CHALMERS | GÖTEBORG UNIVERSITY

MASTER'S THESIS

Modeling Electricity prices in the German market

Henrik Sangö

Department of Mathematical Statistics
CHALMERS UNIVERSITY OF TECHNOLOGY
GÖTEBORG UNIVERSITY
Göteborg, Sweden 2008

Thesis for the Degree of Master of Science (20 credits)

Modeling Electricity prices in the German market

Henrik Sangö

CHALMERS | GÖTEBORG UNIVERSITY

Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE – 412 96 Göteborg, Sweden
Göteborg, August 2008

Modeling Electricity prices in the German market

Henrik Sangö

August 2008

Abstract

In this thesis we model electricity spot prices in the German electricity market. The models being investigated incorporates mean-reversion, jumps, seasonality and GARCH behavior. Several different models were estimated to compare the relative importance of the factors. The work with this thesis was done during my stay at the Technical University of Munich.

Acknowledgements

I would like to thank Thorsten Schmidt at the Technical University of Munich for the suggestion of the problem and guiding to solving it. Thanks also to my supervisor in Gothenburg, Patrik Albin for his support. Special thanks to the organizations that supported me financially during my stay in Munich. Personal thanks to my family, Therese and my friends, who have stood by me throughout my education.

Contents

1 Introduction

Since the deregulation of electricity markets around the world, modeling of electricity prices has become a growing research area. Electricity prices as well as many other commodities exhibit strong seasonality. mean-reversion, jumps and stochastic volatility are other characteristics of electricity prices. In this master thesis we will investigate some models for electricity prices in the German market. In Section 2 the mathematical background for our model is given. In Section 3 the concept of maximum likelihood and the BHHH algorithm which is used for the parameter estimation are explained. Section 4 gives a short introduction to electricity markets. A data analysis is performed in Section 5 and in Section 6 we define our model and test the parameter estimation. In Section 7 several different versions of the model are investigated and we also look at how it behaves when simulating it. In Section 8 the pricing of futures is discussed. And finally in Section 9, we make some concluding remarks, as well as discuss future research areas.

2 Mathematical background

In this section we will go through some basic concepts that will be used through out the thesis.

2.1 Compound Poisson process

Let $N(t)$ be a Poisson process with intensity λ and let $Y_1, Y_2, ...$ be a sequence of independent identically distributed random variables with mean $\beta = \mathbb{E}[Y_i]$ that are independent of $N(t)$. We define the compound Poisson process $J(t)$ as

$$J(t) = \sum_{i=1}^{N_t} Y_i, t \geq 0.$$

The mean of the compound Poisson process is

$$\begin{aligned}
\mathbb{E}[J(t)] &= \sum_{k=1}^{\infty} \mathbb{E}[\sum_{i=1}^{k} Y_i \mid N(t) = k]\mathbb{P}[N(t) = k] \\
&= \sum_{k=1}^{\infty} \beta k \frac{(\lambda t)^k}{k!} e^{-\lambda t} \\
&= \beta k e^{-\lambda t} \sum_{k=1}^{\infty} \frac{(\lambda t)^{k-1}}{(k-1)!} \\
&= \beta \lambda t.
\end{aligned}$$

There are λt jumps in the time interval [0,t] on average and the average jump size is β

2.2 Vasicek interest rate model

In 1997 Vasicek [20] introduced a model for interest rates that captures the essential mean-reversion characteristic for interest rates. Let $W(t)$, $t \geq 0$, be a Brownian motion. The Vasicek model for the interest rate process $R(t)$ is given by

$$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t,$$

where κ , θ and σ are positive constants. Here κ is called the mean-reversion speed, θ is the long-run equilibrium level and σ is the volatility. The stochastic differential equation has an explicit solution given by

$$r_t = r_s e^{-\kappa(t-s)} + \theta(1 - e^{-\kappa(t-s)}) + \sigma \int_s^t e^{-\kappa(t-u)} dW_u,$$

where $s \leq t$. It can easily be showed that

$$\mathbb{E}[r_t \mid r_s] = r_s e^{-\kappa(t-s)} + \theta(1 - e^{-\kappa(t-s)}),$$

$$Var[r_t \mid r_s] = \frac{\sigma^2}{2a}(1 - e^{-2\kappa(t-s)}).$$

Since we now know the distribution we can use the maximum likelihood method, se Section 3.1 to estimate the parameters of the model.See also for example the book by Brigo and Mercurio [5].

2.3 GARCH models

Autoregressive conditional heteroscedasticity (ARCH) models were first introduced by Engle [8]. Bollerslev [4] developed the generalized ARCH (GARCH). A stochastic process X_t is called GARCH(p,q) if it follows the following equation

$$X_t = \sigma_t Z_t,$$

where Z_t is i.i.d $\mathcal{N}(0,1)$ and σ_t comes from the equation

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2,$$

where the constants $\omega, \alpha_i, \beta_j \geq 0$. The case $q = 0$ corresponds to ARCH(p) process. Bollerslev showed in [4] that the GARCH process to is second-order stationary if and only if

$$\omega > 0, \sum_{i=0}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1.$$

For further reading on GARCH we recommend Bollerslev [4] and the book by Straumann [18]

3 Parameter estimation methods

3.1 Maximum likelihood

The maximum likelihood is a method to estimate unknown parameters in statistical model for a given data set. The method assumes that the probability density function of the observed data is known, except for some unknown parameters. The parameters are estimated through maximizing the probability of getting the observed data from the given probability density function. Denote the probability density function for a random variable Y, conditioned on a set of parameters $\theta \in \Theta$, where Θ is domain in $\mathbb{R}$ by $f(y \mid \theta)$. Our observed sample values are $y_1, y_2, ..., y_n$, the probability of getting them is $f(y_1, y_2, ..., y_n \mid \theta)$ which is the joint probability density function of the entire sample. If the observations are independent and identically distributed (i.i.d.) we have

$$f(y_1, y_2, ..., y_n \mid \theta) = \prod_{i=1}^{n} f(y_i \mid \theta).$$

The likelihood function for the sample data is defined by

$$L(\theta \mid \mathbf{y}) = \prod_{i=1}^{n} f(y_i \mid \theta),$$

which we want to maximize, this maximization is however the same as maximizing the logarithm of the likelihood function because the logarithm is a monotonic function. Set

$$l(\theta \mid \mathbf{y}) := ln(L(\theta \mid \mathbf{y})).$$

It is for numerical reasons usually simpler to maximize the log-likelihood function and it is used instead of likelihood function.

$$l(\hat{\theta} \mid \mathbf{y}) = arg \max_{\theta \in \Theta} l(\theta \mid \mathbf{y}).$$

To find the maximum likelihood estimate we take the partial derivatives and set them equal to zero.

$$\frac{\partial l(\theta \mid \mathbf{y})}{\partial \theta} = \frac{\partial(ln(L(\theta_i \mid \mathbf{y})))}{\partial \theta_i} = 0 \qquad\qquad i = 1, ..., k.$$

Since we want to find a maximum the following condition also has to be satisfied

$$\frac{\partial^2 l(\theta \mid \mathbf{y})}{\partial \theta^2} = \frac{\partial^2(ln(L(\theta_i \mid \mathbf{y})))}{\partial \theta_i^2} < 0 \qquad\qquad i = 1, ..., k.$$

The set of parameters satisfying the two conditions above are a maximum likelihood estimation. The calculation of the derivatives might be very hard and one often needs some special algorithm to do that. In the paper from Arnericit et al. [2] it is very good explained how the maximization can be made, we will follow that work. We need an algorithm that after each iteration moves to a new value of the parameters at which $ln(L(\theta))$ is higher than at the previous step. The current value at iteration k is denoted by θ_k, how should then the next θ_{k+1} be chosen to get a higher value. To determine the best value of θ_{k+1} , a second-order Taylor approximation of $ln(L(\theta_{k+1}))$ around $ln(L(\theta_k))$ is used,

$$\begin{aligned} ln(L(\theta_{k+1})) & \\ &= ln(L(\theta_k)) + (\theta_{k+1} - \theta_k)^T \underbrace{\frac{\partial ln(L(\theta_k))}{\partial \theta_k}}_{=g_k} \\ &+ \frac{1}{2}(\theta_{k+1} - \theta_k)^T \underbrace{\frac{\partial^2 ln(L(\theta_k))}{\partial \theta_k^2}}_{=H_k}(\theta_{k+1} - \theta_k). \end{aligned} \tag{3.1}$$

Now we will maximize Equation 3.1 with respect to θ_{k+1},

$$\begin{aligned} \frac{\partial ln(L(\theta_{k+1}))}{\partial \theta_{k+1}} &= g_k + H_k(\theta_{k+1} - \theta_k) \\ H_k(\theta_{k+1} - \theta_k) &= -g_k \\ \theta_{k+1} &= \theta_k + (-H_k)^{-1} g_k. \end{aligned} \tag{3.2}$$

The Newton procedure uses Equation 3.2. From the current value of θ_k the step $(-H_k)^{-1}g_k$ is taken to get to the new value θ_{k+1}. However one normally also has a scalar λ_k, that guaranties that each step of the procedure provides an increase in $ln(L(\theta_k))$

$$\theta_{k+1} = \theta_k + \lambda_k(-H_k)^{-1} g_k, \tag{3.3}$$

where $(-H_k)^{-1}g_k$ is called the direction, denoted d_k, and λ_k is called the step size. Equation 3.3 is often referred to as the Newton-Raphson algorithm when the Hessian is determined analytically. Calculation of the Hessian is usually computation-intensive, i.e. analytical Hessian is rarely available. Therefore we need an alternative calculation of the Hessian which leads us to the BHHH algorithm which is described in the next section.

