Modelling electricity prices with Ornstein-Uhlenbeck processes

Alain Angeralides and Ottmar Cronie

September 18, 2006

Abstract

This Masters thesis deals with aspects of the modelling of deperiodized and deseasonalized electricity spot prices by means of infinitely divisible Ornstein-Uhlenbeck processes. Further, models for periodic (weekly) and seasonal components of the electricity prices are investigated.

Acknowledgements

We would like to deeply thank our supervisor Patrik Albin for his support, guidance, and fantastic ability to inspire us throughout the process of writing this thesis. Our discussions have given us the knowledge, development, and ambition needed in order to reach the level we have finally accomplished. Patrik has always taken his time for us and been there to support us whenever we have needed him. He really has made a big impact on our lives. One should not forget all the good laughs he has given us as well.

We would also deeply like to thank all those who have helped us through out the process of understanding the material deeper as well as explaining the theory of stochastics on a higher level. Without these instruments it would have been an impossibility to overcome some of the obstacles confronted with along the way.

A big special thanks to Anders Muszta who has helped us a lot in understanding much of the theory. He has always been an inspiration for us and a tremendous support.

We would also like to thank Erik Brodin for fruitful discussions on this thesis and all the support he has given us. Thank you Mattias Sundén, Jan Lennartsson, Cătălin Stărică, and others who have supported and helped us in this process. Thank you very much Mikael Jednell, Plusenergi, for always being able to answer our questions when needed! Another big thanks to Östkraft AB for supporting us with such an informative website.

Alain's personal thanks: I would mostly like to thank my mom, Elisabeth, from the bottom of my heart for all the support she has given me to reach this level. A big thanks to my brother Angelo who always supports me when I need him. A thank you also goes out to my dad, Yorgo, for all his support. And a big thanks also to my uncle Elie who always supports me when I need him. Special thanks to my friends Toni Demir and Sargon Yousif who always support me when I need them. I want to thank Daniel Ahlberg for all the support he has given me during my studies. Thank you Jonas Persson, VD Östkraft AB, for explaining the concept of "lack of effect" and for giving me the job :) A big special thanks to Ottmar Cronie for inspiring and fruitful collaboration. Ottmar has always supported me

when I needed him in all different ways. Good luck in your PhD studies in the mathematical statistic department with the most wonderful colleagues.

Ottmar's personal thanks: First of all I would like to thank my family for being the wonderful people they are. No people on this earth mean more to me and I certainly would never have been able to finish this thesis without their infinite support, consideration, patience, and love. Ivan, mam, and dad, thank you! I would also like to thank all my friends (who are all too many to be mentioned here and too important to be left out) for supporting me through this process and being as understanding and loving as you all have been. A big thank you to all people who have inspired me along the way. Thank you Michael Gurebo, Jan-Erik Etander, Aflahia Radford, Kit Nip, and David Thompson for making me realize how much I like mathematics and for inspiring me.

A big special thanks to Alain for being the wonderful, inspiring, and hard working friend you are. Thank you for your patience, good collaboration, and happy mood. You really have taught we what hard work is all about.

Thank you all!!

Contents

1 Introduction

On the first of January 1996 the Swedish electricity market was deregulated, meaning that the prices of electricity are determined by a market with free price establishment and competition among different sellers and buyers. At the same time the Swedish state, via Svenska Kraftnät (*SvK*) which is the administrator of the national electrical grid, became part owner in the Norwegian power exchange *Nord Pool*. Hereby the Swedish and Norwegian electricity markets became one common free market which the other Nordic countries, with the exception of Island, later took part in.

When the Swedish market was still regulated, i.e. before the year 1996, the consumers were forced to buy electricity from one distributor. The electricity companies had the right to sell their electricity only to fixed areas. Thus the consumers had no opportunity to influence the electricity prices. After the deregulation of the electricity market, the purpose was to give the consumers alternatives among different electricity distributors and with free competition. The aim of the deregulation was to create a more effective price establishment.

We have decided to model the electricity spot prices of the years 1996-2001, which will be discussed more in detail further on. The reason that we left out the modelling of the years 2002-2003 was that these years were extremely dry. The extreme situation, which occurred during the autumn and winter 2002/2003 (week 21 year 2002 until week 12 year 2003), is very rare and years as dry as these have not occurred for the past 20 years. During that period the flow of water in the reservoirs was very low, having lower electricity production as a consequence. This made the electricity prices rise to the extreme levels observed in the extremely high spikes of this period. The total flow of water to the reservoirs during this period, of the year 2003, was 20 TWh lower than the average of the totals of the years 1980-2003 of the same period. In the period in question there were moments where the temperature became extremely low. This in turn increased the demand of electricity and thereby led to the increase of electricity prices.

Since we decided not to include the years 2002-2003 it was not possible to include the year 2004. The reason for this being that problems such as change of level and effect of dependency will take place in the data if we would have linked the years 1996-2001 with the year 2004. In the beginning of 2005, emission allowances were introduced and they play a big role in today's electricity market. They have nearly doubled the electricity prices since introduced. Emission allowances will be discussed later in the report. Thus, when having access to only a small set of historical data of emission allowances, this will not be enough for modelling. This is the reason why we left the years 2005-2006 aside.

The mean reverting nature of spot prices and the existence of jumps or spikes in the prices are two distinctive features which are present. We know that in stock markets, prices are allowed to evolve freely but this is not true for electricity prices. When facing abnormal market conditions, price spreads are observed in the short run, but in the long run supply will be adjusted and prices will move towards, and gravitate around the level of production costs.

2 Electricity market

2.1 Nord Pool

The Nordic electricity market, Nord Pool, is the power exchange servicing the Nordic countries of Europe, including Norway, Denmark, Sweden, and Finland, who were all at some point deregulated and connected through interconnectors (electric lines and electrical plants and meters for the transfer of electricity to or from the transmission system). Anyone who is connected to any part of a national network in these countries can, in principle, buy (or sell) electricity from (or to) anyone else connected to the network.

The Norwegian power market was deregulated in 1991 and it took two more years, i.e. 1993, before Nord Pool was established. When introduced it was the first of its kind and was actually the first power exchange in the world. It quickly became a world wide trading place with many participants. The Nordic countries produce a total of approximately 380-390 TWH per year. Sweden accounts for approximately 40%, Norway 30%, Finland 20%, and Denmark 10%.

Storage of electricity is very difficult, sometimes impossible, or otherwise often very expensive. Hence, markets must be kept in balance on a second-by-second basis. Hydroelectricity, which can be considered a storable form of electricity, is traded in large quantities in Nord Pool. Water can be stored in reservoirs and lakes. The reservoirs fill up during the spring and at the beginning of the summer, caused by the melting of the snow, and at the beginning of the autumn, caused by rain.

Approximately 50% of the electricity produced in Sweden comes from hydroelectricity, see [12]. According to [7] the optimal use of the stored water over a year time basis is to start using it at the end of march. The storability makes hydroelectricity more similar to other commodities. It lowers the height of spikes and reduces the price volatility.

Electricity that is consumed and produced must be in balance at every instant, which is achieved by balance control. In every country there exists a TSO, transmission system operator, who is responsible for the task of maintaining this balance and for maintaining the national grid. The TSO arranges the transmission of electricity from power stations to consumers through a network of power lines (national grids, regional networks and local networks).

Nord Pool includes the following markets:

- *Elspot* is the spot price market. The spot price is the price set daily on the spot market. It is determined by supply and demand. This price is called system price or spot price and is an average price for the whole power market (Sweden, Norway, Finland, Denmark). The producers and power trading companies, i.e. the different participants, inform daily, before 12:00, how much they are going to supply (producers) and buy (power trading companies) the coming day. The daily price is determined by the intersection between the demand curve and the supply curve. This spot price is therefore

an average price, an index, which is used as a reference of the electricity market. See Figure 1 below.

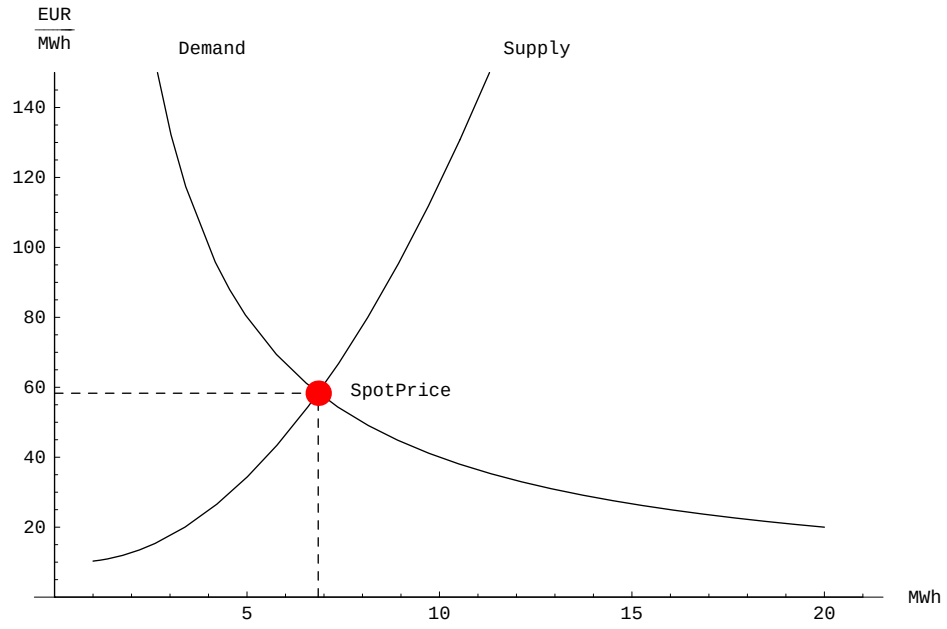


Figure 1: The x-axis is the turnover. The values of the y-axis do not represent real electricity price level.

- *Eltermin* is a purely financial market involving futures and forward markets for speculation and hedging of power contracts (with no physical delivery) over periods such as days, weeks, months, and years. A contract is an agreement which involves that you are obliged, in the future, to buy electricity for a fixed price, previously decided. This price reflects what the market believes the future spot price will be.

- *Elbas* is a market for adjusting imbalances for short term physical delivery. Participants who have earlier taken their positions on Elspot can adjust their positions up to 2 hours prior to delivery.

- *Eloption* trade consists of European-style settlement power options with forward contracts as their underlying instruments. It is a financial market for risk management and for forecasting future income and costs related to trade in electricity contracts.

- *Currency Exchange Rates* has its daily Nord Pool published exchange rates for each SEK (Swedish Krone), FIM(Finnish Mark), DKK (Danish Krone), and NOK (Norwegian Krone) versus the Euro.

2.2 Emission allowances

From December 1 through 11, 1997, more than 160 nations met in Kyoto, Japan, to negotiate binding limitations on greenhouse gases for the developed nations. The outcome of the meeting was the Kyoto protocol, in which the developed nations agreed to limit their greenhouse gas emissions, relative to the levels emitted in 1990. Countries that ratify this protocol commit to reduce their emissions of carbon dioxide together with five other greenhouse gases. If they desire to increase their emissions of these gases then they will be committed of buying emission allowances.

Producers must try to minimize the emissions of greenhouse gases. Trading with emission allowances started at the beginning of January 2005 and comprises all European Union (EU) countries. The system of emission allowance was created in order for the EU to fulfill its undertaking of the Kyoto protocol. During the first phase, years 2005-2007, trading concerns only carbon dioxide while during the second phase, years 2008-2012, there are possibilities to include other greenhouse gases as well.

The system works in the following way. A roof is set for how much emissions of carbon dioxide is allowed during a specific period. Those countries that are members of the so called EU-15 must, on average, reduce their emissions by 8% compared to the levels of the year 1990. The remaining member countries, which joined the EU in may 2004, have their own individual goals that are adjusted to the Kyoto protocol. Every country divides its own, by the EU individually allotted, number of emission allowances between the producing companies of the country. Naturally this concerns only companies which emit carbon dioxide. This division follows so called National Allocation Plans (NAP) which the countries themselves establish. The National Allocation Plans must first be approved by the EU commission. There are at the moment 750 electricity producers in Sweden that are connected to this system. All country is responsible to set up a particular register where all transactions that are carried out must be registered. In Sweden it is the power authorities that are responsible for this register, which is called *Svenskt utsläppsrättssystem, SUS* (the Swedish emission allowance system).

Those who own one emission allowance have the right to emit one ton of carbon dioxide during a stated trading period. During the first trading period, at least 95% of the emission allowances have to be divided freely to those connected to the system. The Swedish state has decided that all Swedish companies shall get their emission allowances for free. The companies that emit less carbon dioxide than the quantity of emission allowances that they have received, can either save the emission allowances for later use, or sell the surplus to other companies. On the other hand, companies that emit more carbon dioxide must buy more emission allowances.

