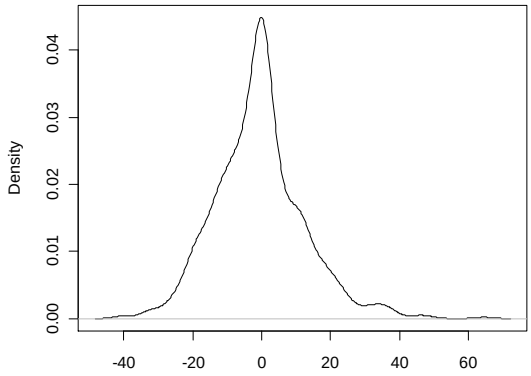

FIGURE A.5. DENSITY OF HDD RESIDUALS

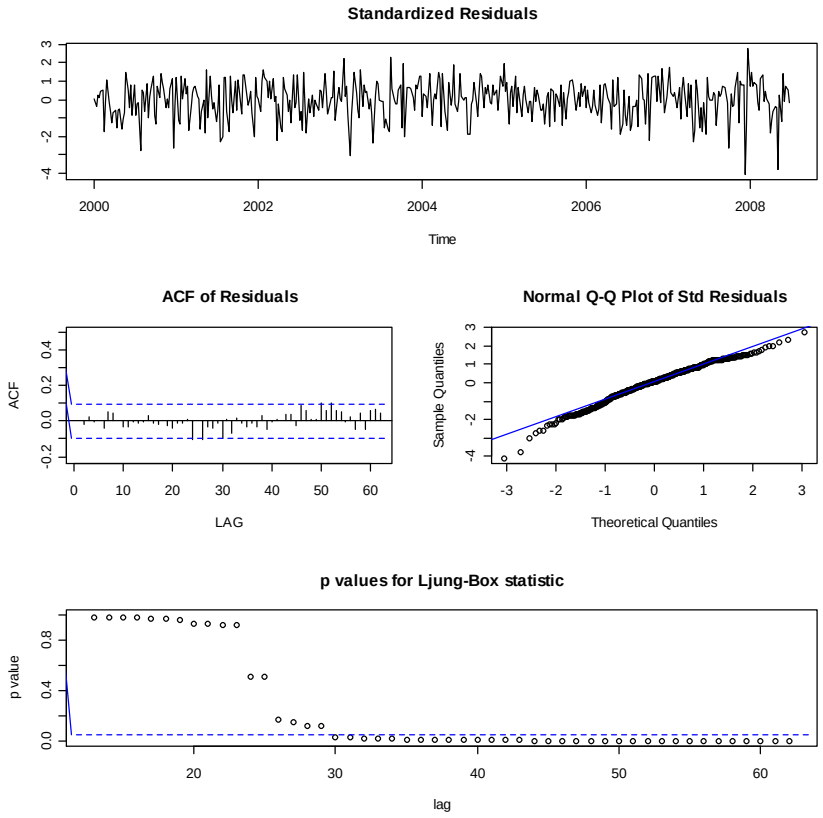

FIGURE A.6. CHARACTERISTICS OF WIND POWER GENERATION RESIDUALS

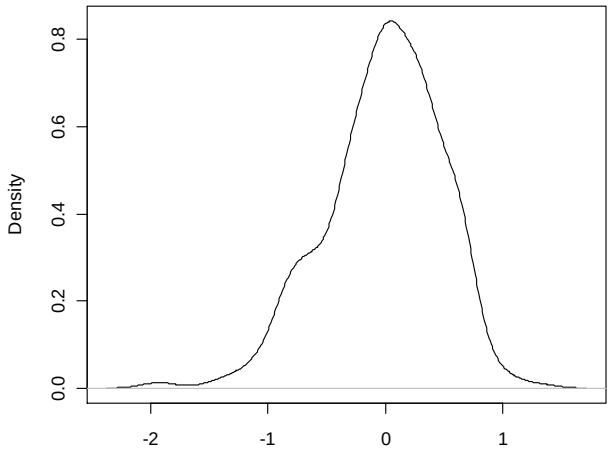

FIGURE A.7. DENSITY OF WIND POWER GENERATION RESIDUALS

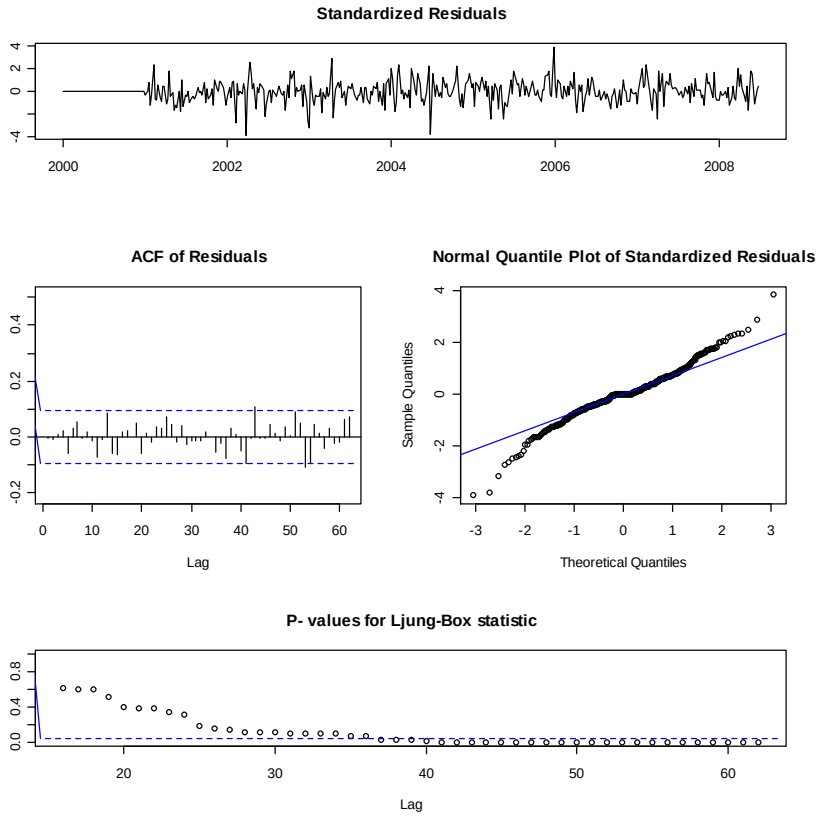

FIGURE A.8 CHARACTERISTICS OF CONSUMPTION RESIDUALS

A.2 FIT PLOT FOR THE CHP DISTRICT MODEL

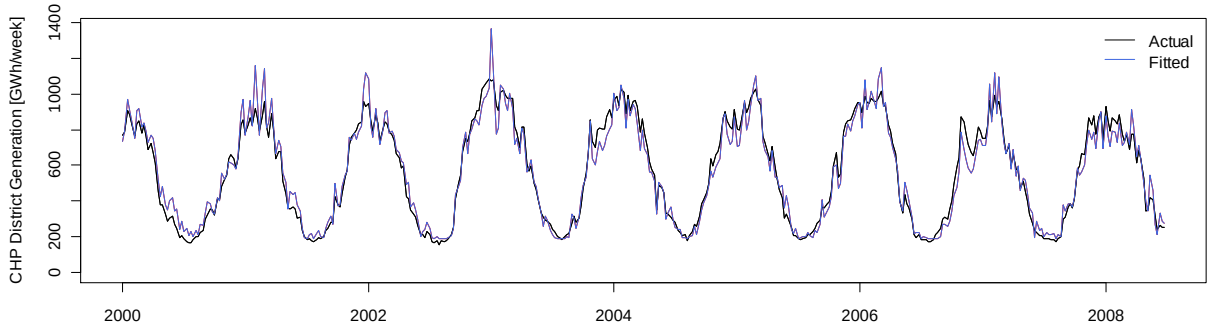