3.2 BHHH algorithm

The BHHH algorithm was introduced by Berndt, Hall, Hall and Hausman in [3] and is an extension of the Newton-Raphson algorithm. In our description of the algorithm we will follow the work from Arnericit et al. [2]. The algorithm uses an information identity. The iterative procedure is defined as

$$\begin{aligned} \theta_{k+1} &= \theta_k + \lambda_k d_k \\ d_k &= -H_k^{-1} g_k \\ H_k &= \sum_{t=1}^{T} g_t g^T \\ g_k &= \sum_{t=1}^{T} g_t. \end{aligned} \tag{3.4}$$

According to the relations in Equation 3.4, the information identity means that minus the expected Hessian at the true parameters is equal to the covariance matrix of the first derivatives. Which means that minus the Hessian can be approximated as an outer product of gradient (OPG). For further properties of the OPG see Arnericit et al. [2]. The numerical optimization procedure of the BHHH algorithm can be summarized in following steps:

1. Determine initial vector of parameters θ_s and convergence criteria.
2. Calculate a direction vector $[-H(\theta_k)]^{-1}$ where $H(\theta_k)$ is calculated by the OPG
3. Calculate a new vector $\theta_{k+1} = \theta_{k+1} + \lambda d_k$, where λ is scalar. Start with $\lambda = 1$. If $f(\theta_k + d_k) > \theta_k)$ try with $\lambda = 2$. If $f(\theta_k + 2d_k) > \theta_k)$ try with $\lambda = 3$, etc. until a λ is found for which $f(\theta_k + \lambda d_k)$ is a maximum.
4. If the convergence criteria is satisfied the algorithm stops, if not step 2 to 4 is repeated

The BHHH algorithm is implemented in R in the package Maxlik. In this work that package is used, therefore we will not go into further details of the algorithm.

4 Electricity market

In the last two decades a lot of countries around the world have started to deregulate their electric power sectors. Nowadays there are several countries with electricity market-places. The Nordic power market Nordpool was the first market place, where electricity could be traded across borders. In Germany electricity is traded at the European Energy Exchange since 2000.

Electricity prices are determined by the supply (generation) and the demand (consumption) on the market. Since electricity can not be stored in a direct way, generation and consumption have to be continuously balanced. The demand of electricity is very inelastic because consumers find it as a necessary commodity. Due to the inelastic demand and non storability of electricity, electricity prices have extreme spikes. The way in which electricity is generated varies between different markets. The Nordic market has a large part that comes from hydro and is therefore less influenced by fuel prices (coal, oil, natural gas) as compared with the German market, see Table 4.1. It is possible to save water for hydro generation in reservoirs, therefore the Nordic market have less spikes and prices are highly depended on reservoir levels.

Table 4.1: Composition of generation in Germany and Sweden 2006 (Source: Nordel [14] and VDEW [19])

	Germany	Sweden
Renewable	3.23%	46%
Nuclear	33.3%	44%
Thermal power	63.5%	10%

The supply and demand of electricity is influenced by a very large number of factors. The influence from weather and business cycles are reasons for the seasonal patterns in electricity markets. Political decisions also have a big influence on electricity markets. In 2005 the European Union introduced a system for CO_2- emissions which has a big impact on electricity generation based on fuels. The list of factors influencing electricity prices is very long and will not be discussed further, however we conclude that electricity prices have some special characteristics such as

1. Seasonality.
2. Spikes.
3. Mean-reversion.
4. Stochastic volatility.

5 Data analysis

We will work with data from the European Energy Exchange (EEX). The data is daily average spot prices base load from the period 1.1.2002-30.6.2008, which consists of 2373 data points. In Figure 5.1 our data set is plotted.

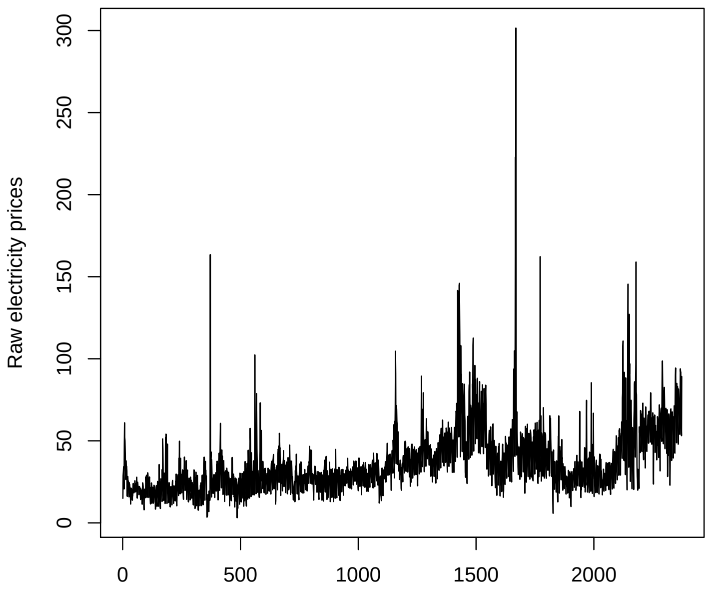


Figure 5.1: The electricity spot prices from EEX 1.1.2002-30.6.2008

We notice that our data is extremely volatile. In Table 5.1 some descriptive statistics for the data can be found. The data exhibits both positive and negative spikes/jumps. The mean is higher than the median which indicates that the data is skewed to the right.

Table 5.1: Descriptive statistics

Min	Median	Mean	Max	Var	SD	Skewness	Kurtosis
3.12	37.03	37.80	301.5	400.7	20.02	2.712	22.30

The positive skewness value confirms the possibility of a positively skewed distribution. The positive kurtosis value indicates a relatively peaked distribution. In Figure 5.2 is a plot of a kernel density estimate for the data. By looking at the kernel density estimate we see that the distribution is peaked and skewed to the right.

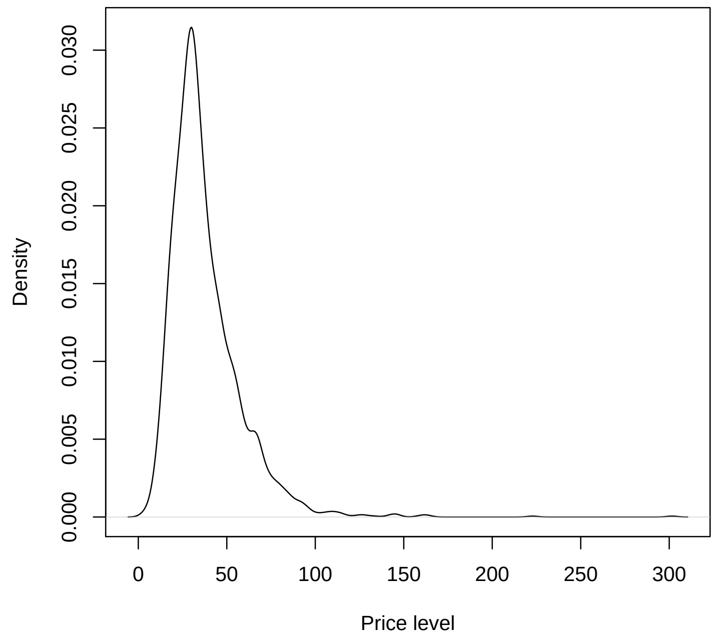


Figure 5.2: A kernel density estimate for the data

Figure 5.3 shows a normal qq-plot of the data. If the data would be normally distributed the graph would be a straight line. We can clearly see that our data does not follow the normal distribution since the tails are much heavier than the normal distribution.

Table 5.2: Median of data for weekdays and the difference from the median of the data

Weekday	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Median	35.72	37.03	38.38	37.21	34.8	27.83	21.08
Diff	2.82	4.13	5.48	4.31	1.9	-5.07	-11.82

From Table 5.2 we conclude that the data has a very strong weekly seasonality. Since weekends have much lower prices it seems that the prices are very sensitive to the demand. Therefore we also took a closer look at the data regarding if holidays could

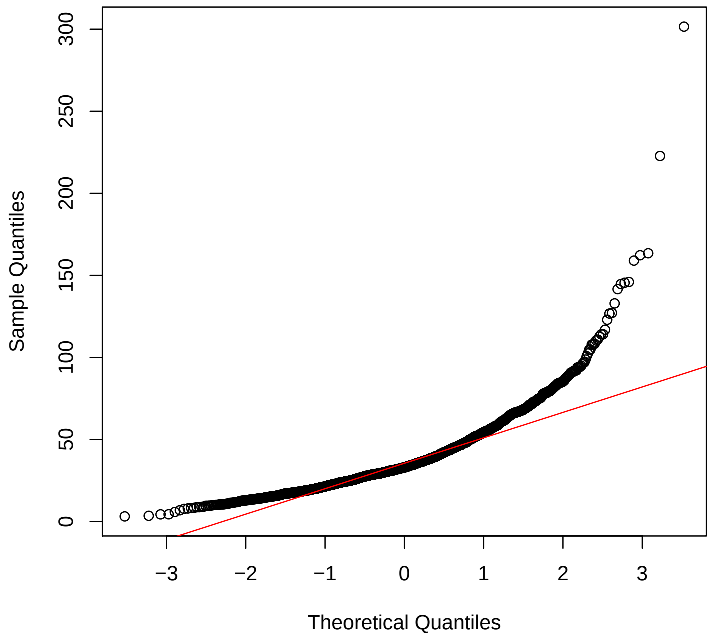


Figure 5.3: Normal qq-plot of the data

have the same effect. We found 11 holidays[1] that seems to have major effect on the prices. In Table 5.3 one can find the median for the holidays and also for the weekdays when the holidays were removed. We see that the holiday also has a major effect on the prices.

Table 5.3: Median of the data for weekdays and holidays

Mon	Tue	Wed	Thu	Fri	Sat	Sun	Holiday
35.61	37.025	38.295	37.14	34.695	27.825	21.06	21.03

We also conclude that it seems like the weekly seasonality can not be modeled by a sinus function and that dummy variables seem to be more appropriate.

In Table 5.4 it can be seen that the data has monthly seasonality but it is not as strong as the weekly seasonality. It does not look like the monthly seasonality can be modeled by a sinus function and that again dummy variables seem more appropriate.

[1] Neujahr, Karfreitag, Ostermontag, Maifeiertag, Christi Himmelfahrt, Pfingstmontag, Tag der Deutschen Einheit,1. Weihnachtstag, 2. Weihnachtstag, Heiligabend, Silvester

Table 5.4: Median of the data for months

Jan	Feb	Mar	Apr	May	Jun
33.015	35.46	31.325	30.54	30.235	34.775
Jul	Aug	Sep	Oct	Nov	Dec
31.69	31.36	34.08	35.78	35.28	34.87

In Figure 5.4 the running moving average for the median of the data with window size 91 days is plotted. As the running moving average at point j with window size k we use

$$MA(k) = median((X_i)_{i=j-k}^{j}).$$

One can see a trend in the data, which depends on factors influencing the electricity prices. The long run trend has probably a high degree of dependence from CO_2 and fuel prices. The big downfall in 2006 was due to that the CO_2 prices fell dramatically, then in 2008 they rose quickly again. There has also been a large increase in fuel prices over the last years. To be able to model the trend in a good way one probably needs to consider both CO_2 and fuel prices. Future prices might give an indication of the trend expected on the market.