The price of the emission allowances is determined by supply and demand. A very important parameter in the pricing of the emission allowances is the total amount of emission allowances handed out by the EU. A greater supply of emission allowances means lower prices and vice versa. Other factors that affect

the price are, among others, weather, the economical development, and political (in)security.

2.3 Electricity certificates

In may 2003 the so called electricity certificates were introduced. It is a market based system created to support the production of renewable electricity. It is more expensive to produce electricity from renewable energy than from the traditional production types used in Sweden. Thus, the government created this electricity certificate system in order to help the producers. Power sources that are entitled to electricity certificates are wind power, sun power, geothermal power, bio-fuel, wave power, certain water power types, and peat.

Producers of renewable electricity are allotted a certificate for every Megawatt hour (1 MWh=1000 kWh) of electricity they produce and they can then sell them to electricity trading companies. The price of electricity certificates is set by the market according to the principle of supply and demand. The compensation producers receive, by selling these electricity certificates, is supposed to cover the extra cost for producing the renewable electricity, compared to the traditional way of producing electricity. Furthermore, this compensation should also stimulate the producers to invest further in renewable forms of electricity.

2.4 Factors affecting electricity prices

The most important factors that affect the electricity spot price are supply and demand.

Demand: Temperature and business cycles are the major factors that affect the demand of electricity. When the temperature is very low, the consumer's demand will increase which will lead to an increase of electricity prices which is due to the lack of produced electricity. Temperature affects how much water there is available in the reservoirs. Depending on how much water is available in the reservoirs will affect the electricity price. Hence, the more water the lower the electricity price. When it comes to business cycles, consumers demand more electricity during economic booms than during economic recessions.

Political decisions: Political decisions regarding taxes and fees also play a role. Another example of politically related factors are emission allowances.

Supply: Emission allowances for carbon dioxide play a big role today since they affect the marginal costs for carbon condense production. During a large part of a normal water level year the Danish and Finish carbon condense production prices determine the marginal cost on the Nordic market. The more water available for water power production, the less carbon condensing will be a price determining factor. The marginal costs for a carbon condense power plant varies with time, partly because of varying fuel prices and partly

because of varying prices of the emission allowances. Today's prices of emission allowances have nearly doubled the marginal costs for carbon condense power plants, both in the Nordic countries and in the remainder of Europe. Looking at the years 1996-2006 below one can clearly see the effect of the emission allowances. See Figure 2 below.

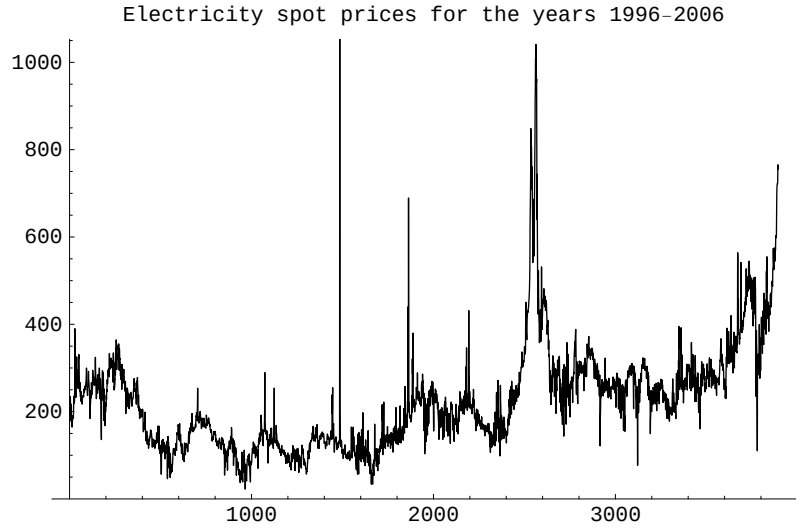


Figure 2: Electricity prices 1996-2006.

When it comes to lack of production of electricity, import of electricity from other countries will take place. When having simultaneously a low demand of electricity and a high production of electricity, export of electricity will take place.

3 Ornstein-Uhlenbeck processes

As has already been mentioned our aim is to see how *Ornstein-Uhlenbeck* processes (OU processes) can be used to model electricity spot prices. We here introduce some theory needed in order to define OU processes, together with some instruments used when fitting these processes to data.

The motivation for choosing OU processes to fit to electricity spot prices is, among others, their particular jumping behaviour. They move up entirely by jumps and then tail off exponentially, where this decay takes place between the jumps of the process. Another motivation is that they have a so called mean reverting behaviour, meaning that they oscillate around a process mean value. This corresponds well with the features of electricity spot prices. OU processes are driven by Lévy processes, which are processes with the ability of jumping. The OU processes jump according to the jumps of the driving Lévy processes.

3.1 Building Lévy processes

Definition 3.1 (Lévy process) *A stochastic process* $\{X(t)\}_{t\geq 0}$ *(*$\{X(t)\}_{t\in\mathbb{R}}$*) is a* Lévy process *if the following conditions hold:*

1. *X is right continuous and has left limits with probability one;*
2. *$X(0) = 0$ with probability one;*
3. *X has independent increments, i.e. $X(t_1) - X(t_0), \ldots, X(t_n) - X(t_{n-1})$ are independent for all $t_0 \leq t_1 \leq \ldots \leq t_n$;*
4. *X has stationary increments, i.e. $X(t) - X(s) \stackrel{D}{=} X(t-s)$ for all $s \leq t$;*
5. *X is stochastically continuous, i.e. $\lim_{h\to 0} \mathbf{P}\{|X(t+h) - X(t)| \geq \epsilon\} = 0$, for $\epsilon > 0$, for all t.*

If X is a Lévy processes, then each process value $X(t)$ can be represented as a sum of n independent identically distributed (IID) random variables, whose distribution is that of $X(t/n)$. In other words, $X(t)$ can be divided into n IID parts. Distributions with this property are called infinitely divisible distributions:

Definition 3.2 (Infinite divisibility) *A random variable Y is* infinitely divisible *if, for each $n \in \mathbb{N}$, and some IID random variables $Y_1, \ldots, Y_n$, we have $Y \stackrel{D}{=} Y_1 + \ldots + Y_n$.*

The random variables $\{Y_k\}_{k=1}^n$ that divide Y must have common characteristic function (CHF)

$$\varphi_{Y_1}(u) \equiv \mathbf{E}\left[\mathrm{e}^{iuY_1}\right] = \mathbf{E}\left[\mathrm{e}^{iu(Y_1+\cdots+Y_n)}\right]^{1/n} = \mathbf{E}\left[\mathrm{e}^{iuY}\right]^{1/n} \equiv \varphi_Y(u)^{1/n} \quad \text{for } u \in \mathbb{R}.$$

Conversely, to each infinitely divisible distribution D, there exists a Lévy process X such that $X(1) \stackrel{D}{=} D$.

Theorem 3.3 (Lévy-Khintchine) *A random variable Y is infinitely divisible if and only if there exists a unique so called triplet (σ^2, ν, γ), with*

$$\begin{cases} \sigma^2 \geq 0 & \text{the Gaussian coefficient,} \\ \nu & \text{the Lévy measure on } \mathbb{R} \text{ satisfying } \nu(\{0\}) = 0 \text{ and } \int_{\mathbb{R}} y^2 \wedge 1 \, d\nu(y) < \infty, \\ \gamma \in \mathbb{R} & \text{the drift coefficient,} \end{cases}$$

such that the CHF satisfies the Lévy-Khintchine formula

$$\phi_Y(u) = \exp\left\{ i\gamma u - \frac{1}{2}\sigma^2 u^2 + \int_{-\infty}^{\infty} (\mathrm{e}^{iux} - 1 - iux1_{\{|x|<1\}})\nu(dx) \right\} \quad \text{for } u \in \mathbb{R}.$$

To clerify the role of the Lévy measure actually does, we give an alternative equivalent (as it turns out) definition of it:

Definition 3.4 (Lévy measure) *The* Lévy measure *of a Lévy process X is the measure ν on $\mathbb{R}$ defined by*

$$\nu(A) = \mathbf{E}\left[\#\left\{ t \in [0,1] : X(t) - X(t^-) \in A \setminus \{0\} \right\} \right] \quad \text{for } A \subseteq \mathbb{R}.$$

Definition 3.5 (Compound Poisson process) *Let $\{N(t)\}_{t\geq 0}$ be a Poisson process with intensity λ and $\{Y_i\}_{i-1}^{\infty}$ IID random variables with cumulative probability distribution function (CDF) F that are independent of N. A* compound Poisson process *with intensity $\lambda > 0$ and jump size distribution F is given by*

$$X(t) = \sum_{i=1}^{N(t)} Y_i \quad \text{for } t \geq 0.$$

Then $X(t)$ has CHF

$$\phi_{X(t)}(u) = \exp\left\{ t\lambda \int_{-\infty}^{\infty} \left(\mathrm{e}^{iux} - 1\right) F(dx) \right\} \quad \text{for } u \in \mathbb{R}.$$

Note that, for a compound Poisson process X, the Lévy measure of $X(1)$ is given by $\nu(A) = \lambda \int_A dF$.

From the Lévy-Khintchine formula, we see that an infinitely divisible random variable consists of three independent parts: a constant, a zero-mean normal distributed part, and a part of compound Poisson distribution type (that more exactly is the limit of a compound Poisson distribution, but for simplicity we will call the compound Poisson part).

Now consider a Lévy process built up by the infinitely divisible random variable in question. The Lévy measure ν dictates how the jumps occur. Jumps of sizes in the set A occur according to a compound Poisson process with intensity parameter $\nu(A)$. Hence, each Lévy process is the sum of a Brownian motion with drift (continuous part), plus an independent jump process of compound Poisson processes type.

3.2 Specific Lévy processes

In this section we introduce the Lévy processes that feature in this thesis (besides Brownian motion and the Poisson process, the definitions of which are elementary). These processes all have zero Gaussian component $\sigma^2 = 0$.

Definition 3.6 (Inverse Gaussian Process) *The* Inverse Gaussian distribution *(IG) with parameters $a, b > 0$ and $\mu \in \mathbb{R}$, is the infinitely divisible distribution with CHF*

$$\phi_{\mathrm{IG}}(u; a, b, \mu) = \exp\left\{-a(\sqrt{-2iu + b^2} - b) + iu\mu\right\}.$$

The probability density function (PDF) of the IG distribution is given by

$$f_{\mathrm{IG}}(x; a, b, \mu) = \frac{a}{\sqrt{2\pi}\,(x-\mu)^{3/2}} \exp\left\{ab - \frac{a^2}{2(x-\mu)} - \frac{b^2(x-\mu)}{2}\right\} \quad \text{for } x > \mu.$$

An IG process *is a Lévy process $\{X(t)\}_{t\geq 0}$ such that $X(1)$ is $\mathrm{IG}(a, b, \mu)$ distributed. Then $X(t) - X(s) \stackrel{D}{=} \mathrm{IG}(a(t-s), b, \mu(t-s))$ for $0 \leq s \leq t$.*

Definition 3.7 (Normal Inverse Gaussian Process) *The* Normal Inverse Gaussian distribution, $\mathrm{NIG}(\alpha, \beta, \delta, \mu)$, *with parameters $\alpha > 0$, $-\alpha < \beta < \alpha$, and $\delta > 0$, is the infinitely divisible distribution with CHF*

$$\phi_{\mathrm{NIG}}(u; \alpha, \beta, \delta, \mu) = \exp\left\{-\delta(\sqrt{\alpha^2 - (\beta + iu)^2} - \sqrt{\alpha^2 - \beta^2}) + iu\mu\right\}.$$

The PDF of the NIG distribution is given by

$$f_{\mathrm{NIG}}(x; \alpha, \beta, \delta, \mu) = \frac{\alpha}{\pi} \exp\left\{\delta\sqrt{\alpha^2 - \beta^2} + \beta(x-\mu)\right\} \frac{K_1(\alpha\sqrt{\delta^2 + (x-\mu)^2})}{\sqrt{1 + (x-\mu)^2/\delta^2}} \quad \text{for } x \in \mathbb{R},$$

where K is the modified Bessel function of the third kind

$$K_\lambda(x) = \frac{1}{2}\int_0^\infty y^{\lambda-1}\, \mathrm{e}^{-x(y-1/y)/2}\, dy \quad \text{for } x > 0.$$

A NIG process *is a Lévy process $\{X(t)\}_{t\geq 0}$ such that $X(1)$ is $\mathrm{NIG}(\alpha, \beta, \delta, \mu)$ distributed. Then $X(t) - X(s) \stackrel{D}{=} \mathrm{NIG}(\alpha, \beta, \delta(t-s), \mu(t-s))$ for $0 \leq s \leq t$.*

Definition 3.8 (Generalized Inverse Gaussian Process) *The* Generalized Inverse Gaussian distribution, *$GIG(\lambda, a, b, \mu)$, with parameters $a, b > 0$ and $\mu, \lambda \in \mathbb{R}$ is the infinitely divisible distribution with CHF*

$$\phi_{\mathrm{GIG}}(u; \lambda, a, b, \mu) = \frac{1}{K_\lambda(ab)} \left(1 - \frac{2iu}{b^2}\right)^{\lambda/2} K_\lambda(ab\sqrt{1 - 2iub^{-2}})\, \mathrm{e}^{iu\mu}.$$

The PDF of the GIG distribution is given by

$$f_{\mathrm{GIG}}(x; \lambda, a, b, \mu) = \frac{(b/a)^\lambda (x-\mu)^{\lambda-1}}{2K_\lambda(ab)} \exp\left\{-\frac{a^2}{2(x-\mu)} - \frac{b^2(x-\mu)}{2}\right\} \quad \text{for } x > \mu.$$

Special cases of the GIG distribution includes the IG distribution.