FIGURE A.9. ACTUAL AND FITTED CHP DISTRICT GENERATION

A.3 ESTIMATION RESULTS FOR THE CONDENSE MODEL

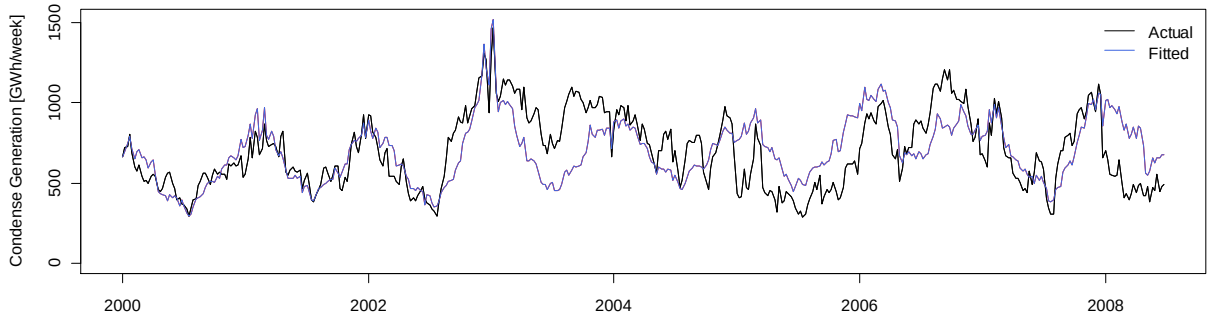

FIGURE A.10. ACTUAL AND FITTED CONDENSE GENERATION

TABLE A.1. ESTIMATION OUTPUT FOR THE CONDENSE MODEL

	C_1	C_2
Coefficient	7.88	0.112
Std. Error	0.843	0.010
t-Statistic	9.35	11.10
P-value	0.00	0.00
R^2		47.4%

A.4 FIT PLOT FOR THE NET IMPORT MODEL

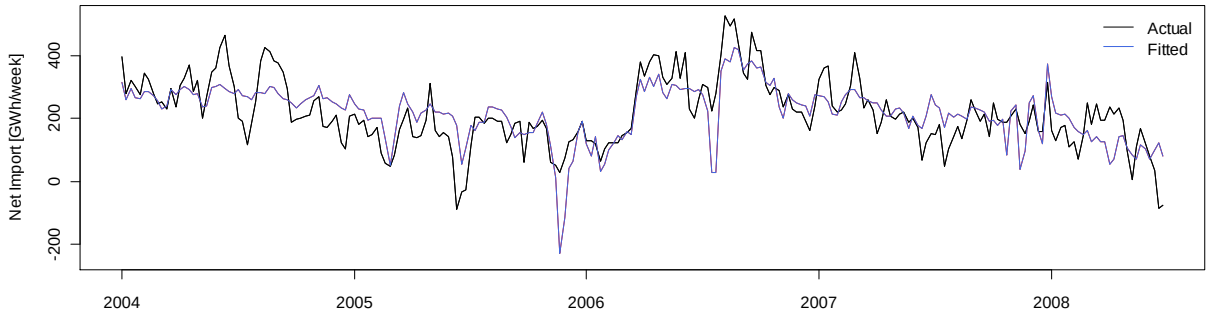

FIGURE A.11. ACTUAL AND FITTED NET IMPORT