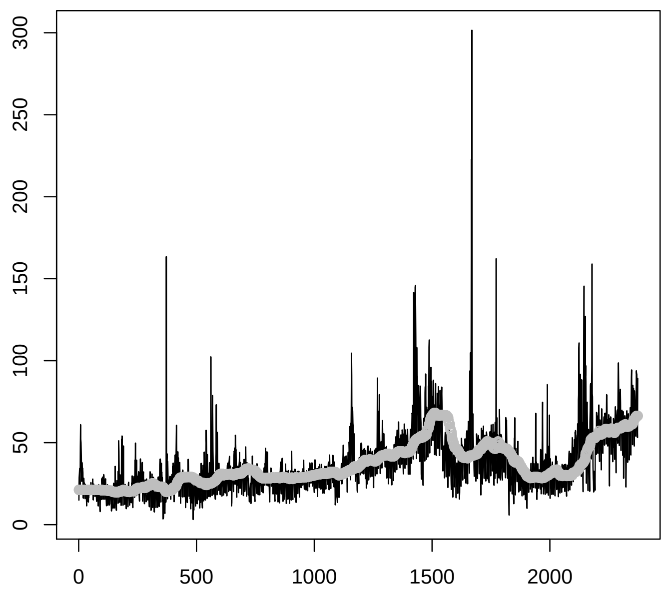


Figure 5.4: Running moving average for the data set with window size 91

Table 5.5 shows the result of a linear regression for the data. The R^2 value is 0.3348, which says that the linear trend can explain 33 percent of the fluctuations. In Figure 5.5 the trend together with the data is plotted. We see that a linear trend can be used as an approximation of the trend but it is not a particulary good one.

Table 5.5: Estimation liner trend

Coefficient	Estimate	Std.Error	t value	Pr(>|t|)
a	20.66	0.5112	40.41	<2e-16
b	0.01278	0.0003761	33.99	<2e-16

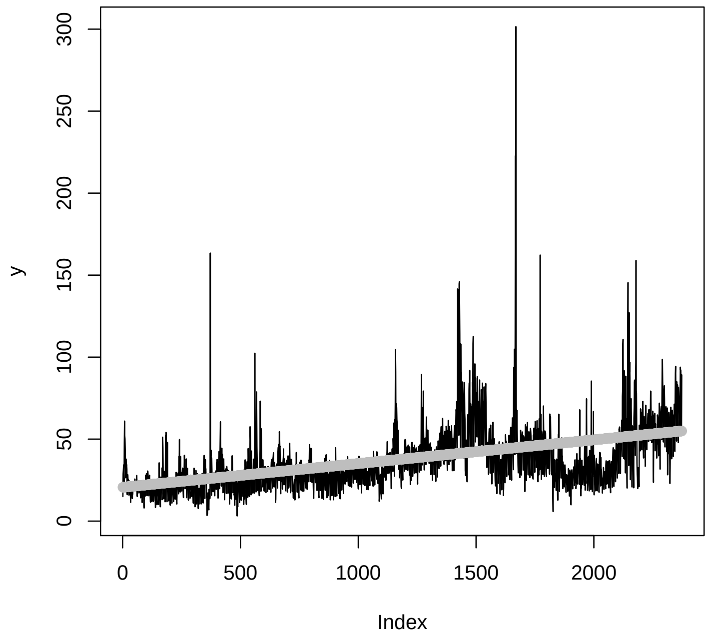


Figure 5.5: Linear trend for the data

6 Model setup

The model that we will investigate contains of two parts, a jump process and a diffusion process with stochastic volatility. Similar models have been investigated in Knittel and Roberts [11] and Escribano et al. [9].

Continuous time model

The diffusion process in continuous time without stochastic volatility is defined as

$$dD_t = \kappa(\theta_t - D_t)dt + \sigma dW_t,$$

where W_t is a Brownian motion. The process is a so called Vasicek process, see Section 2.2. The jump process is a compounded Poisson process, see Section 2.1.

$$J_t = \sum_{i=1}^{N_t} Y_i,$$

where Y_i are normally distributed with parameter μ_J and σ_J. N_t is poisson process with intensity λ. Merging the two parts gives us

$$dS_t = \kappa(\theta_t - S_t)dt + \sigma dW_t + dJ_t.$$

Discrete time model

We will now derive the model in discrete time. The Euler scheme with discretization factor Δ for the diffusion part is

$$D_t = D_{t-\Delta} + \kappa(\theta - D_{t-\Delta})\Delta + \sigma Z_t,$$

where Z_t are i.i.d $\mathcal{N}(0, \sqrt{\Delta})$.

If we look at the compounded Poisson process in a very short interval Δ we will either have a jump or do not a jump

$$\Delta J_t = \begin{cases} 0 \\ Y_i. \end{cases}$$

The probability of exactly one jump in an interval Δ is

$$\mathbb{P}(N_{t+\Delta} - N_t = 1) = \underbrace{e^{-\lambda\Delta}}_{\approx 1}\lambda\Delta = \lambda\Delta.$$

From Section 2.1 we also know that there is λt jumps in the time interval [0,t] on average, which corresponds to the approximation above. Merging the two parts in discrete time with $\theta = 0$ gives us

$$S_t = \begin{cases} S_{t-\Delta}(1-\kappa\Delta) + \sigma\sqrt{\Delta}Z_{1,t}, & \textit{with probability } 1-\lambda\Delta \\ S_{t-\Delta}(1-\kappa\Delta) + \sigma\sqrt{\Delta}Z_{1,t} + \mu_J + \sigma_J Z_{2,t}, & \textit{with probability } \lambda\Delta, \end{cases}$$

where $Z_{1,t}$ are i.i.d $\mathcal{N}(0, \sqrt{\Delta})$ and $Z_{2,t}$ are i.i.d $\mathcal{N}$(0,1).

Setting $\Delta = 1$ and adding GARCH(1,1) behavior gives us the following model

$$S_t = \begin{cases} S_{t-1}(1-\kappa) + \sigma_t Z_{1,t}, & \textit{with probability } 1-\lambda \\ S_{t-1}(1-\kappa) + \sigma_t Z_{1,t} + \mu_J + \sigma_J Z_{2,t}, & \textit{with probability } \lambda, \end{cases}$$

where $Z_{2,t}$ and $Z_{1,t}$ are i.i.d $\mathcal{N}(0,1)$ and σ_t^2 follows

$$\sigma_t^2 = w + \alpha(\sigma_{t-1} Z_{1,t-1})^2 + \beta \sigma_{t-1}^2.$$

Due to independence of $Z_{1,t}$ and $Z_{2,t}$ and the convolution of the normal cumulative distribution function we can write

$$S_t = \begin{cases} S_{t-1}(1-\kappa) + \sigma_t Z_t, & \textit{with probability } 1-\lambda \\ S_{t-1}(1-\kappa) + \mu_J + \sqrt{\sigma_t^2 + \sigma_J^2} Z_t, & \textit{with probability } \lambda, \end{cases} \tag{6.1}$$

where Z_t are i.i.d $\mathcal{N}(0,1)$ and σ_t^2 follows

$$\sigma_t^2 = w + \alpha(\sigma_{t-1} Z_{t-1})^2 + \beta \sigma_{t-1}^2.$$

Remark The reason for setting $\theta = 0$ is that we will incorporate θ in $f(t)$, a deterministic function that adjust for the seasonality in the data.

Remark If we will work on a different timescale than $\Delta = 1$ we must adjust the parameters.

Remark The discrete time Vasicek model is actually nothing but an AR(1) model.

6.1 Parameter estimation

The probability density function for the model in Equation 6.1 is

$$\begin{aligned} f(S_t, S_{t-1}) = \lambda \exp\Bigg[&\frac{-(S_t - (1-\kappa)S_{t-1} - \mu_J)^2}{2(\sigma_t^2 + \sigma_J^2)} \frac{1}{\sqrt{2\pi(\sigma_t^2 + \sigma_J^2)}}\Bigg] \\ &+ (1-\lambda)\exp\Bigg[\frac{-(S_t - (1-\kappa)S_{t-1})^2}{2\sigma_t^2} \frac{1}{\sqrt{2\pi\sigma_t^2}}\Bigg]. \end{aligned}$$

We will use the BHHH algorithm, see Section 3.2 for the parameter estimation of the model. We calculate σ_t iteratively with the approximation that we have λ jumps with hight μ_J on average. First an initialization has to be made for the first values of σ_t then we use the following formula

$$\sigma_t^2 = \omega + \alpha(S_{t-1} - (1-\kappa)S_{t-2} - \lambda\mu_J) + \beta\sigma_{t-1}^2.$$

6.2 Simulation

We will now make an illustration of our model. We start by simulating a path of length $n = 2000$ for the parameter vector $(\omega, \alpha, \beta, \kappa, \mu_J, \lambda, \sigma_J) = (5, 0.2, 0.45, 0.25, 10, 0.05, 25)$, the path can be view in Figure 6.1.

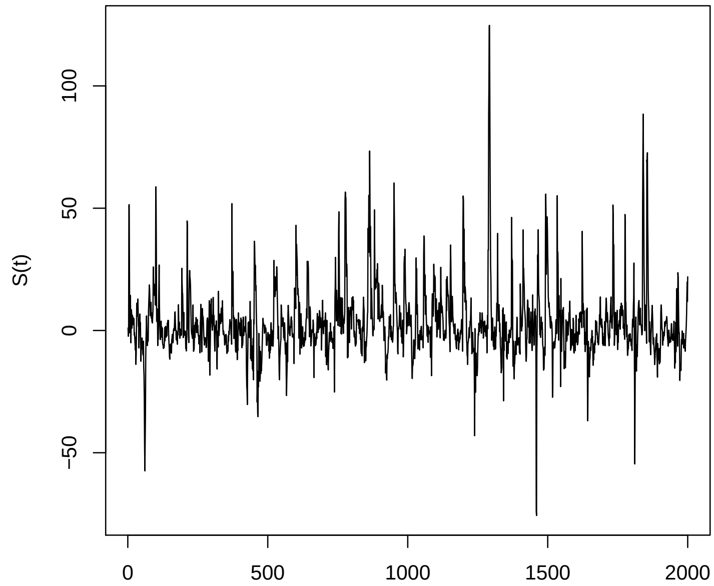


Figure 6.1: Simulation from the model

6.3 Restimation

Further on we will estimate the parameters in our model from the simulated paths. By doing so one gets an idea how good the parameter estimation works. 100 simulations with the following estimation were done then we take the 5%- and 95% quantile, the result can be found in Table 6.1.