A GIG process *is a Lévy process $\{X(t)\}_{t\geq 0}$ such that $X(1)$ is GIG distributed.*

Definition 3.9 (Gamma Process) *The* Gamma distribution, *Gamma*(a, b, μ), *with parameters* $a, b > 0$ *and* $\mu \in \mathbb{R}$ *is the infinitely divisible distribution with PDF*

$$f_{\text{Gamma}}(x; a, b, \mu) = \frac{b^a}{\Gamma(a)}(x-\mu)^{a-1}\exp(-(x-\mu)b) \quad \text{for } x > \mu.$$

The Gamma process *is a Lévy process* $\{X(t)\}_{t\geq 0}$ *such that* $X(1)$ *is Gamma*(a, b, μ) *distributed. Then* $X(t) - X(s) \stackrel{D}{=} \text{Gamma}(a(t-s), b, \mu(t-s))$ *for* $0 \leq s \leq t$.

Definition 3.10 (Variance Gamma Process) *The* Variance Gamma distribution, *VG*(a, b, c, μ) *with parameters* $a, b, c > 0$ *and* $\mu \in \mathbb{R}$ *is the infinitely divisible distribution with CHF*

$$\phi_{\text{VG}}(u; a, b, c, \mu) = \left(\frac{bc}{bc + (c-b)iu + u^2}\right)^a \mathrm{e}^{iu\mu},$$

and is distributed as the difference between two independent Gamma distributed random variables. The PDF of the VG distribution is given by

$$\begin{aligned} & f_{\text{VG}(a,b,c,\mu)}(x) \\ &= \frac{\sqrt{b+c}}{\Gamma(a)\sqrt{\pi\, bc(x-\mu)}}\, \mathrm{e}^{(b-c)(x-\mu)/(2bc)} \left(\frac{x-\mu}{b+c}\right)^a K_{a-1/2}\left(\frac{(b+c)(x-\mu)}{2bc}\right) 1_{\{x>\mu\}} \\ &\quad + \frac{\sqrt{b+c}}{\Gamma(a)\sqrt{\pi\, bc(\mu-x)}}\, \mathrm{e}^{(b-c)(x-\mu)/(2bc)} \left(\frac{\mu-x}{b+c}\right)^a K_{a-1/2}\left(-\frac{(b+c)(x-\mu)}{2bc}\right) 1_{\{x<\mu\}} \end{aligned}$$

for $x \in \mathbb{R}$.

A VG process *is a Lévy process* $\{X(t)\}_{t\geq 0}$ *such that* $X(1)$ *is* $\text{VG}(a, b, c, \mu)$ *distributed. Then* $X(t) - X(s) \stackrel{D}{=} \text{VG}(a(t-s), b, c, \mu(t-s))$ *for* $0 \leq s \leq t$. *Further, the VG process is the difference of two independent Gamma processes.*

Definition 3.11 (Meixner Process) *The* Meixner distribution, *Meixner*$(\alpha, \beta, \delta, \mu)$, *with parameters* $\alpha > 0$, $-\pi < \beta < \pi$, $\delta > 0$ *and* $\mu \in \mathbb{R}$ *is the infinitely divisible distribution with CHF*

$$\phi_{Meixner}(u; \alpha, \beta, \delta, \mu) = \left(\frac{\cos(\beta/2)}{\cosh((\alpha u - i\beta)/2)}\right)^{2\delta} \mathrm{e}^{iu\mu}.$$

The PDF of the Meixner distribution is given by

$$f_{Meixner}(x; \alpha, \beta, \delta, \mu) = \frac{(2\cos(\beta/2))^{2\delta}}{2\alpha\pi\Gamma(2\delta)}\, \mathrm{e}^{\beta(x-\mu)/\alpha} \left|\Gamma\left(\delta + \frac{i(x-\mu)}{\alpha}\right)\right|^2 \quad \text{for } x \in \mathbb{R}.$$

A Meixner process *is a Lévy process* $\{X(t)\}_{t\geq 0}$ *such that* $X(1)$ *is Meixner*$(\alpha, \beta, \delta, \mu)$ *distributed. Then* $X(t) - X(s) \stackrel{D}{=} \text{Meixner}(\alpha, \beta, \delta(t-s), \mu(t-s))$ *for* $0 \leq s \leq t$.

Definition 3.12 (Generalized Hyperbolic Process) *The* Generalized Hyperbolic distribution, *GH*$(\alpha, \beta, \delta, \nu, \mu)$, *with parameters*

$$\delta \geq 0, \quad |\beta| < \alpha \quad if \quad \nu > 0,$$

$$\delta > 0, \quad |\beta| < \alpha \quad if \quad \nu = 0,$$

$$\delta > 0, \quad |\beta| \leq \alpha \quad if \quad \nu > 0,$$

is the infinitely divisible distribution with CHF

$$\phi_{\mathrm{GH}}(u; \alpha, \beta, \delta, \nu, \mu) = \left(\frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta + iu)^2}\right)^{\nu/2} \frac{K_\nu\left(\delta\sqrt{\alpha^2 - (\beta + iu)^2}\right)}{K_\nu\left(\delta\sqrt{\alpha^2 - \beta^2}\right)} \mathrm{e}^{iu\mu}.$$

The PDF of the GH distribution is given by

$$f_{\mathrm{GH}}(x; \alpha, \beta, \delta, \nu, \mu)$$
$$= \frac{(\alpha^2 - \beta^2)^{\nu/2}(\delta^2 + (x-\mu)^2)^{\nu/2 - 1/4}}{\sqrt{2\pi}\,\alpha^{\nu-1/2}\delta^\nu K_\nu(\delta\sqrt{\alpha^2 - \beta^2})} K_{\nu-1/2}(\alpha\sqrt{\delta^2 + (x-\mu)^2})\,\mathrm{e}^{\beta(x-\mu)} \quad for\ x \in \mathbb{R}.$$

Special cases of the GH distribution includes the VG distribution as well as the NIG distribution.

The GH process *is a Lévy process* $\{X(t)\}_{t\geq 0}$ *such that* $X(1)$ *is GH distributed. The GH process is a*

Some distribution have no known closed form probability density functions. For such distributions one can find an approximation of the PDF by numerical Fourier invers transform of the CHF in question.

Proposition 3.13 (Inverse transform of CHF) *If a random variable* ξ *has an integrable CHF* ϕ_ξ*, then it has a continuous PDF given by*

$$f_\xi(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-itx}\,\phi_\xi(t)\,dt \quad for\ x > 0.$$

3.3 Ornstein-Uhlenbeck processes

Definition 3.14 (Ornstein-Uhlenbeck process) *Let* $\{L(t)\}_{t\in\mathbb{R}}$ *be a Lévy process, which is usually refered to as the* Background Driving Lévy Process (BDLP). *For a constant* $\lambda > 0$*, the* Ornstein-Uhlenbeck process *(OU process) is given by*

$$Z(t) = \int_{-\infty}^{t} \mathrm{e}^{-\lambda(t-s)}\,dL(s) \quad for \quad t \in \mathbb{R}.$$

If Z is an OU process, then it will be a stationary process, that is, it will have translation invariant finite dimensional distributions. This is so because

$$\{Z(t+h)\}_{t\in\mathbb{R}} = \left\{\int_{-\infty}^{t+h} \mathrm{e}^{-\lambda(t+h-s)}\, dL(s)\right\}_{t\in\mathbb{R}} = \left\{\int_{-\infty}^{t} \mathrm{e}^{-\lambda(t-s)}\, dL(s+h)\right\}_{t\in\mathbb{R}} \stackrel{D}{=} \{Z(t)\}_{t\in\mathbb{R}}$$

for $h \in \mathbb{R}$, by the stationary increments of L.

Definition 3.15 (AR Process) *Given an $p \in \mathbb{N}$, a process $\{X(t)\}_{t\in\mathbb{Z}}$ is called an* Autoregressive process *with index p,* AR(p) process*, if it is a stationary process, that for some constants $a_0 = 1$ and $a_1, \ldots, a_p \in \mathbb{R}$ satisfies*

$$\begin{cases} e(t) = \displaystyle\sum_{k=0}^{p} a_k X(t-k) = X(t) + \sum_{k=1}^{p} a_k X(t-k) \\ e(t) \text{ is independent of } X(t-1), X(t-2), \ldots \end{cases} \qquad \text{for } t \in \mathbb{Z}.$$

It turns out that there are restrictions on the coefficients $a_1, \ldots, a_p$ for an AR(p) process. For example, for an AR(1) process, we must have $|a_1| < 1$ (see [3]).

As for an OU process

$$Z(t) = \mathrm{e}^{-\lambda(t-t_0)}\, Z(t_0) + \int_{t_0}^{t} \mathrm{e}^{-\lambda(t-s)}\, dL(s) \quad \text{for } t \geq t_0,$$

so that $Z(t)$ only depends on the history $\{Z(t)\}_{t\leq t_0}$ through the value of $Z(t_0)$, $Z(t)$ is a Markov process. In addition, taking $t_0 = t-1$, this shows that an OU process is an AR(1) process with parameter $a_1 = \mathrm{e}^{-\lambda} \in (0,1)$ and IID noise

$$\{e(t)\}_{t\in\mathbb{Z}} = \left\{\int_{t-1}^{t} \mathrm{e}^{-\lambda(t-s)}\, dL(s)\right\}_{t\in\mathbb{Z}},$$

which is infinitely divisible (by approximating the intregral with Riemann sums).

The OU process is the solution to the Langevin stochastic differential equation

$$dZ(t) = -\lambda Z(t)dt + dL(t) \quad \text{for } t > 0, \quad Z(0) = \int_{-\infty}^{0} \mathrm{e}^{-\lambda s}\, dL(s)$$

(see [2], page 29). This equation explains the so called *mean-reverting* feature of the OU process, i.e., its tendency to move back to its equilibrium state.

3.4 Approximated infinitely divisible distribution

We will build an infinitely divisible distribution that is styled to fit with the features of our data, which we call the *Approximated Infinitely Divisible Distribution* (AIDD). This distribution consists of a sum of Gaussian random variable and an

independent sum of independent rescaled Poisson distributed random variables with different intensities

$$\mathrm{N}(\mu,\sigma^2)+\sum_{i=1}^{n} b_i \mathrm{Po}(a_i), \quad \mu\in\mathbb{R},\ \sigma^2>0,\ a_1,\ldots,a_n>0,\ b_1,\ldots,b_n\in\mathbb{R}. \tag{3.1}$$

We will now explain how the AIDD can be used to approximate any infinitely divisible distribution D: By the Lévy-Khintchine formula, the CHF ϕ_D of D is given by

$$\phi_D(u)=\exp\left\{i\gamma u-\frac{\sigma^2u^2}{2}+\int_{-\infty}^{\infty}(\mathrm{e}^{iux}-1-iux1_{\{|x|<1\}})\nu(dx)\right\},$$

where $\gamma\in\mathbb{R}$ and $\sigma^2\geq 0$ are constants and ν the Lévy measure. Now we make an approximation $\hat{\nu}$ of the Lévy measure ν by point masses

$$\frac{d\hat{\nu}(x)}{dx}=\sum_{k=1}^{n} a_k\delta(x-b_k).$$

This gives

$$\begin{aligned}
\phi_D(u) &\approx \exp\left\{i\gamma u-\frac{\sigma^2u^2}{2}+\int_{-\infty}^{\infty}(\mathrm{e}^{iux}-1-iux1_{\{|x|<1\}})\hat{\nu}(dx)\right\}\\
&= \exp\left\{i\gamma u-\frac{\sigma^2u^2}{2}+\sum_{k=1}^{n}(\mathrm{e}^{iub_k}-1-iub_k1_{\{|b_k|<1\}})a_k\right\}\\
&= \exp\left\{i\mu u-\frac{\sigma^2u^2}{2}+\sum_{k=1}^{n}(\mathrm{e}^{iub_k}-1)a_k\right\},
\end{aligned}$$

where $\mu=\gamma-a_k\sum_{k=1}^{n}b_k1_{\{|b_k|<1\}}$. This in turn is the CHF of the distribution (3.1).