Table 6.1: Result from the parameter estimation

Parameter	Original value	5% quantile	95% quantile
ω	5	4.375	6.370
α	0.2	0.1852	0.2937
β	0.45	0.3688	0.4875
κ	0.25	0.2285	0.2698
μ_J	10	5.558	13.975
λ	0.05	0.04222	0.06541
σ_J	25	22.00	28.04

We notice that the original value is always between the two quantiles, therefore the estimation procedure works very well. The procedure has been tested for lots of different values and also for different starting values and the estimation works well.

6.4 Modeling seasonality

Electricity prices often has several seasonal patterns; daily, weekly and monthly seasonality are common. Seasonal patterns are often described with a deterministic function, which usually is done additive when the spot price is modeled directly and multiplicative when $\ln S(t)$ is modeled. Sinusoidal functions with different periodicity (yearly, half yearly, weekly etc) are often used for the modeling. Another approach is to use dummy variables for weekdays and months. In Lucia and Schwartz [7] both dummy variables and sinusoidal functions were tested on the Nordic power market. The result was that dummy variables increased the likelihood for the model. However dummy variables requires more variables than sinusoidal modeling. Linear trend might also be incorporated for modeling seasonality, this was done in Geman and Roncoroni [10]. In Burger et. al [6] the seasonality is modeled using load forecasting and the availability for plants. Since our initial data analysis does not indicate any sinusoidal patterns we will use dummy variables to model the seasonality. The data analysis also indicated a linear trend which will be incorporated in the model.

7 Model investigation

In this section several different models on our data will be tested. We will both model the spot price and the logarithm of the spot price. To be able to conclude which model that performs best we will use several criteria:

1. Schwarz criterion, see Section 7.1.
2. Ex post goodness of fit criteria, see Section 7.5.
3. Comparing the distributional characteristics of the original data and simulated prices from the models, see Section 7.6.

7.1 Schwarz criterion

Schwarz criterion (SC) was introduced in [17], it is sometimes called the Bayesian information criterion. Let n denote the number of observations, p the number of free parameters to be estimated and LL the maximized value of the log-likelihood function for the estimated model. The formula for the SC is

$$SC(model) = -2 \cdot LL(model) + \log(n) \cdot p.$$

The criterion says that the model with the smallest SC-value should be selected. Lower SC implies either fewer explanatory variables, better fit, or both. Another similar criteria is the Akaike information criterion [1], where $\log(n)$ is replaced by 2. Since $\log(n) > 2$ in our case, the SC penalizes more parameters more strongly than the Akaike information criterion.

7.2 Modeling spot prices

Model 7.1.

$$
\begin{aligned}
P_t &= S_t + f_t \\
S_t &= \begin{cases} S_{t-1}(1-\kappa) + \sigma_t Z_t, & \textit{with probability } 1-\lambda \\ S_{t-1}(1-\kappa) + \mu_J + \sqrt{\sigma_t^2 + \sigma_J^2} Z_t, & \textit{with probability } \lambda \end{cases} \\
\sigma_t^2 &= w + \alpha(\sigma_{t-1} Z_{1,t-1})^2 + \beta \sigma_{t-1}^2,
\end{aligned}
$$

where Z_t are[2] i.i.d $\mathcal{N}$(0,1).

We will try four different seasonality functions

$$
\begin{aligned}
f_{a,t} &= B_0 + B_1 t + \sum_{i=1}^{7} (D_i \times \textit{weekday}) + D_8 \times \textit{holiday}, \\
f_{b,t} &= B_0 + B_1 t + \sum_{i=1}^{6} (D_i \times \textit{weekday}) + D_7 \times \textit{holiday/sunday}, \\
f_{c,t} &= B_0 + B_1 t + D_0 \times \textit{holiday/weekend}, \\
f_{d,t} &= B_0 + B_1 t + \sum_{i=1}^{6} (D_i \times \textit{weekday}) + D_7 \times \textit{holiday/sunday} + \sum_{i=1}^{12} (M_i \times \textit{month}).
\end{aligned}
$$

Remark In seasonality function $f_{a,t}$ the value for a holiday depends both on which weekday it occurs and the dummy variable for holidays. But in $f_{b,t}$ the weekday that a holiday occurs on does not influence the value.

The probability density function for Model 7.1 is

$$
\begin{aligned}
f(P_t, P_{t-1}) = \lambda \exp&\left[\frac{-(P_t - f_t - (1-\kappa)(P_{t-1} - f_{t-1}) - \mu_J)^2}{2(\sigma_t^2 + \sigma_J^2)} \frac{1}{\sqrt{2\pi(\sigma_t^2 + \sigma_J^2)}} \right] \\
&+ (1-\lambda) \exp\left[\frac{-(P_t - f_t - (1-\kappa)(P_{t-1} - f_{t-1}))^2}{2\sigma_t^2} \frac{1}{\sqrt{2\pi\sigma_t^2}} \right].
\end{aligned}
$$

We will use the BHHH algorithm, see Section 3.2 for the parameter estimation. The result from the parameter estimation can be found in Table 7.1.

[2]from now on Z_t are always i.i.d $\mathcal{N}$(0,1)

Table 7.1: Estimated Parameters for Model 7.1

Variables	*Model 1 a*	*Model 1 b*	*Model 1 c*	*Model 1 d*
ω	3.814	3.752	3.345	3.891
α	0.2969	0.2913	0.1715	0.3010
β	0.5169	0.5273	0.7421	0.5163
κ	0.1546	0.1542	0.2292	0.1561
μ_J	12.26	13.35	18.41	14.52
λ	0.04745	0.04421	0.02696	0.04199
σ_J	21.86	21.86	28.70	22.013
B_0	19.64	19.72	20.00	20.24
B_1	0.01140	0.01136	0.01204	0.01125
D_0			-9.412	
D_1	2.358	2.540		2.437
D_2	3.864	3.734		3.937
D_3	4.206	3.942		4.290
D_4	3.857	3.579		3.953
D_5	2.128	1.738		2.184
D_6	-4.751	-4.867		-4.643
D_7	-10.70	-10.50		-10.69
D_8	-10.34			
M_1				-1.769
M_2				-0.2800
M_3				-1.353
M_4				-0.7006
M_5				-1.666
M_6				0.1526
M_7				-0.507
M_8				-1.058
M_9				0.9404
M_{10}				2.712
M_{11}				-1.695
M_{12}				-2.775
LL	-7645	-7631	-8208	-7621
SC	15347	15316	16453	15336

We conclude that Sundays and holidays should be treated in one dummy variable together. Adding the monthly seasonality dummy variables does not improve the model. The second seasonality function $f_{b,t}$ had the lowest SC-value, therefore from now on we will only work with the seasonality function $f_{b,t}$.

We will now remove the GARCH behavior from the model to be able to see how that effects the model.

Model 7.2.

$$P_t = S_t + f_t$$
$$S_t = \begin{cases} S_{t-1}(1-\kappa) + \sigma \cdot Z_t, & \textit{with probability } 1-\lambda \\ S_{t-1}(1-\kappa) + \mu_J + \sqrt{\sigma^2 + \sigma_J^2} Z_t, & \textit{with probability } \lambda \end{cases}$$
$$f_t = B_0 + B_1 t + \sum_{i=1}^{6} (D_i \times weekday) + D_7 \times holiday/sunday.$$

The result from the parameter estimation can be found in Table 7.2. We will now remove the jump part and then later on also the GARCH behavior to see how that effects the model.

Model 7.3.

$$P_t = S_t + f_t$$
$$S_t = S_{t-1}(1-\kappa) + \sigma_t Z_t$$
$$\sigma_t^2 = w + \alpha(\sigma_{t-1} Z_{1,t-1})^2 + \beta \sigma_{t-1}^2$$
$$f_t = B_0 + B_1 t + \sum_{i=1}^{6} (D_i \times weekday) + D_7 \times holiday/sunday.$$

Model 7.4.

$$P_t = S_t + f_t$$
$$S_t = S_{t-1}(1-\kappa) + \sigma \cdot Z_t$$
$$f_t = B_0 + B_1 t + \sum_{i=1}^{6} (D_i \times weekday) + D_7 \times holiday/sunday.$$

The result from the parameter estimation can be found in Table 7.2

Table 7.2: Estimated Parameters for Model 7.1b, 7.2, 7.3 and 7.4

Variables	*Model 7.1b*	*Model 7.2*	*Model 7.3*	*Model 7.4*
ω	3.752		4.195	
α	0.2914		0.7116	
β	0.5273		0.4886	
σ		4.590		10.96
κ	0.1543	0.1883	0.1123	0.2922
μ_J	13.36	7.976		
λ	0.04421	0.1031		
σ_J	21.86	30.99		
B_0	19.72	18.39	18.37	19.58
B_1	0.01136	0.01124	0.01501	0.01455
D_1	2.541	4.902	3.726	4.435
D_2	3.734	6.155	5.166	7.799
D_3	3.942	6.594	5.052	6.878
D_4	3.579	6.098	4.670	6.493
D_5	1.738	4.000	2.535	2.763
D_6	-4.867	-3.344	-4.000	-5.923
D_7	-10.50	-9.473	-9.871	-13.16
LL	-7631	-7900	-7874	-9049
SC	15316	15847	15792	18135

In Table 7.2 it can be seen that both the spikes and the GARCH behavior have large effect on the log-likelihood value. We also notice that the stationarity condition $\alpha+\beta < 1$ does not hold in Model 7.3. From the SC-value we conclude that Model 7.1 with $f_{b,t}$ is the best one, which confirms that a good model should incorporate both jumps and GARCH behavior.