As any Lévy measure can be approximated arbitrarily well by point masses, in the sense that approximated CHF can be made to converge to the true one, we have shown that any infinitely divisible distribution can be approximated arbitrarily well by an AIDD distribution.

4 Initial data analysis

We started by analyzing the electricity prices from the date 1996-01-01 until 2006-08-29, see Figure 3 below.

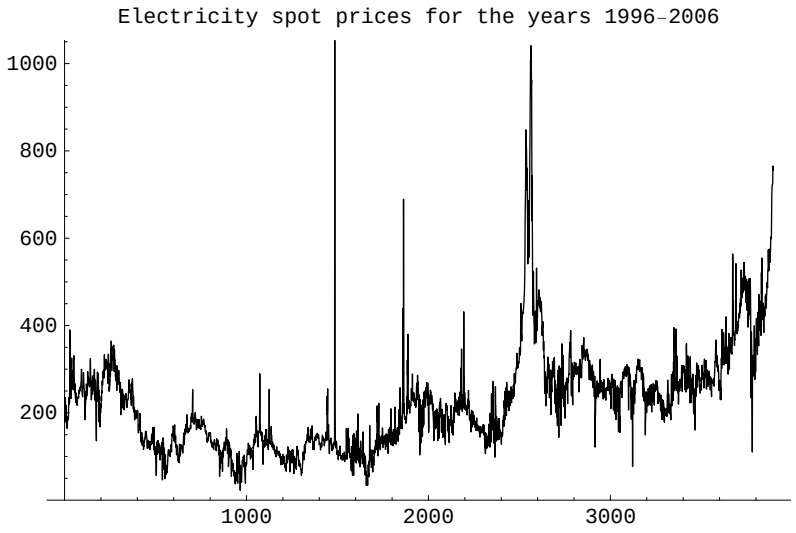


Figure 3: Electricity prices 1996-2006.

As we mentioned earlier we did not include the years 2002-2006 in our data set when modelling the electricity prices. Since the extreme case of the winter of 2002-2003 is very unusual, we decided to ignore this abnormality and thus eliminate the years 2002-2003. We also did not want to get into modelling the years 2005-2006 because of emission allowances, as there is not enough data available for the trade of emission allowances to model their behaviour. Stationary models for current electricity prices will not behave representatively until the trade of emission allowances has stabilized. The year 2004 was also eliminated since it is squeezed between those years. Thus our decision was to model the electricity spot prices for the years 1996-2001.

As can be seen in Figure 4 below, there is a strong mean reversion of the spot price present, so that it oscillates around a mean level with periods of high volatility characterizing the behaviour of the spikes observed in the market.

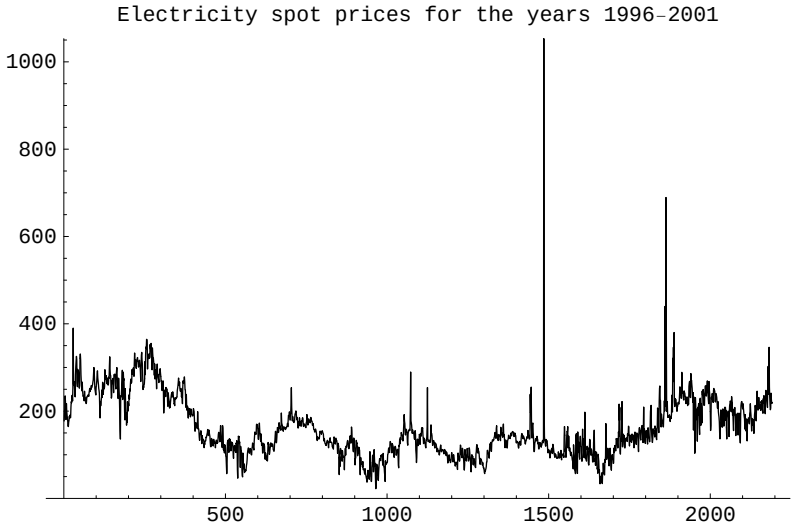


Figure 4: Electricity prices 1996-2001.

It seems cleat the above electricity spot prices are not normally distributed, as the thin tails of the normal distribution will not be able to capture the extreme values that appear in the data. This conclusion is supported by the normal-plot in Figure 5 below.

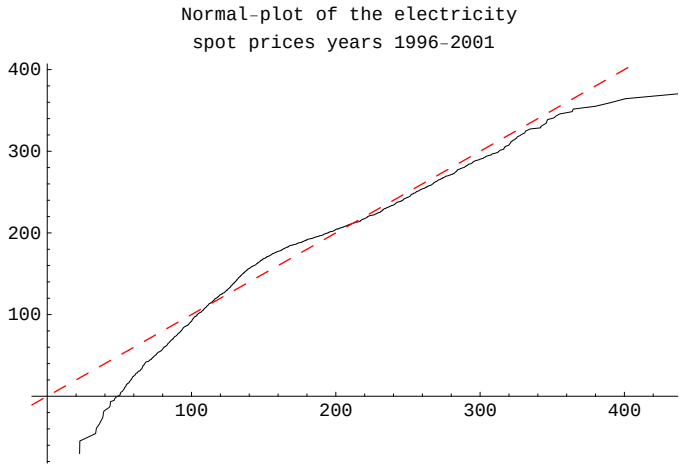


Figure 5: Normal-plot of the electricity prices 1996-2001.

Another fact that speaks against the normal distribution is the high skewness and kurtosis coefficients of our data, that are now described.

Definition 4.1 (Skewness and kurtosis) *The* skewness *of a random variable X, which measures the degree to which the distribution of that random variable is asymmetric, is defined as*

$$\frac{\mathbf{E}\big[(X - \mathbf{E}[X])^3\big]}{(\mathbf{Var}[X])^{3/2}}.$$

The kurtosis*, which is a way of measuring the heaviness of the tail of the distribution, is defined as*

$$\frac{\mathbf{E}\big[(X - \mathbf{E}[X])^4\big]}{(\mathbf{Var}[X])^2}.$$

Table 1 below displays the mean, variance and standard deviation of our data set, together with the skewness and kurtosis.

Mean	162.532
Variance	4782.27
Standard Deviation	69.1539
Skewness	1.73072
Kurtosis	16.4609

Table 1: Stylized facts of the data.

As one can see, the data exhibits very high spikes. These are naturally not appreciated neither by consumers nor by the electricity traders. This phenomenon is known as *lack of effect* in the electricity market, which (not suprisingly) means that there is a lack of produced effect on the electricity market!

As can be seen, on the 24th of January 2000, the electricity price jumped from a level of approximately 100-150 Kr/MWh to 1000 Kr/MWh. On the 5th and 6th of February 2001 the price was nearly as high as well. These high spikes were due to lack of effect .

We now give an example of how the eletricity market may work in the situation of lack of effect: It is January and very cold. Since some nuclear reactors are put out of function, people responsible for controlling the balance of the electricity production of Sweden know that all production resources must be in use in order to be on the safe side the coming morning, in order to avoid the phenomenon of extremely high price spikes, due to demand exceeding supply.

There is a risk that lack of effect will take place the coming morning. Everybody tries to adjust to this situation: The producers try to produce as much electricity as posible, and the consumers, who directly or indirectly, feel the day-market price try to minimize their use of electricity.

In order to achieve a balance between the buying and selling of electricity orders, the day-market prices for those critical hours in the morning will be very high. When we come near the morning hours it turns out that there is not enough electricity produced, i.e., the demand will exceed the available production capacity. The price of the balance adjustments will rise and transactions carried out will have prices that are much higher than the daily market price set the day before. Thus, further stakes will take place in order for more production of electricity to occur.

In [13], a report which was sent to SvK on the 29th of August 2000, we can see a situation where a load of about 28000 MW may take place. A scenario which, according to the report, happens approximately once every 10 years. These high spikes jump up and stay at that level for a couple of hours. The incident that took

place on the 24th of January 2000, which had a very high load level, is shown in Table 2 below, for different time intervals.

Time	MWh/h	Time	MWh/h	Time	MWh/h
06-07	22285	11-12	23775	16-17	23736
07-08	24292	12-13	23492	17-18	23829
08-09	24595	13-14	23229	18-19	23316
09-10	24328	14-15	23229		
10-11	24192	15-16	23501		

Table 2: Load on the 24th of January year 2000, according to Nord Pool.

In the table we can see that the highest load level was between 08-09, a time that is usual for top levels. During the hours 07-11, we had a maximum variation of 403 MWh/h between the top load level, between the hours 08-09, and the minimum load level, between 10-11. Thus, as a conclusion, according to SvK:s data, the electricity consumption scenario given above, if one uses the interval [top load, top load-500 MW], has a probability of occurring in a ratio of 4 hours per 10 years i.e. a probability of 0.0046.

Comments on lack of effect and how SvK acts in such situations:

- There is not enough statistical data on lack of effect scenarios to be able to predict these situations.
- When calculating the risk of lack of effect one must take into consideration that the market reacts differently on situations depending on which price prevails.
- Low electricity prices and high delivery safety can hardly be combined on a deregulated electricity market.
- Closing down production units will increase the risk of lack of effect. In these situations one must be able to cover the lack of energy needed in order to fulfill the consumer's demand.
- There is a clear risk that the perception of the market is that SvK takes the responsibility of supplying enough running power reactors, regardless of the markets behaviour. If this is the case, the market clearly does not seem to take the responsibility of supplying these power plants themselves.

5 Random numbers

There are a couple of general methods to simulate random numbers with a desired distribution. As these general methods usually are quite slow and/or unstable, due to involving numerical integration and numerical equation solving, it is desirable to try to find special algorithms that work for a specific distribution.

5.1 Inverse method

Let X be a random variable with continuous CDF F. Then

$$\mathbf{P}[F(X) < x] = \mathbf{P}[X < F^{-1}(x)] = F(F^{-1}(x)) = x \quad \text{for } x \in (0,1)$$

(where F^{-1} denoes a generalized inverse, should the inverse not exist). Hence the random variable $F(X)$ is uniformly distributed over $(0,1)$. It follows that we can generate a random number X with the CDF F by first generating a uniformly distributed random number U, and then applying the inverse CDF $X = F^{-1}(U)$. See also [10].

However, often there is no closed form expression available for F^{-1}. In that situation one has to rely on numerical methods to solve the equation $F(X) = U$ for X.

5.2 Rejection method

Another method that is a bit quicker than the inverse method, but unfortunately also a bit less stable, is the *rejection method*, see [10]. Here the idea is to look at the ratio between two distributions, where we already know how to generate random number from one of them.

Let g be a PDF which we do how to generate random numbers from, and let f be the PDF we want to generate random numbers from. Then this can be achieved if the following properties hold:

- g and f are defined on the same domain.
- We have
$$c \equiv \sup_x \frac{f(x)}{g(x)} < \infty,$$
where the supremum is taken over all x in the domain of g and f. For most of the distributions considered in this thesis, it is a good choice to pick the PDF g with polynomially decaying tails, for numerical reasons.

The rejection method works as follows:

1. Generate a random number x_g with PDF g, and an independent random number u uniformly distributed over $(0,1)$.

2. If
$$u \leq \frac{f(x_g)}{c\, g(x_g)},$$
then go to Step 3, otherwise return to Step 1.

3. Accept x_g as a random number with PDF function f.

5.3 IG and NIG random numbers

In [4] the following algorithm for generating NIG$(\alpha, \beta, \delta, \mu)$ distributed random numbers is suggested, making use of an IG distributed random number:

1. Generate an IG$(\delta^2, \alpha^2 - \beta^2, 0)$ distributed random number in the following way:
   (a) Let V be a $\chi^2(1)$ distributed random number.
   (b) Let $\xi = \delta/\sqrt{\alpha^2 - \beta^2}$.
   (c) Put
$$W = \xi + \frac{\xi^2 V}{2\delta^2} + \frac{\xi}{2\delta^2}\sqrt{4\xi\delta^2 V + \xi^2 V^2}.$$
   (d) Let U be a uniformly distributed random number over $(0, 1)$ independent of V.
   (e) Put
$$Z = \begin{cases} W & \text{if } U \leq \xi/W, \\ \xi^2/W & \text{if } U > \xi/W. \end{cases}$$

2. Let Y be a N$(0, 1)$ distributed random number independent of Z.

3. Then the random variable $X = \mu + \beta Z + \sqrt{Z}Y$ is NIG$(\alpha, \beta, \delta, \mu)$ distributed.

5.4 GH random numbers

In [1] the following algoritm for generating GH distributed random numbers is suggested, taking off from a GIG distributed random number:

1. Let Z come be a GIG$(v, \delta, \sqrt{\alpha^2 - \beta^2}, 0)$ distributed random number.

2. Let Y be a N$(0, 1)$ distributed random number independent of Z.

3. Then the random variable $X = \mu + \beta Z + \sqrt{Z}Y$ is GH$(\alpha, \beta, \delta, v, \mu)$ distributed.

Note that the transformation of GIG to GH is the same as that of IG to NIG. GIG random numbers we generate using the rejection method.

5.5 Numerical integration

When generating random number one usually needs to make use of the CDF F of the distribution of the random numbers in question. Having access only to the PDF f of the distribution, and no closed formula for the CDF, one needs to employ numerical integration, i.e. to find a numerical approximation of

$$F(x) = \int_{-\infty}^{x} f(y)\, dy. \tag{5.1}$$

But how do we numerically approximate $-\infty$?

The answer, in the case of the distributions considered here, is given by a study of the tails: All distributions considered in this thesis, except for the lighter tailed Gaussian distribution, are so called *semi-heavy tailed* distributions. This means that their tails behave according to

$$f(x) \sim \begin{cases} C_- |x|^{\rho_-} \mathrm{e}^{-\eta_- |x|} & \text{as } x \to -\infty \\ C_+ |x|^{\rho_+} \mathrm{e}^{-\eta_+ |x|} & \text{as } x \to \infty \end{cases} \tag{5.2}$$

for some constants $\rho_-, \rho_+ \in \mathbb{R}$ and $C_-, C_+, \eta_-, \eta_+ > 0$.

The property (5.2) is very useful, as we can use it in our numerical integration procedure (5.1) to determine what finite interval $[x^*, x]$ to integrate over, instead of $[-\infty, x]$, for example by selecting x^* such that $f(x^*) = 10^{-6}$, ensuring that the contribution of the omitted integration domain $[-\infty, [x^*]$ to the integral is neglible.

As an example, in the case of the NIG distribution, it can readily be established from the form of the PDF that

$$f(x) \sim \sqrt{\frac{\alpha}{2\pi}} \exp\left\{\delta\sqrt{\alpha^2 - \beta^2}\right\} \frac{\delta}{|x|^{3/2}}\, \mathrm{e}^{-(\alpha+\beta)|x|} \quad \text{as } x \to -\infty.$$

And for this distribution we selected $x^* = \mu - 50/(\alpha + \beta)$.

6 Fitting OU processes to electricity spot prices

When fitting Ornstein Uhlenbeck processes to data, we used that the OU process is a special type of AR(1) process. This means that if our observed data $\{x_t\}_{t=1}^n$ is modeled by an OU process, then the sequence $e(t) = x_t - \alpha\, x_{t-1}$, $t = 1, \ldots, n$, is IID for some $\alpha \in (0, 1)$. Hence we try to find this α, as well as fitting an infinitely divisible distribution to the sequence $\{e(t)\}_{t=1}^n$.

The parameter estimation is done by the maximum likelihood (ML) estimation. In our specific case, the likelihood function takes the form

$$L\big(\{x_t\}_{t=1}^n; \alpha, \Theta\big) = \prod_{i=1}^{n} f\big(x_i - \alpha\, x_{i-1}; \Theta\big),$$

where $f(y; \Theta)$ is the PDF of a parametric infinitely divisible distribution that is consideration.

To evaluate our fit we use the *Kolmogorov-Smirnov* (KS) goodness-of-fit test:

Definition 6.1 (Kolmogorov-Smirnov (KS)) *Given a data set* $\{x_t\}_{t=1}^n$ *with* empirical distribution *function*

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{x_i \le t\}},$$

the KS goodness-of-fit test *of a fitted CDF* $\hat{F}$ *is based on the KS test statistic* KS_n*, also called the* KS distance *between* F_n *and* $\hat{F}$*, given by*

$$\mathrm{KS}_n = \sup_{x \in \mathbb{R}} \big| F_n(x) - \hat{F}(x) \big|.$$

A useful numerical form of KS_n*, to be used in computations, is that*

$$\mathrm{KS}_n = \max_{1 \le i \le n} \max \left\{ \left| \hat{F}(x_{(i)}) - \frac{i-1}{n} \right|, \left| \hat{F}(x_{(i)}) - \frac{i}{n} \right| \right\},$$

where $x_{(1)}, \ldots, x_{(n)}$ *denotes the ordered data set.*

6.1 Fitting GH OU processes

When trying to fit an infinitely divisible distributions to eletricity spot prices, in the fashion above, the only distribution that gave us a reasonably good fit was the GH distribution: The ML estimate of the $\mathrm{GH}(a, b, d, v, \mu)$ parameters are given in Table 3 below, together with the Kolmogorov-Smirnov (KS) distance.

α	a	b	d	v	μ	KS
0.95	0.16688	0.01371	0.33574	3.28605	4.36503	0.0653054

Table 3: Estimated parameters of the GH distribution and the KS distance.

As a general rule, the KS distance should be at most in order for the fit to be accepted for a data set of our size (a point to be discussed later on), so our value 0.065 is way too large. And even if the fit were accepted anyway, the GH-OU process would not necessarily be a good model for the raw electricity prices, as we have not yet checked that $\{x_t - \alpha\, x_{t-1}\}_{t=1}^n$ are independent.

The estimated GH PDF is shown in Figure 6 below together with a histogram of the data set $\{x_t - \alpha\, x_{t-1}\}_{t=1}^n$. Further, the empirical distribution function of the data is plotted together with the estimated GH CDF.

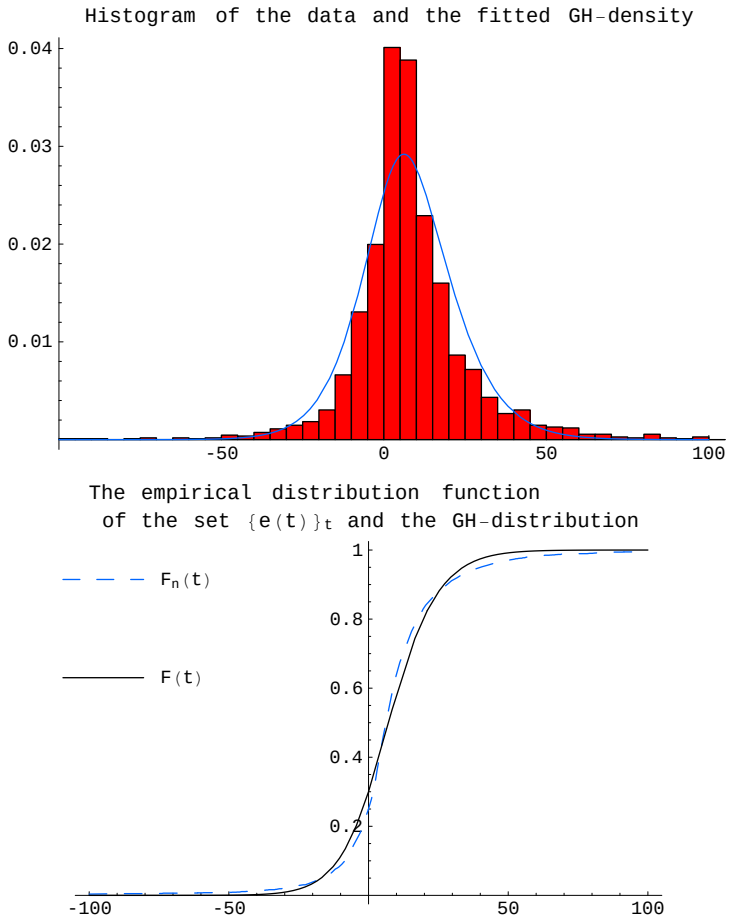


Figure 6: Upper: GH PDF. Lower: Empirical distribution function of $x_t - \alpha\, x_{t-1}$ together with the estimated GH CDF.

6.2 Testing independence

There are various ways to test mutual independence for the members of an observed time series $\{y_i\}_{i=1}^n$.

Auto correlation plots are one common method for testing correlation between $f(y_t)$ and $f(y_{t+k})$, where usually $f(y) = |y|$, or $f(y) = y^i$ for $i = 1, \ldots, m$ and $k = 1, \ldots, n-t$, for some $m \in \mathbb{N}$, i.e., it is a way of investigating if different moments are correlated, see [8].

The *Ljung-Box test* is another way of testing independence, based on correlation methodology. It is a modification of the so called *Portmanteau test*, and is described next.

Definition 6.2 (Sample auto correlation function) *The* auto correlation function *of a stationary stochastic process $\{Y(t)\}_{t\geq 0}$ is defined as*

$$\rho_Y(h) = \frac{\mathbf{Cov}[Y(t), Y(t+h)]}{\sqrt{\mathbf{Var}[Y(t)]\,\mathbf{Var}[Y(t+h)]}} = \frac{\mathbf{Cov}[Y(0), Y(h)]}{\mathbf{Var}[Y(0)]}.$$

Given an observed stationary time series $\{y_i\}_{i=1}^n$, we can estimate the auto correlation function by the sample auto correlation function

$$\hat{\rho}_y(k) = \sum_{t=1}^{n-k} \frac{(y_t - \bar{y})\,(y_{t+k} - \bar{y})}{(n-1)s_y^2} \quad \text{for } k = 0, \ldots, n-1,$$

where $\bar{y}$ is the sample mean and s_y^2 the sample variance.

Definition 6.3 (Ljung-Box (LB)) *Let $\{y_i\}_{i=1}^n$ be an observed time series. In the* Ljung-Box (LB) test *the null hypothesis is that the data are uncorrelated, which is tested against the alternative that the data are not uncorrelated.*

The LB test statistic *is given by*

$$Q_{\mathrm{LB}}(M) = n(n+2) \sum_{k=1}^{M} \frac{\hat{\rho}^2(k)}{n-k},$$

where M is the number of lags being tested, that is a number that has to be choosen by the statistician. Under the null hypothesis $Q_{\mathrm{LB}}(M)$ is χ^2_M distributed, so that the null hypothesis is rejected on the test level β (which we usually take to be 0.05), showing that data are not uncorrelated, if

$$Q_{LB}(M) > \chi^2_{1-\beta,M},$$

where $\chi^2_{1-\beta,M}$ is the $1-\beta$ quantile of the χ^2_M distribution.

The LB test statistic is a weighted sum of squared sample auto correlations for different lags k. Of course, if the data set is independent, then all correlations and thus the sum should be near zero.

The following Figure 7 shows a plot of the time series $\{x_t - \alpha\, x_{t-1}\}_{t=1}^n$ and its sample auto correlation function, with α taken from the GH fit of the previous section.

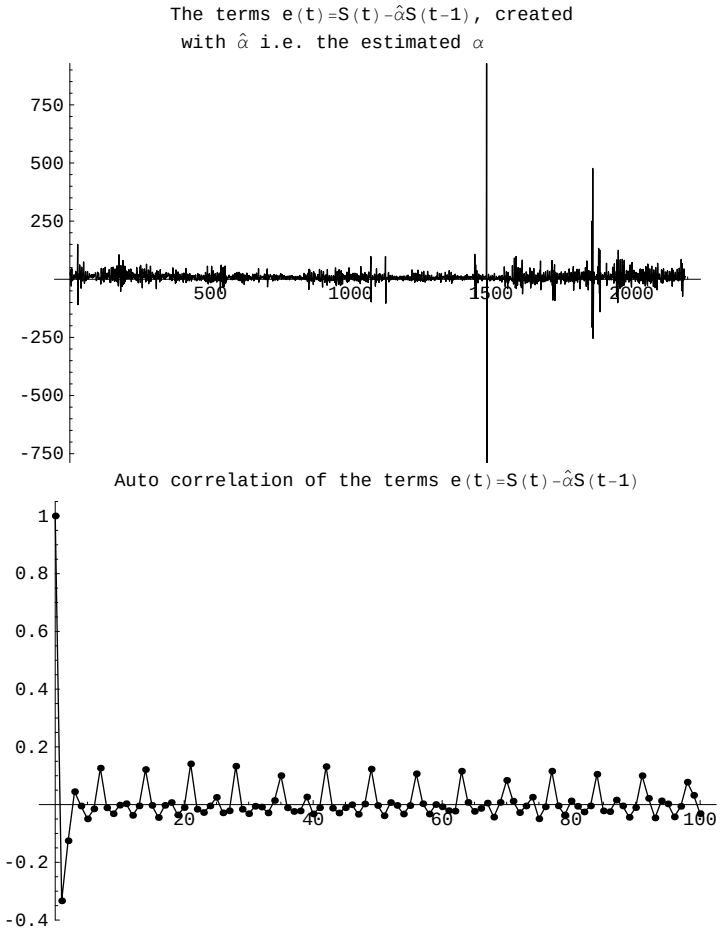


Figure 7: Upper: Plot of $x_t - \alpha\, x_{t-1}$.
Lower: Auto correlation of $x_t - \alpha\, x_{t-1}$.