Some different GARCH processes were tested to see the effects on the model. The result can be found in Table 7.3

Table 7.3: Estimated Parameters for different GARCH models

Variables	*GARCH(1,1)*	*ARCH(1)*	*GARCH(2,2)*
ω	3.752	13.24	5.818
α_1	0.2914	0.3576	0.3020
β_1	0.5273		-0.05123
α_2			0.1494
β_2			0.3171
κ	0.1543	0.1809	0.1553
μ_J	13.36	9.715	13.23
λ	0.04421	0.08515	0.04458
σ_J	21.86	19.65	21.77
B_0	19.72	20.44	19.84
B_1	0.01136	0.01011	0.01126
D_1	2.541	2.625	2.500
D_2	3.734	3.931	3.715
D_3	3.942	4.441	3.937
D_4	3.579	4.048	3.554
D_5	1.738	1.949	1.701
D_6	-4.867	-5.205	-4.926
D_7	-10.50	-11.00	-10.55
LL	-7631	-7709	-7631
SC	15316	15469	15320

The ARCH(1) improves the model a lot, compared to Model 7.2 with constant volatility but the GARCH(1,1) outperforms ARCH(1). We further notice that our estimates for the GARCH(2,2) fails the condition $\alpha_i \ \beta_j > 0$, therefore our estimates is not a valid GARCH process. Since the GARCH(2,2) model does not have better log-likelihood value we will not try to change our estimator to obtain α_i, $\beta_j > 0$. We conclude that GARCH(1,1) seems to be the best option.

The next step in our investigation will be to see how good our estimation of the trend is, we will now try a moving average instead of the linear trend. Figure 7.1 shows a plot of the new data that we are modeling. We have removed the running moving average for the median of the data using a window size of 91 days from the spot price.

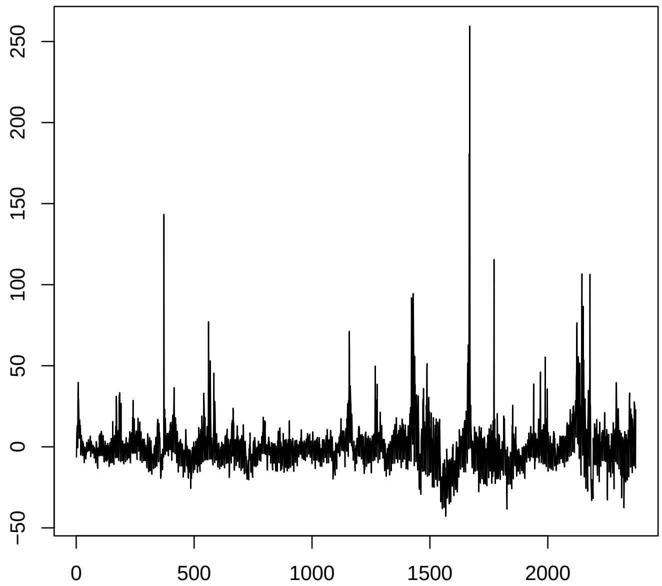


Figure 7.1: Spotprice-MA(91)

Model 7.5.

$$
\begin{aligned}
P_t &= S_t + f_t + MA(91, P_t) \\
S_t &= \begin{cases} S_{t-1}(1-\kappa) + \sigma_t Z_t, & \textit{with probability } 1-\lambda \\ S_{t-1}(1-\kappa) + \mu_J + \sqrt{\sigma_t^2 + \sigma_J^2} Z_t, & \textit{with probability } \lambda \end{cases} \\
\sigma_t^2 &= w + \alpha(\sigma_{t-1} Z_{1,t-1})^2 + \beta \sigma_{t-1}^2 \\
f_t &= B_0 + \sum_{i=1}^{6} (D_i \times weekday) + D_7 \times holiday/sunday,
\end{aligned}
$$

where MA(91) is running moving average for the median of the data using a window size of 91 days.

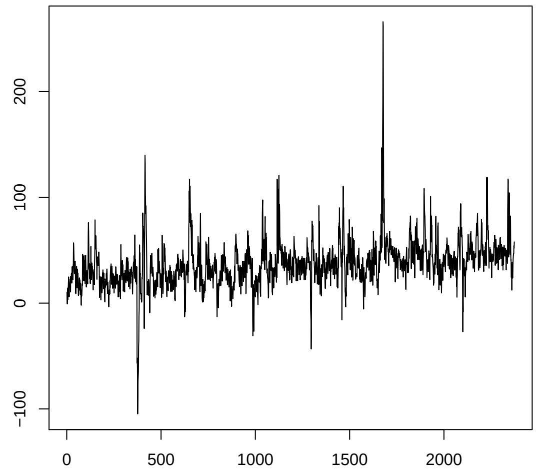
Table 7.4: Estimated Parameters for Model 7.1 and 7.5

Variables	*Model 7.1*	*Model 7.5*
ω	3.752	2.581
α	0.2914	0.2725
β	0.5273	.6027
κ	0.1543	0.2725
μ_J	13.36	15.03
λ	0.04421	0.03316
σ_J	21.86	25.13
B_0	19.72	2.005
B_1	0.01136	
D_1	2.541	0.05765
D_2	3.734	1.122
D_3	3.942	1.319
D_4	3.579	0.9494
D_5	1.738	-0.813
D_6	-4.867	-7.295
D_7	-10.50	-12.88
LL	-7631	-7543

In Table 7.4 we see that Model 7.5 has a much higher log-likelihood value than Model 7.1. One can also notice that κ, the mean reversion has a much higher value for Model 7.5, this is due to that the moving average models the trend much better than our linear trend in Model 7.1 does.

7.2.1 Simulation

Figure 7.2 shows a simulation of Model 7.1 using the estimated parameters. We see that our simulated paths looks similar to the original data expect for that we also get negative values from the model. This is a property that is highly unwanted because it will not happen in reality. The reason for why the model easily get negative values is that the model tries to model all jumps in the data in the same way but in reality most of the negative jumps occur directly after a positive jump. This is a feature which the model does not capture.



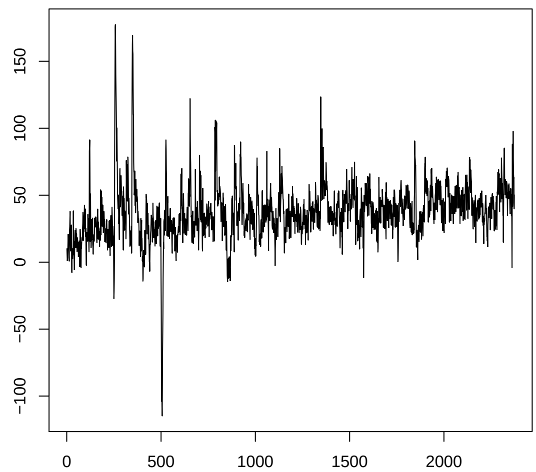


Figure 7.2: Simulation of prices for Model 7.1

7.3 Modeling log-prices

The next step in our investigation will be to model the logarithm of the prices. By modeling the log-prices instead, we can make a transform back to get the real prices, by doing so we will not get negative values from our model. However if we directly try to model the log-prices, the seasonality will be transformed to a different scale. Because of the strong trend that we have in the data the weekly seasonality will be hard to model. Therefore we will adjust for the weekly seasonality before we take the logarithm. The adjustment for the weekly seasonality is done by adjusting for the median of the data for each weekday. In Figure 7.3 a plot of the adjusted data can be viewed.

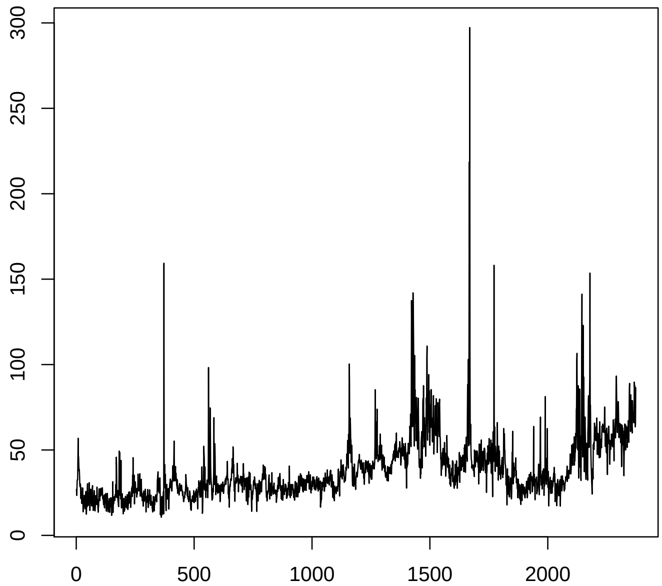


Figure 7.3: Spot prices adjusted for weekly seasonality

In Figure 7.4 a plot of logarithm of the adjusted data can be viewed. As one can see the new data shares the same characteristics as the data we modeled before but the trend seems to have bigger impact.

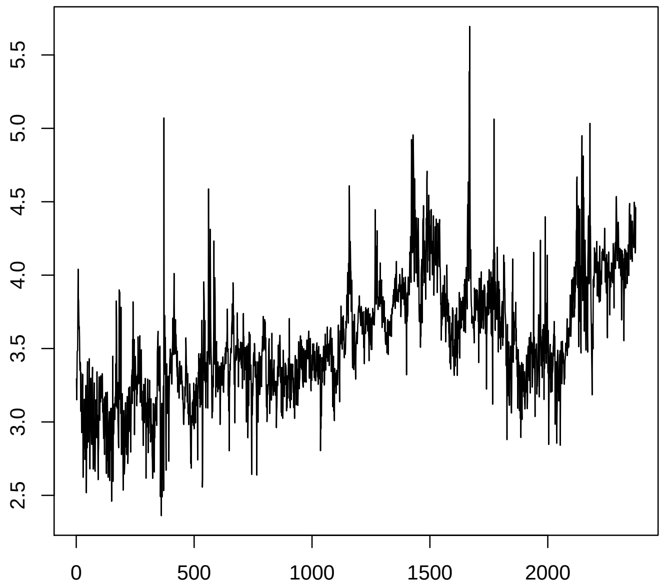


Figure 7.4: The logarithm of the adjusted data

The first model that we will test is

Model 7.6.

$$
\begin{aligned}
\log(P_t - f_{1,t}) &= S_t + f_{2,t} \\
S_t &= \begin{cases} S_{t-1}(1-\kappa) + \sigma_t Z_t, & \textit{with probability } 1-\lambda \\ S_{t-1}(1-\kappa) + \mu_J + \sqrt{\sigma_t^2 + \sigma_J^2} Z_t, & \textit{with probability } \lambda \end{cases} \\
\sigma_t^2 &= w + \alpha(\sigma_{t-1} Z_{1,t-1})^2 + \beta \sigma_{t-1}^2 \\
f_{1,t} &= Median(P_t) - \sum_{i=1}^{7} (Median(weekday)) - Median(holiday) \\
f_{2,t} &= B_0 + B_1 t.
\end{aligned}
$$

We will use the same estimation procedure as before and the results can be found in Table 7.5. The next step is to try a moving average instead of the linear trend.

Model 7.7.

$$S_t = \log(P_t - f_{1,t}) - MA(91, \log(P_t - f_{1,t}))$$
$$S_t = \begin{cases} S_{t-1}(1-\kappa) + \sigma_t Z_t, & \textit{with probability } 1-\lambda \\ S_{t-1}(1-\kappa) + \mu_J + \sqrt{\sigma_t^2 + \sigma_J^2} Z_t, & \textit{with probability } \lambda \end{cases}$$
$$\sigma_t^2 = w + \alpha(\sigma_{t-1} Z_{1,t-1})^2 + \beta\sigma_{t-1}^2$$
$$f_{1,t} = Median(P_t) - \sum_{i=1}^{7}(Median(weekday)) - Median(holiday),$$

where MA(91) is running moving average for the median of the data using a window size of 91 days.

Table 7.5: Estimated Parameters for Model 7.6 and 7.7

Variables	*Model 7.6*	*Model 7.7*
ω	0.004483	0.003993
α	0.2253	0.2030
β	0.5273	0.5480
κ	0.1641	0.2866
μ_J	0.0514	0.07442
λ	0.1018	0.0947
σ_J	0.3065	0.3273
B_0	3.057	
B_1	0.0003976	
LL	837	922

We notice that also here κ has a much larger value then we are working with the model with the removed moving average.

Remark Since we are working on a different data set now we can not compare this model to the previous once. However we will do a post goodness of fit test to be able to compare them later on.

7.3.1 Simulation

In Figure 7.5 Model 7.6 is simulated using the estimated parameters. We see that it is no longer possible to get negative values from the model. The simulated paths looks very similar to our original data set.

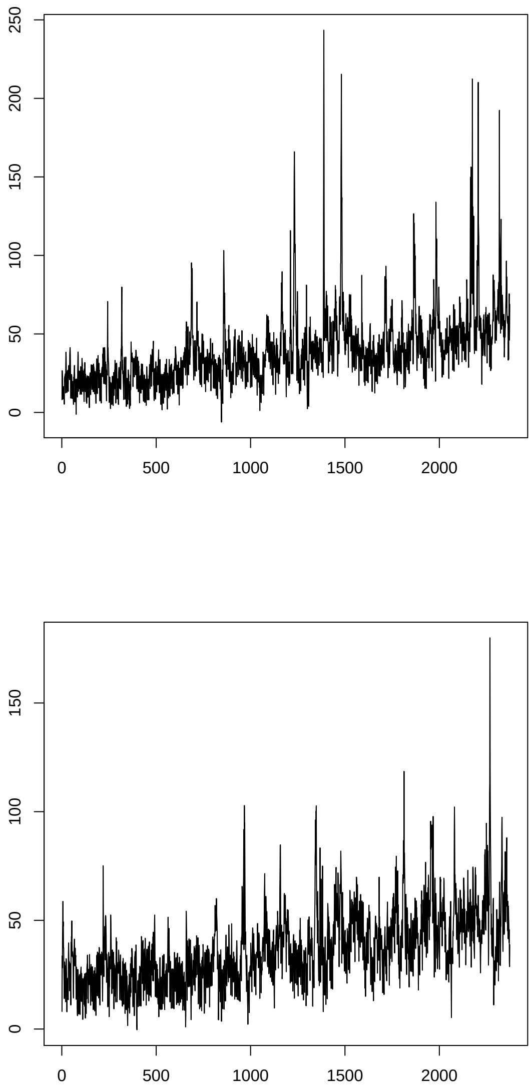


Figure 7.5: Simulation of prices for Model 7.5

7.4 Extreme value approach

We will now take on a different approach when modeling the jumps. Since the negative jumps mostly come after a positive jump, we will ignore the negative jumps. Lets start by defining the model in discrete time

Model 7.8.

$$
\begin{aligned}
P_t &= S_t + f_t \\
f_t &= Median(P_t) - \sum_{i=1}^{7}(Median(weekday)) - Median(holiday) - MA(91, P_t) \\
S_t &= \begin{cases} S_{t-1}(1-\kappa) + \sigma_t Z_t, & \textit{with probability } 1-\lambda \\ S_{t-1}(1-\kappa) + \sigma_t Z_t + J_t, & \textit{with probability } \lambda, \end{cases}
\end{aligned}
$$

where J_t are the positive jumps with some distribution.

The estimation of the model is now more complex since we cannot use a mixed normal model for the estimation. We will instead separate the estimation into three parts. First, we adjust for the weekly seasonality and the moving average. Let $(X_i)_{i=1}^n$ denote the adjusted prices, which is plotted in Figure 7.6.

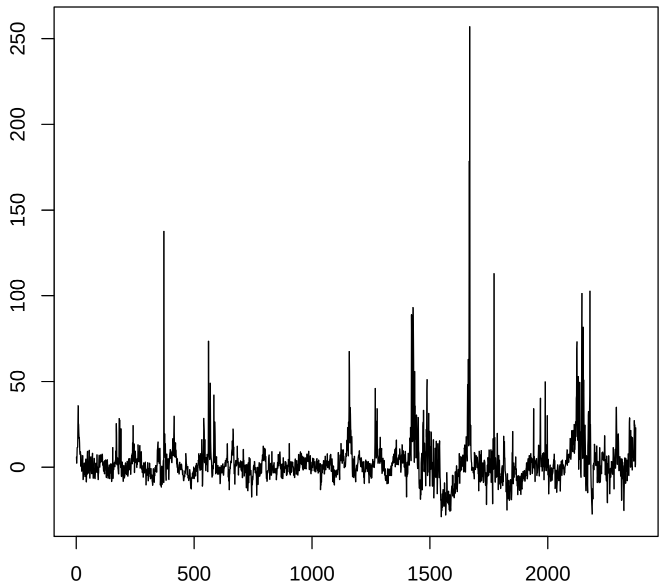


Figure 7.6: Plot of the adjusted data $(X_i)_{i=1}^n$

From $(X_i)_{i=1}^n$ we will extract the jumps $(J_i)_{i=1}^n$ To do this we look at the price differences $(X_i - X_{i-1})_{i=2}^n$, we first estimate the standard deviation from this series. If a price changes deviates more than 1 st.dev we consider it as a jump. That gives us a jump frequency of 0.094. All jumps is being placed of by the cut off values, then we estimate the parameters of mean-reverting diffusion from the filtered series $(X_i - J_i)_{i=1}^n$, which is plotted in Figure 7.7. The result can be viewed in Table 7.6.

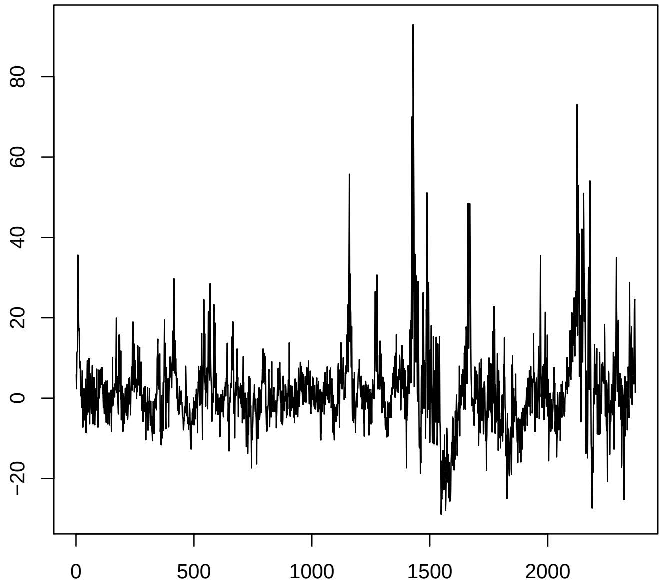


Figure 7.7: Plot of the adjusted data $(X_i - J_i)_{i=1}^n$

Table 7.6: Estimated parameters for Model 7.8

σ	κ	LL
6.119	0.1805	-7658

From the series of removed jumps (J_i) we only consider the positive jumps, the positive jumps has a jump frequency of 0.0476, a histogram of the jumps can be found in Figure 7.8.

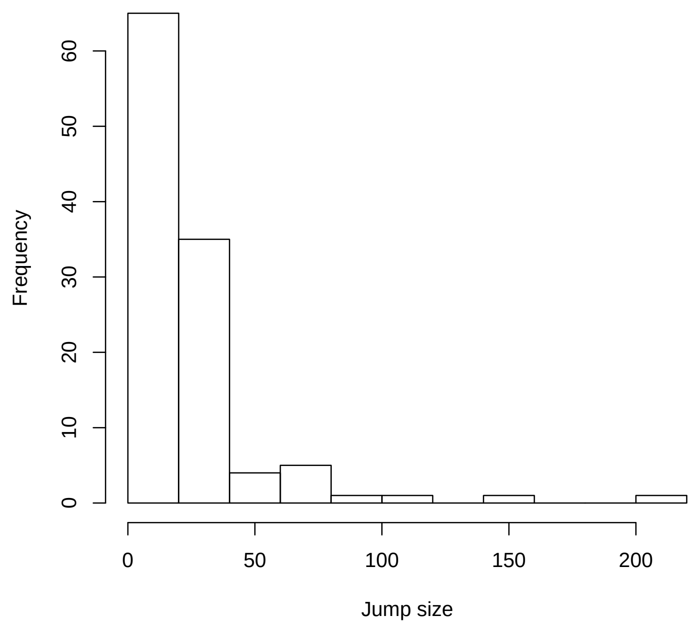


Figure 7.8: Histogram of the positive jumps

The exponential distribution is often used to model the tails. Using Matlab we fitted the data to the exponential distribution. In Figure 7.9 is a probability plot for the exponential cumulative distribution function. We see that it is a poor fit.