Note that the auto correlation plot indicates dependence between data.

Table 4 below shows the LB test statistic, calculated for the data set $\{x_t - \alpha\, x_{t-1}\}_{t=1}^{n}$, at lags $M = 1, \ldots, 25$. A comparison is made with the $\chi^2_{0.95,M}$ quantile.

M	1	2	3	4	5
$\chi^2_{M,0.95}$	3.84146	5.99146	7.81473	9.48773	11.0705
$Q_{LB}(M)$	243.815	278.319	282.717	282.77	288.147
M	6	7	8	9	10
$\chi^2_{M,0.95}$	12.5916	14.0671	15.5073	16.919	18.307
$Q_{LB}(M)$	288.636	323.732	324.006	326.217	326.225
M	11	12	13	14	15
$\chi^2_{M,0.95}$	19.6751	21.0261	22.362	23.6848	24.9958
$Q_{LB}(M)$	326.246	329.302	329.347	361.833	361.85

M	16	17	18	19	20
$\chi^2_{M,0.95}$	26.2962	27.5871	28.8693	30.1435	31.4104
$Q_{LB}(M)$	366.277	366.292	366.405	369.241	369.474
M	21	22	23	24	25
$\chi^2_{M,0.95}$	32.6706	33.9244	35.1725	36.415	37.6525
$Q_{LB}(M)$	413.519	414.089	415.752	415.802	417.193

Table 4: Comparison between $\chi^2_{0.95,M}$ quantiles and LB test statistics $Q_{LB}(M)$ on the data set for $M = 1, \ldots, 25$.

As one can see, there is a very clear indication of dependence for each choice of lag M.

As a result of our investigations, we are all anxious to see how simulated GH OU trajectory based on the ML estimated parameters compares with a plot of the observed eletricity spot prices. Such a comparison is made in Figure 8 below.

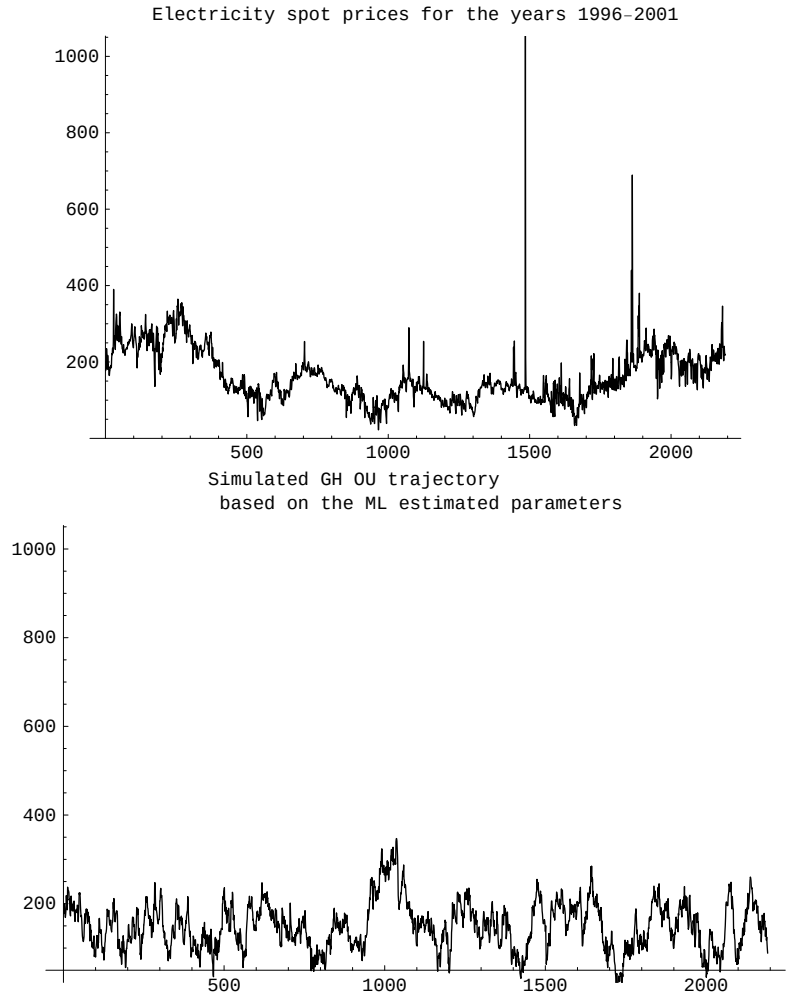


Figure 8: Upper: Electricity spot prices 1996-2001. Lower: Simulated GH OU trajectory.

7 Data filtration

As have been demonstrated in Chapter 7, it is not possible to get a satisfactory fit of OU processes to raw electricity prices.

If one wants to fit an OU process model X to a time series $\{x_t\}_{t=0}^n$, then it has to hold that the differences $\{x_t - \alpha x_{t-1}\}_{t=1}^n$ are independent for a suitable $\alpha \in (0,1)$, see Section 3.1. To get an initial idea of what kind of dependence structure our data posseses, we plot the sample auto correlation function in Figure 9 below.

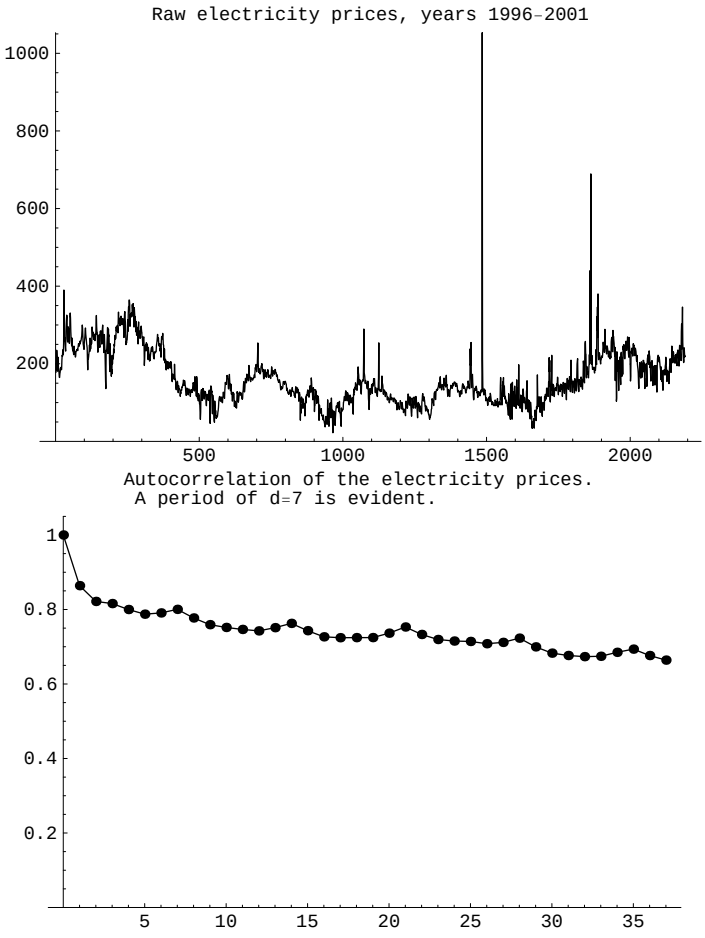


Figure 9: Upper: Raw electricity prices, years 1996-2001. Lower: The sample auto correlation of the raw electricity prices.

At a first glance it is clear that there is a period of $d = 7$ present. Hence, the first thing to do is to remove this period.

In order to filter a data setto remove periodicity from it, different smoothing techniques are needed:

Definition 7.1 (Simple moving average) *Given a number $q \in \mathbb{N}$, the* moving average *of a stochastic process $\{X(t)\}_{t\in\mathbb{N}}$ is given by*

$$m_{sma}(X(t)) = \frac{1}{d}\sum_{j=-q}^{q} X(t+j) \quad \textit{for } q \leq t$$

if the window size $d = 2q+1$ *is odd, and by*

$$m_{sma}(X(t)) = \frac{1}{d}\left(\frac{1}{2}X(t-q) + \sum_{j=-q+1}^{q-1} X(t+j) + \frac{1}{2}X(t+q)\right) \quad \textit{for } q \le t$$

if the window size $d = 2q$ *is even.*

For a finite observed data series $\{x_t\}_{t=0}^{n}$ we have the so called *end-effect* problem, that moving average smoothed data $\{y_t\}_{t=0}^{n-2d}$ will have d points less at the beginning and d points less at the end of the series, as compared with the original data series. This problem can be solved in the following way: If the window size is odd $d = 2q+1$, then the filtered data become

$$\mathrm{m_{sma}}(x_t) = \begin{cases} \dfrac{x_0 + \ldots + x_{2q}}{d} & \text{for} \quad 0 \le t \le q-1 \\ \dfrac{1}{d}\displaystyle\sum_{j=-q}^{q} x_{t+j} & \text{for} \quad q \le t \le n-q \\ \dfrac{x_{n-q} + \ldots + x_n}{d} & \text{for} \quad n-q+1 \le t \le n \end{cases},$$

while if the window size is even $d = 2q$, the filtered data become

$$\mathrm{m_{sma}}(x_t) = \begin{cases} \dfrac{\frac{1}{2}x_0 + x_1 + \ldots + x_{2q-1} + \frac{1}{2}x_{2q}}{d} & \text{for} \quad 0 \le t \le q-1 \\ \dfrac{1}{d}\left(\dfrac{1}{2}x_{t-q} + \displaystyle\sum_{j=-q+1}^{q-1} x_{t+j}\frac{1}{2}x_{t+q}\right) & \text{for} \quad q \le t \le n-q \\ \dfrac{\frac{1}{2}x_{n-2q} + x_{n-2q+1} + \ldots + x_n + \frac{1}{2}x_n}{d} & \text{for} \quad n-q+1 \le t \le n \end{cases}.$$

7.1 Deperiodization

As mentioned, our firsr task is to remove the possible periods. We have taken the following approach to carry this out from [5]:

Method 1 (Period removal) *Let* $\{x_t\}_{t=1}^{n}$ *be a data set that has period* d*. Apply the following steps in order to remove the period:*

1. *Apply the moving average filter to get the filtered data set* $\{m_{sma}(x_t)\}_{t=q+1}^{n-q}$.
2. *Let* $q = \lfloor d/2 \rfloor$*. Compute the mean deviations*

$$w_k = \frac{1}{n-2q}\sum_{j=\lfloor (q+1-k)/d \rfloor}^{\lfloor (n-q-k)/d \rfloor} (x_j - m_{sma}(x_j)) \quad \textit{for } k = 1, \ldots, d.$$

3. The periodical component $\hat{s}_k$ is computed according as

$$\hat{s}_k = w_k - \frac{1}{d}\sum_{i=1}^{d} w_i \quad \textit{for } k = 1, \ldots, d,$$

while $\hat{s}_k = \hat{s}_{k-d}$ for $k > d$.

4. The deperiodized data is computed as $d_t = x_t - \hat{s}_t$ for $t = 1, \ldots, n$.

The previously described end-effect problem, which one with this procedure for removal of period, is resolved in above described manner.

For our data, $\{w_k\}_{k=1}^{d}$, $d = 7$, represents the averages of the week days, as the period is the number of days in a week, so called *intra-week periodicity.*

Figure 10 below shows a plot of the deperiodized data together with the original electricity price data set.