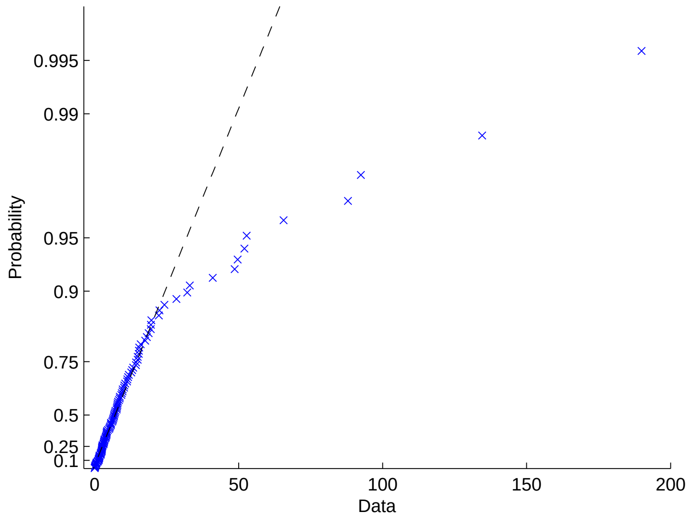


Figure 7.9: Probability plot for the exponential distribution

The Generalized Pareto distribution (GPD) is another distribution that is often used to model the tails. We use the R package POT when working with the GPD. The GPD has three parameters μ (location), σ (scale) and ξ (shape). The cumulative distribution function is defined as

$$G(x) = 1 - \left[1 - \frac{\xi(x-\mu)}{\sigma}\right]^{-1/\xi}$$

for $1 - \xi(x-\mu)/\sigma > 0$, $\sigma > 0$ and $x > \mu$. The mean is given by $\mu + \sigma/(1-\xi)$.

For location parameter $\mu = 1$ we get the estimates $\sigma = 7.6602$ (scale) and $\xi = 0.4941$ (shape). In Figure 7.10 is a probability plot for the GDP. We see that it is a very good fit.

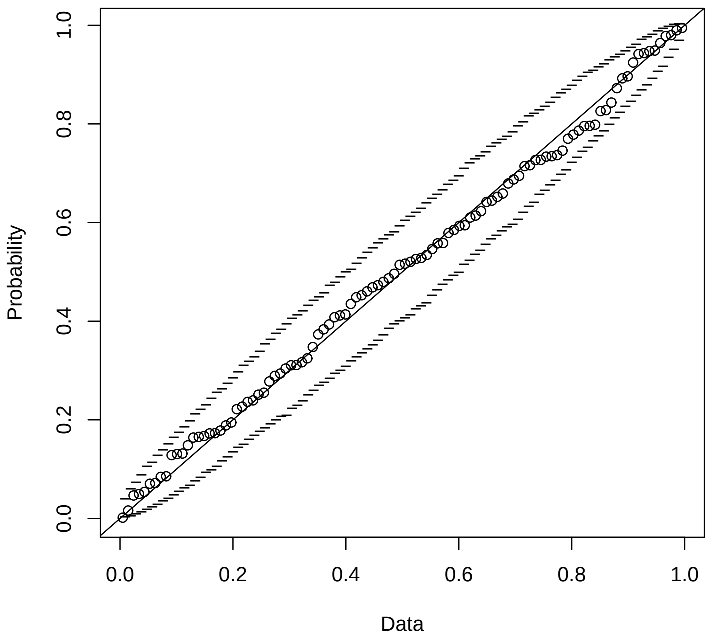


Figure 7.10: Probability plot for the Generalized Pareto distribution

7.4.1 Simulation

In Figure 7.11 Model 7.8 with GDP jumps is simulated using the estimated parameters. The simulated paths looks very similar to our original data set. We notice that it is possible to get negative values therefore modeling the log prices might be more appropriate.

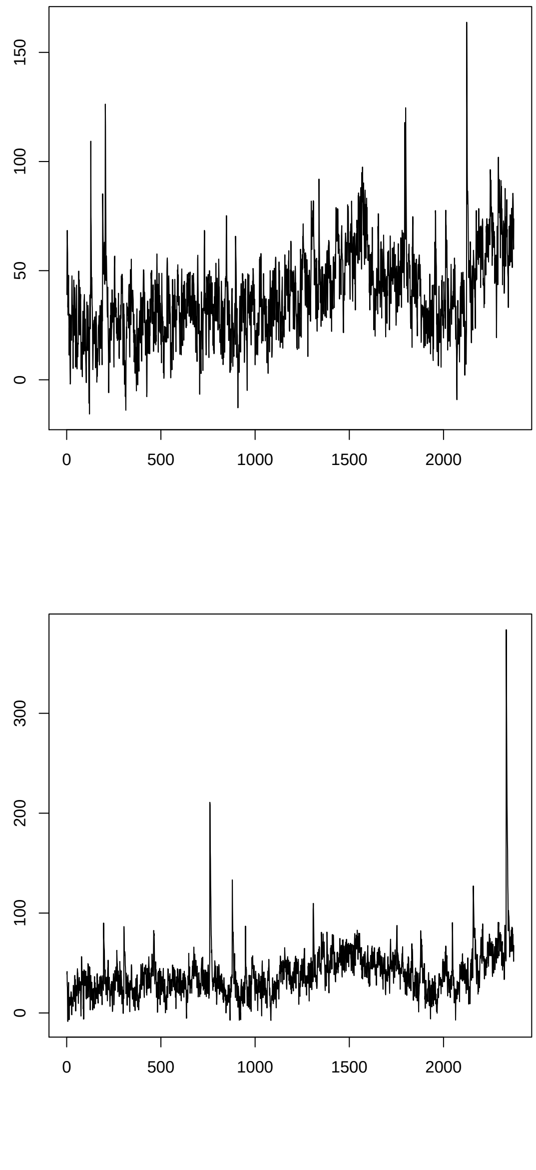


Figure 7.11: Simulation of prices for Model 7.8 GPD

7.5 Ex post goodness of fit

To be able to compare the model for the log-prices with our other models we will use two ex-post goodness-of-fit-criteria which we define as

$$
\begin{aligned}
GoF1(m_i) &= \sum_{t=2}^{n} (P_t - \mathbb{E}_{m_i}[P_t \mid \mathcal{F}_{t-1}])^2 \\
GoF2(m_i) &= \sum_{t=2}^{n} abs(P_t - \mathbb{E}_{m_i}[P_t \mid \mathcal{F}_{t-1}]),
\end{aligned}
$$

where m_i is model i with parameters θ_i.

7.5.1 GoF for model 7.1

The GoF for Model 7.1 is

$$
\begin{aligned}
\mathbb{E}[P_t \mid \mathcal{F}_{t-1}] &= \mathbb{E}[f_t + S_t \mid \mathcal{F}_{t-1}] = f_t + \mathbb{E}[S_t \mid S_{t-1}] \\
&= f_t + \mathbb{E}[\lambda(S_{t-1}(1-\kappa) + \sigma_t Z_t) \mid S_{t-1}] \\
&\quad + \mathbb{E}[(1-\lambda)(S_{t-1}(1-\kappa) + \mu_J + (\sigma_t^2 + \sigma_J^2)^{1/2} Z_t) \mid S_{t-1}] \\
&= f_t + S_{t-1}(1-\kappa) + \lambda\mu_J + \lambda \underbrace{\mathbb{E}[\sigma_t Z_t \mid S_{t-1}]}_{=0*} \\
&\quad + (1-\lambda) \underbrace{\mathbb{E}[(\sigma_t^2 + \sigma_J^2)^{1/2} Z_t \mid S_{t-1}]}_{=0*} \\
&= f_t + S_{t-1}(1-\kappa) + \lambda\mu_J.
\end{aligned}
$$

* Since σ_t is known to the filtration S_{t-1} and Z_t is $\mathcal{N}(0,1)$

7.5.2 GoF for model 7.6

The GoF for Model 7.6 is

$$
\begin{aligned}
\mathbb{E}[P_t \mid \mathcal{F}_{t-1}] &= f_{1,t} + \mathbb{E}[e^{f_{2,t}+S_t} \mid \mathcal{F}_{t-1}] = e^{f_{2,t}}\mathbb{E}[e^{S_t} \mid S_{t-1}] \\
&= f_{1,t} + e^{f_{2,t}}\mathbb{E}[e^{\lambda(S_{t-1}(1-\kappa)+\sigma_t Z_t)+(1-\lambda)(S_{t-1}(1-\kappa)+\mu_J+(\sigma_t^2+\sigma_J^2)^{1/2} Z_t)} \mid S_{t-1}] \\
&= f_{1,t} + e^{f_{2,t}+S_{t-1}(1-\kappa)+\lambda\mu_J}\mathbb{E}[e^{\lambda\sigma_t Z_t+(1-\lambda)(\sigma_t^2+\sigma_J^2)^{1/2} Z_t} \mid S_{t-1}] \\
&= f_{1,t} + e^{f_{2,t}+S_{t-1}(1-\kappa)+\lambda\mu_J}\mathbb{E}[e^{(\lambda\sigma_t+(1-\lambda)(\sigma_t^2+\sigma_J^2)^{1/2})Z_t} \mid S_{t-1}] \\
&= f_{1,t} + e^{f_{2,t}+S_{t-1}(1-\kappa)+\lambda\mu_J+\frac{1}{2}(\lambda\sigma_t+(1-\lambda)(\sigma_t^2+\sigma_J^2)^{1/2})^2}.
\end{aligned}
$$

We used the fact that $\mathbb{E}[e^{c \cdot Z_t}] = e^{\frac{1}{2}c^2}$ when Z_t is i.i.d $\mathcal{N}(0,1)$ and that σ_t is known to the filtration S_{t-1}.

7.5.3 Result GoF test

We used the exact same procedure to get the GoF for the other models. All the results from the GoF tests can be viewed in Table 7.7.

Table 7.7: GoF values

Model	GoF1 value	GoF2 value
Model 7.1	308348	13033
Model 7.2	298539	12940
Model 7.3	312133	13053
Model 7.4	286149	13583
Model 7.5	285746	12536
Model 7.6	353598	13865
Model 7.7	300051	13716
Model 7.8	295320	12770

We notice that the result from the ex post GoF criteria for our first 5 models is quite different from the Schwarz criterion. The ex post GoF criteria is mainly for testing how good prices can be forecasted. Jumps and GARCH behavior does not pay off in this criteria. Model 7.5 performs good in this test as expected because the trend is much better modeled here compared to in Model 7.1. We notice the same difference when comparing Model 7.7 with Model 7.6, especially in the GoF2 value. Model 7.8 performs best in the test and the models for the spot price outperforms the models for the logarithm of the spot price, however the difference is not that large.

7.6 Distributional characteristics

To check if the models have the same distributional characteristics as the original data. We simulated 1000 paths for each model using the estimated parameters. The result can be found in Table 7.8. Model 7.3 is not stationary and is therefore not in this test.

Table 7.8: Distributional characteristics (st.dev of the estimates between parentheses)

Model	Mean	St.dev	Skewness	Kurtosis
Original data	37.80	20.02	2.712	22.30
Model 7.1	36.68 (1.760)	19.53 (1.224)	0.4158 (0.8130)	8.403 (6.744)
Model 7.2	37.87 (1.204)	21.48 (1.046)	0.4421 (0.2237)	5.230 (0.7324)
Model 7.4	37.74 (0.7425)	19.92 (0.5130)	-0.04787 (0.0697)	2.901 (0.1075)
Model 7.5	37.16 (0.7412)	19.64 (1.685)	0.4737 (0.6208)	6.821 (6.827)
Model 7.6	38.17 (0.9842)	19.00 (5.098)	2.347 (3.301)	31.03 (105.2)
Model 7.7	37.50 (0.5775)	18.84 (1.977)	1.772 (1.757)	15.89 (58.90)
Model 7.8	39.58 (1.224)	22.38 (5.941)	2.517 (3.093)	33.08 (59.98)

The models for the log prices and Model 7.8 captures the skewness and kurtosis much better than the others. Interesting to notice is the high st.dev for the estimation of kurtosis for the log price models and Model 7.8, which means that the kurtosis value varies a lot. The mean and st.dev are very similar for all models and the original data except Model 7.8 which seems to overestimate the mean and st.dev a little bit, this might be due that the estimation is done in several steps.

7.7 Summary model investigation

First we conclude from Section 7.2 that the best model was Model 7.1 which incorporates jumps and GARCH(1,1) behavior. For modeling the seasonality dummy variables for weekdays and holidays was the best option, monthly dummy variables did not improve our model. We also notice that Model 7.5, where the moving average is removed outperforms Model 7.1. It is especially the parameter κ, the mean reversion, that differs in value when comparing Model 7.1 and Model 7.5. This is due to that in Model 7.1 the prices drifts away more from the expected trend than in Model 7.5. If we would like to simulate future paths we need to have an opinion about how the future trend might be. Since we do not know if our opinion on the future trend will be right, we should probably chose a κ value lower than in the Model 7.5.

One problem when simulating Model 7.1 is that negative prices sometimes occur. The reason for the negative values is that the model models all jumps in the data in the same way, but in reality most of the negative jumps occur directly after a positive jump. This is a feature that the model does not capture. However by modeling the log prices instead, we do not get any negative spikes anymore. Another option is to only model the positive jumps as we did in the extreme value approach. In Section 7.5 an ex post goodness of fit test for the models were made, Model 7.8 performed the best. Further on in Section 7.6 we tested if the models share the same distribution characteristics as the original data when simulating them. Here the log price models and Model 7.8 performed the best. The overall conclusion is that Model 7.8 performs best.

8 Futures

In this chapter the valuation of electricity forwards/futures is discussed. There are two general approaches to price electricity forwards/futures. The first one uses the spot price as an underlying and defines the future price as the expected value under the market consistent pricing measure. This approach is used by Lucia and Schwartz in [13], Cartea and Figueroa in [7] and by Schmidt in [15]. The other approach is to model the future curve directly without considering the underlying spot price, instead ideas from interest rate theory is used. Koekebakker and Ollmar [12] modeled the forward/future curve on Nord Pool using the Heath-Jarrow-Morton approach. In this thesis, only the first approach will be discussed.

The price of a derivative is equal to the expected pay-off under Q. For a future expiring at time T the price is obtained as the expected value of the spot price at expiry under an equivalent Q-martingale measure, conditional on the information set available up to time t, that is

$$F(t,T) = \mathbb{E}_t^{\mathbb{Q}}(S(T) \mid \mathcal{F}_t).$$

The futures actually traded in electricity markets are not futures on a single spot rate. Instead, they offer electricity for a certain period of length Δ. More precisely, the future offers delivery of electricity in the period $[T, T+\Delta]$, with the value

$$\sum_{t_i \in [T, T+\Delta]} S_{t_i},$$

where $t_i \in [T, T+\Delta]$ refers to the trading days in the period under consideration. We will set $t_i - t_{i-1} = \delta$ and approximate the sum by an integral

$$\sum_{t_i \in [T, T+\Delta]} S_{t_i} \approx \frac{1}{\delta} \int_T^{T+\Delta} S_u du.$$

The expectation must be calculated under an equivalent $\mathbb{Q}$-martingale measure. In a complete market this measure is unique, ensuring only one arbitrage free price of the future. But the electricity market is an incomplete market, therefore this measure is not unique. There are several different approaches how to overcome the problem of selecting the pricing measure $\mathbb{Q}$. One way is to have a model under the pricing measure $\mathbb{Q}$ and estimate the parameters from derivative prices directly. Another approach is to use the parameters under $\mathbb{P}$ and then add a risk premium to adjust for the price difference, that is

$$F(t,T) = \mathbb{E}_t^{\mathbb{Q}}(S(T) \mid \mathcal{F}_t) = \mathbb{E}_t^{\mathbb{P}}(S(T) \mid \mathcal{F}_t) + \pi(t, T, T+\Delta)$$

Forward prices are determined by supply and demand, and are not necessarily forecasts of electricity prices. If more consumers want to buy protection than producers are willing to sell it, π should be positive and the other way around. Therefore the risk premium will be different for different maturities.

Example 8.1. *We will look at the Vasicek model to see how the pricing of futures work*

$$dS_t = \kappa(\theta_t - S_t)dt + \sigma dW_t, \tag{8.1}$$

where W_t is a Brownian motion.

The price of the future is given by the expectation under the risk-neutral martingale measure $\mathbb{Q}$

$$\delta \times F(t, T, T+\Delta) = \mathbb{E}_t^{\mathbb{Q}}\Big(\int_T^{T+\Delta} S_u du\Big).$$

From Section 2.2 we know that Equation 8.1 has the following solution.

$$S_u = S_t e^{-\kappa(u-t)} + \theta(1 - e^{\kappa(u-t)}) + \sigma \int_t^u e^{-\kappa(u-s)} dW_s.$$

Hence we get

$$\mathbb{E}_t^{\mathbb{Q}}\Big(\int_T^{T+\Delta} S_u du\Big) = \int_T^{T+\Delta} \mathbb{E}_t^{\mathbb{Q}}(S_u) du = \theta\Delta + \frac{\theta - S_t}{\kappa}(e^{-\kappa(T+\Delta-t)} - e^{-\kappa(T-t)}).$$

9 Conclusions

In this thesis several models for electricity spot prices in the German market were tested. We concluded that the data has very strong weekly seasonality, this is due to that the market is very sensitive to the load. Modeling monthly seasonality did not improve our model. The trend is a very important factor on the German market and research on understanding the factors that drives the trend would be very interesting. We also concluded that there is a strong GARCH(1,1) behavior on the German market but modeling the GARCH behavior does only pay off in one of our model selection criteria. Modeling of the jumps is an interesting topic, we saw that the extreme value approach models the upward jumps very good. A future research topic is to also model the decline after the jumps in a good way. Another interesting question is which quantitative measures that can be used to compare models. We used three different measures here and they does not give the same result. Pricing of derivatives is the main application for a spot market model and therefore it would be interesting to compare how good the models can price derivatives.

References

[1] Akaike, H (1974). *A new look at the statistical model identification*, IEEE Transactions on Automatic Control 19, 716−723.

[2] Arneric, J., Babic Z. and Skrabic B. (2007). *Maximization of the likelihood function in financial time series models*, International Scientific Conference Contemporary Challenges of Theory and Practice in Economics.

[3] Berndt, E.K., Hall, B.H., Hall, R.E. and Hausman, J.A. (1974). *Estimation and inference in non-linear structural models*, Annals of Economic and Social Measurement 4, 653−665.

[4] Bollerslev, T.(1986). *Generalized autoregressive conditional heteroscedasticity*, Journal of Econometrics 31, 307−327.

[5] Brigo D. and Mercurio F. (2001). *Interest-Rate Models: Theory and Practice*, Springer.

[6] Burger, M., Klar, B., Müller, A. and Schindlmayr, G. (2004). *A spot market model for pricing derivatives in electricity markets*, Quantitative Finance 4, 109−122.

[7] Cartea, A. and Figueroa, M. G. (2005). *Pricing in electricity markets: A mean reverting jump diffusion model with seasonality*, Applied Mathematical Finance 12, 313−335.

[8] Engle, R. F. (1982) *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation*, Econometrica 50, 987−1008.

[9] Escribano, A. Peña, J. I. and Villaplana, P. (2002). *Modeling electricity prices: International Evidence* Working paper no. 02-27, Economic series 08, Universidad Carlos III de Madrid, Departamento de Economia.

[10] Geman, H. and Roncoroni, A. (2006). *Understanding the fine structure of electricity prices*, Journal of Business 79, 1225-1261.

[11] Knittel, C. R. and Roberts, M. R. (2001). *An empirical examination of deregulatedelectricity prices* POWER working paper no. PWP-087.

[12] Koekebakker, S. and Ollmar, F. (2005). *Forward curve dynamics in the Nordic electricity market* Managerial Finance, 31 (6), 73−94.

[13] Lucia, J. and Schwartz, E. (2002). *Electricity prices and power derivatives: Evidence for the Nordic Power Exchange*, Review of Derivatives Research 5, 5−50.

[14] Nordel (2007). *Annual statistics* www.nordel.org.

[15] Schmidt, T. (2007). *Modelling energy markets with extreme spikes*, Working paper.

[16] Shreve, S.E. (2004) *Stochastic Calculus for Finance II: Continous-Time Models*, Springer.

[17] Schwarz, G. (1996) *Estimating the dimension of a model*, Annals of Statistics 6, 461−464.

[18] Straumann, D. (2005). *Estimation in conditionally heteroscedastic time series models* Lecture Notes in Statistics, Springer.

[19] VDEW (2007). *Stromzahlen, Der deutsche Strom- und Energimarkt auf einen blick*, www.strom.de.

[20] Vasicek, O. (1977). *An equilibrium characterization of the term structure*, Journal of Financial Economics 5, 177–188.