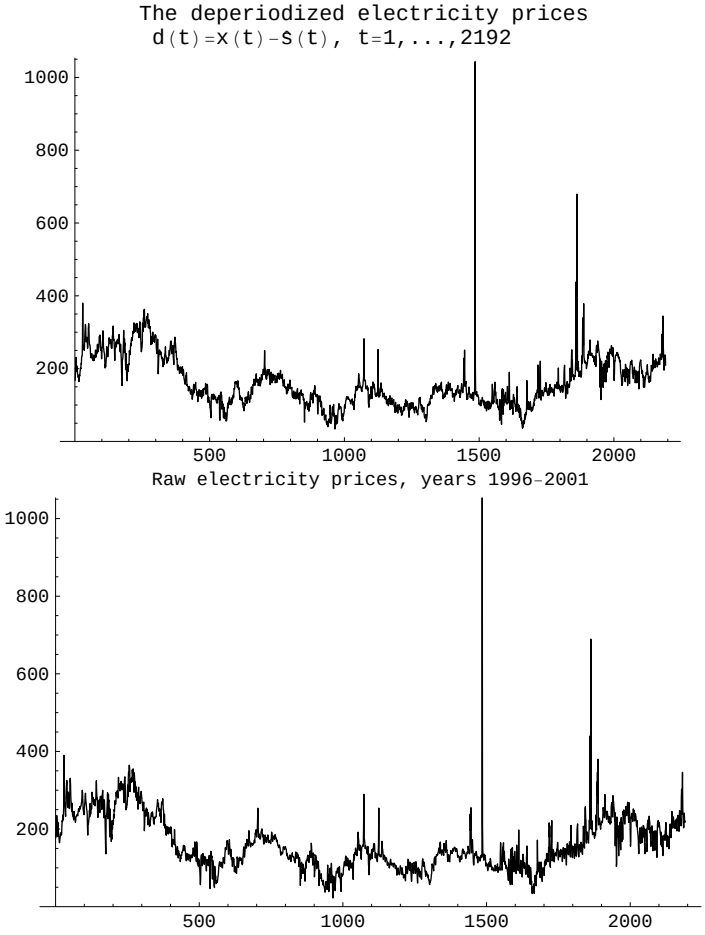


Figure 10: Upper: Deperiodized electricity prices, years 1996-2001. Lower: Raw electricity prices, years 1996-2001.

As one can see from the auto correlation plot of the deperiodized data $\{d_t\}_{t=1}^{n}$ in Figure 11 below, the intra-week periodicity that was previously present in the

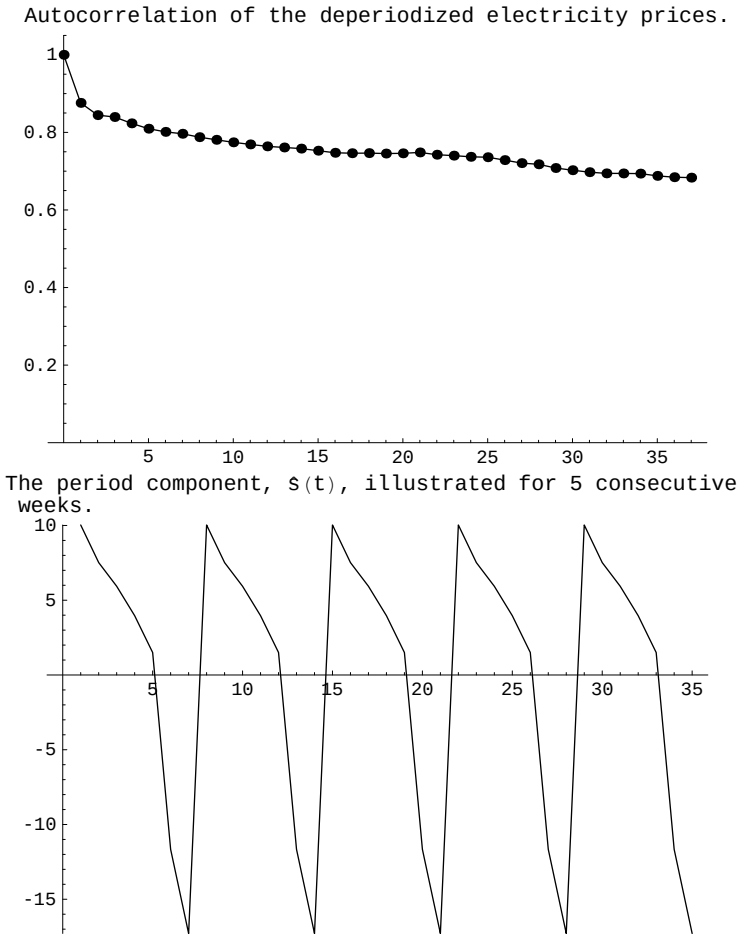


Figure 11: Upper: The auto correlation (for lags $k = 0, \ldots, 37$) of the deperiodized electricity prices.
Lower: The period component series $\{\hat{s}_t\}_{t=1}^{35}$.

data has now been removed. That periodicity is extracted and represented by the sequence $\{\hat{s}_t\}_{t=1}^{n}$.

The next thing to do is to filter the deperiodized data $\{d_t\}_{t=1}^{n}$, in such a way that we end up with two components; a so *seasonal component* $\{s_t\}_{t=1}^{n}$, and a *noise component* $\{y_t\}_{t=1}^{n}$ that can be modeled as an OU process, i.e., that is such that $\{y_t\}$ is stationary and $\{y_t - \alpha\, y_{t-1}\}_{t=1}^{n}$ is an IID sequence for some $\alpha \in (0, 1)$.

7.2 Exponential filtration

The following filter is presented in [6]:

Definition 7.2 (Exponential smoothing) *Given a number $\theta \in (0, 1)$, the* exponential moving average *of a stochastic process $\{X(t)\}_{t\in\mathbb{Z}}$ is given by*

$$m_{\exp}(X(t), \theta) = \sum_{j=0}^{\infty} \theta(1-\theta)^j X(t-j).$$

In our case with n observed data $\{x_t\}_{t=1}^n$, the exponential filtration becomes

$$\mathrm{m}_{\exp}(x_t, \theta) = \sum_{j=0}^{t-1} \theta(1-\theta)^j x_{t-j} \quad \text{for } t = 1, \ldots, n.$$

The advantage of this filter, as compared to other filters which are based on smoothing with fixed window sizes, is that it does not have the end-effect problem.

A drawback on the other hand, is that the first couple of smoothed values depend too strongly on themselves together with a few values preceeding them. In the case of electricity prices, this effect is recognized from the resulting noise component $d_t - \sum_{j=0}^{t-1} \theta(1-\theta)^j d_{t-j}$, which is too regular for first couple of values.

When smoothing $\{d_t\}_{t=1}^n$ using the exponential moving average technique, we start by creating the seasonal component

$$s_t(\theta) = \mathrm{m}_{\exp}(d_t, \theta) = \sum_{j=0}^{t-1} \theta(1-\theta)^j d_{t-j},$$

giving the noise component

$$y_t(\theta) = d_t - s_t(\theta) = d_t - \mathrm{m}_{\exp}(d_t, \theta).$$

Since our aim is to find a noise which can be modeled as an OU process, we want to find values of α and θ such that $\epsilon_t(\theta, \alpha) = y_{t+1}(\theta) - \alpha\, y_t(\theta)$, $t = 1, \ldots, n-1$, are IID.

More specifically, we wish to find values of α and θ, such that the LB test statistic

$$(n-1)(n+1) \sum_{k=1}^{M} \frac{\hat{\rho}^2\big(\{\epsilon_t(\theta, \alpha)\}_{t=1}^{n-1}, k\big)}{n-1-k} \leq \chi^2_{M,0.95}, \tag{7.1}$$

where M is the number of lags considered in the LB test.

We will carry out the parameter estimate by minimizing the LB test statistic with respect to θ and α. When we have found values of the parameters such that the test statistic is not significant, i.e., such that (7.1) holds, then we have a filtration which we consider satisfactory.

Naturally, a parameter pair θ and α that gives us the lowest value of the LB test statistic for one specific lag, say $M = 5$, do not have to minimize it for another lag size, say $M = 15$. This problem is resolved by trying to pick values of θ and α that make the test statistic non-significant for all lags considered.

Our specific choises of parameter values where $\theta = 0.235$ and $\alpha = 0.221$, for which the LB test showed independence for all lags checked. See Figure 12 below. Also note the seemingly stationarity stationary look of the noise, which is one of the features we are looking for.

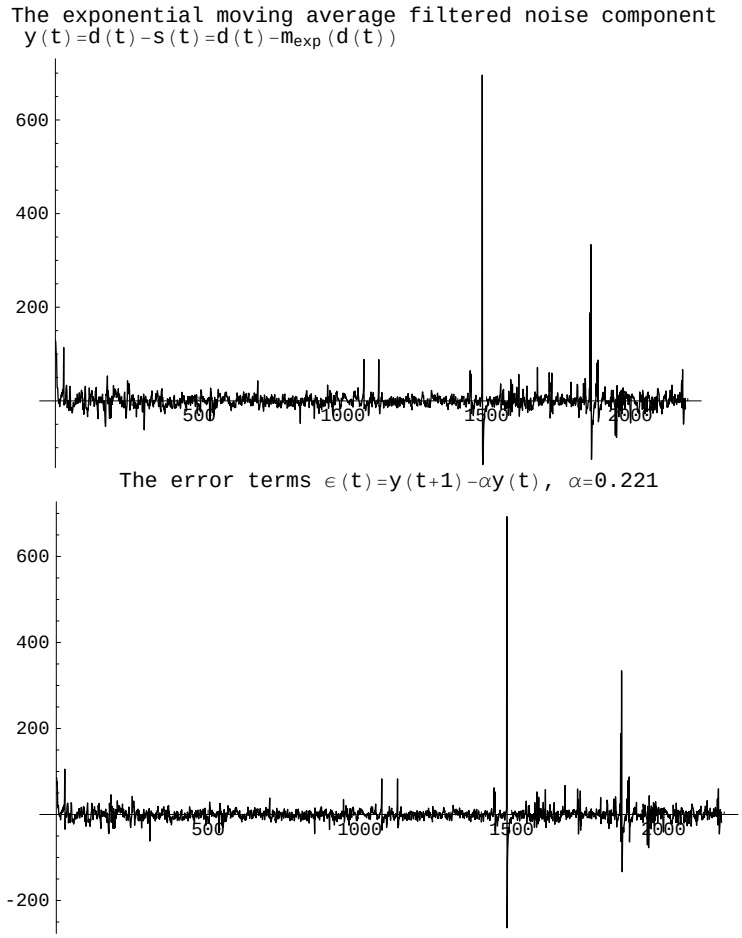
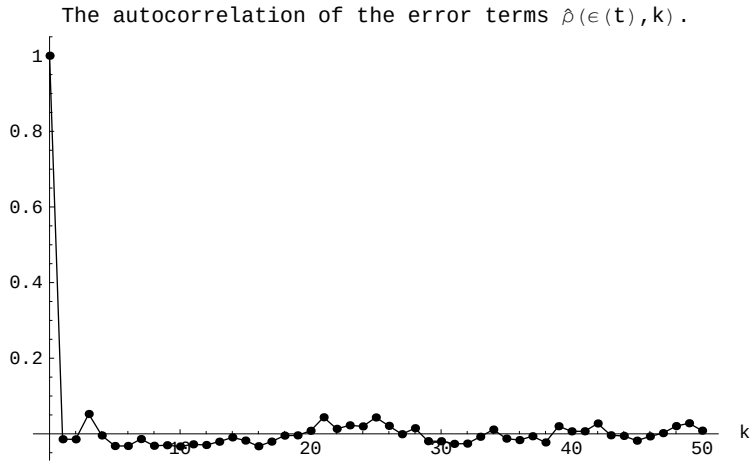


Figure 12: Upper: The exponential moving average filtered noise $y_t(\theta)$. Lower: The error terms of the exponential moving average filtered noise $\epsilon_t(\theta, \alpha)$.

We also checked is whether the sample auto correlations of the squared error terms $\hat{\rho}\big(\{\epsilon_t^2(\theta,\alpha)\}_{t=1}^{n-1}, k\big)$ had the same behaviour as the sample auto correlation of the error terms themselves. See Figure 13 below.



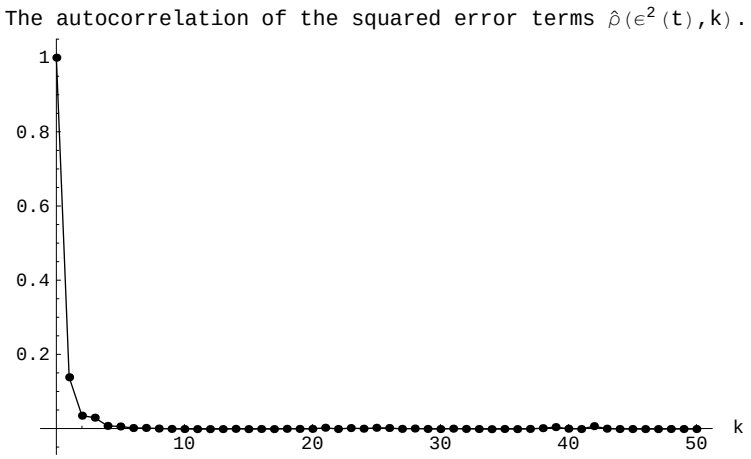


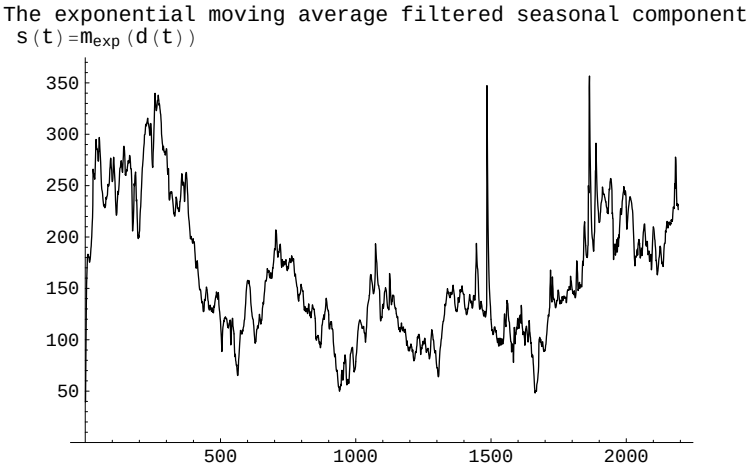
Figure 13: Upper: The sample auto correlation function of the error terms $\hat{\rho}(\{\epsilon_t(\theta, \alpha)\}_{t=1}^{n-1}, k)$. Lower: The sample auto correlation function of the squared error terms $\hat{\rho}(\{\epsilon_t^2(\theta, \alpha)\}_{t=1}^{n-1}, k)$.

As there seems to be no difference between these auto correlations, we do again have an indication of independence.

After extracting the noise from the deperiodized electricity prices, the seasonal component remains. Hence, the original electricity prices x_t are broken down as

$$x_t = p_t + s_t + y_t,$$

where $\{p_t\}_{t=1}^n$ is the period component, $\{s_t\}_{t=1}^n$ the seasonal component, and $\{y_t\}_{t=1}^n$ the noise component. See Figure 14 below.



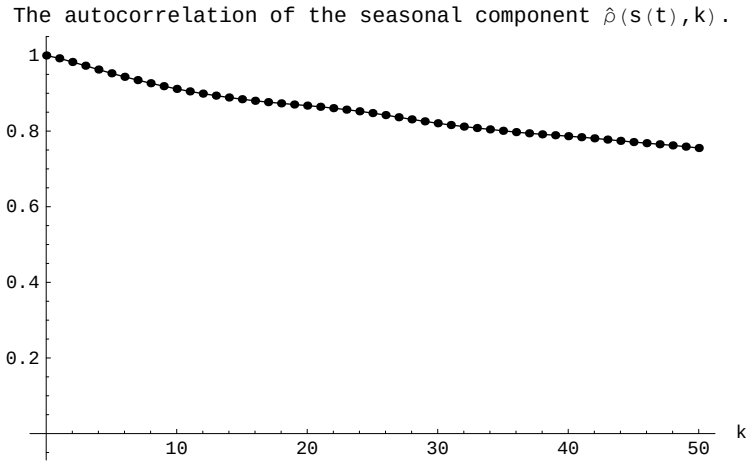


Figure 14: Upper: The exponential moving average filtered seasonal component $s_t = \mathrm{m}_{\exp}(d_t)$. Lower: The sample auto correlation function of the seasonal component $\hat{\rho}(\{s_t\}_{t=1}^n, k)$.

A of the exponential moving average approach, except for the problem with the starting values, is the lack of robustness of the filter. This is the reason why, at certain spikes, the seasonal component partially inherits the spiky behaviour. With these problems in mind, we try to find another filter, that is more robust.

It should be mentioned that the procedure described above has also been executed on logged electricity prices, without obtaing any satisfying result.

7.3 Moving median filtration

When looking for a robust smoothing technique, a natural choice is to use a moving median filter:

Definition 7.3 (Moving median) *The* median *of a data set* $\{x_t\}_{t=1}^n$ *is defined*

$$med\big(\{x_t\}_{t=1}^n\big) = \begin{cases} x_{(k+1)} & \text{if } n \text{ is odd} \quad n = 2k+1, \\ \dfrac{x_{(k)} + x_{(k+1)}}{2} & \text{if } n \text{ is even} \quad n = 2k, \end{cases}$$

where is $x_{(1)} < \ldots < x_{(n)}$ *is the ordered data set.*

The moving median filter *of an observed time series* $\{x_t\}_{t=0}^n$*, and an odd window size* $l = 2q+1$*, is given by*

$$m_{med}(x_t, l) = \begin{cases} med\big(\{x_k\}_{k=0}^{2q}\big) & for \quad 0 \le t \le q, \\ med\big(\{x_k\}_{k=t-q}^{t+q}\big) & for \quad q+1 \le t \le n-q+1, \\ med\big(\{x_k\}_{k=n-2q}^{n}\big) & for \quad n-q \le t \le n, \end{cases}$$

while for an even window size $l = 2q$*,*

$$m_{med}(x_t, l) = \begin{cases} med\big(\{x_k\}_{k=0}^{2q-1}\big) & for \quad 0 \le t \le q, \\ med\big(\{x_k\}_{k=t-q}^{t+q}\big) & for \quad q+1 \le t \le n-q+1, \\ med\big(\{x_k\}_{n-2q}^{n}\big) & for \quad n-q+2 \le t \le n. \end{cases}$$

The end-effect problems for the moving median filter is taken care of in the same way as for the moving average filter.

As for the exponential smoothing, we start by creating the seasonal component $s_t(l) = \mathrm{m}_{\mathrm{med}}(d_t, l)$, after which the noise is obtained as $y_t(l) = d_t - s_t(l) = d_t - \mathrm{m}_{\mathrm{med}}(d_t, l)$. Just as before, we want the noise to display the OU process characteristics of being stationary, and makin the error terms $\epsilon_t(l, \alpha) = y_{t+1}(l) - \alpha\, y_t(l)$, $t = 1, \ldots, n-1$, IID for some $\alpha \in (0, 1)$.

Once again, we look for parameter values l and α such that the LB test statistic is non-significant for each lag M considered, i.e.,

$$Q(M) = (n-1)(n+1) \sum_{k=1}^{M} \frac{\hat{\rho}^2\big(\{\epsilon_t(l,\alpha)\}_{t=1}^{n-1}, k\big)}{n-1-k} \leq \chi^2_{M,0.95}.$$

Our parameter choices where $l = 9$ and $\alpha = 0.173$. See Figure 15 below. Note the stationarity look of the noise.

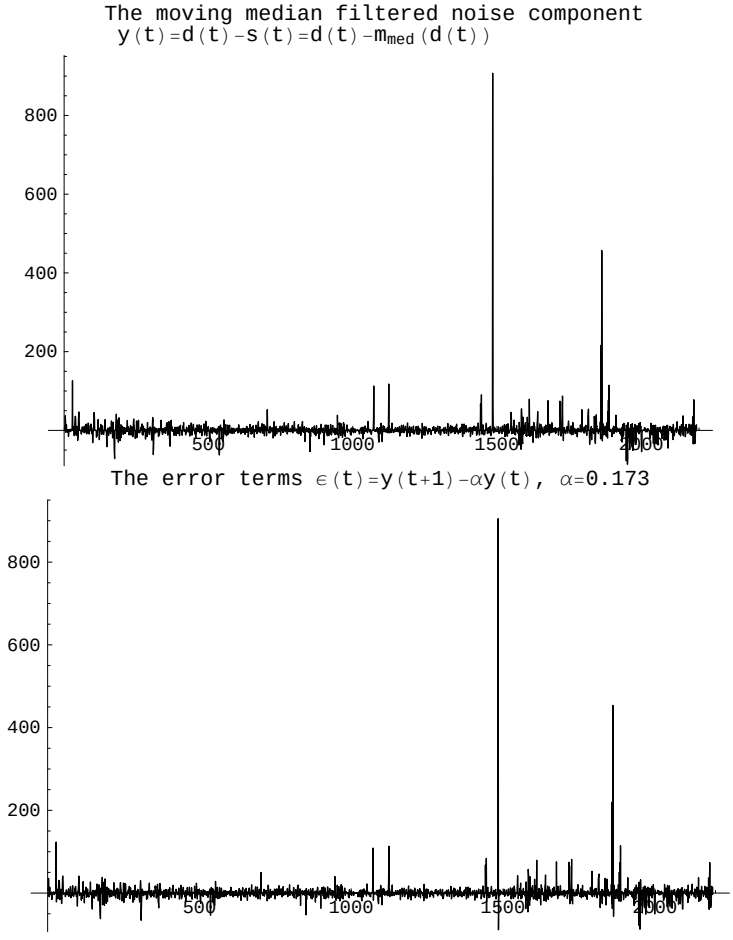


Figure 15: Upper: The moving median filtered noise $y_t(l)$. Lower: The exponential moving average filtered error terms $\epsilon_t(l, \alpha)$.

Figure 16 below shows a plot of the LB test statistic $Q(M)$ and the $\chi^2_{0.95,M}$-quantile for $M = 1, \ldots, 40$.

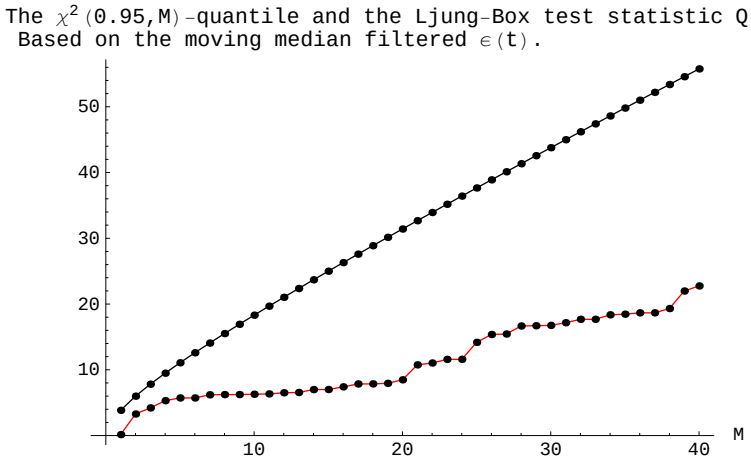


Figure 16: Upper: The $\chi^2_{M,0.95}$-quantile. Lower: The LB test statistic $Q(M)$ for the error terms $\epsilon_t(l,\alpha)$.

Figure 17 below shows plots of the sample auto correlation functions of the error terms and the squared error terms.

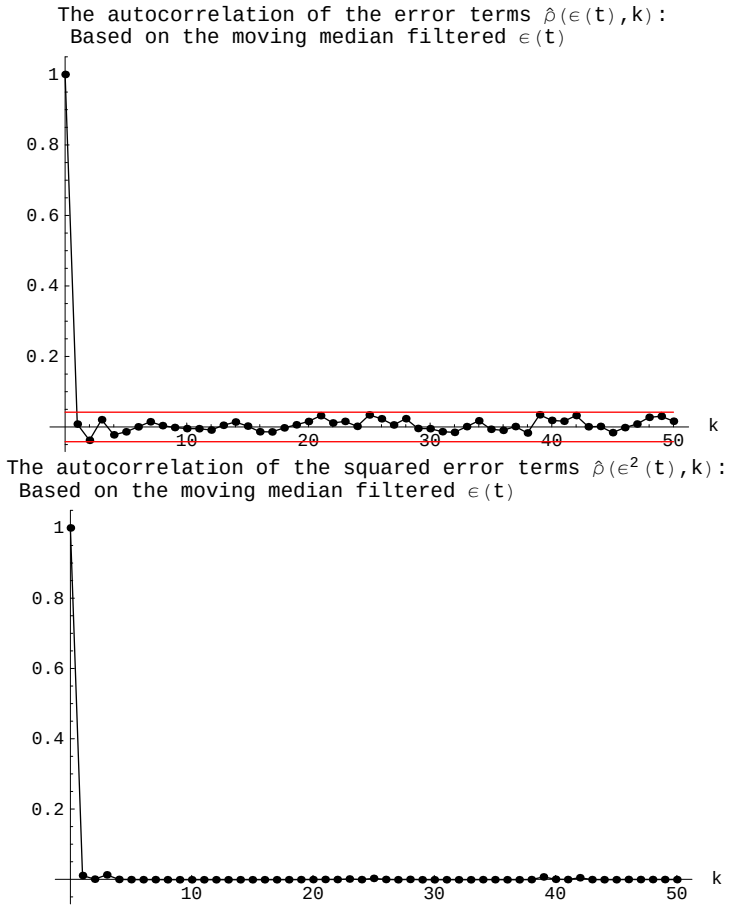


Figure 17: Upper: The sample auto correlation of the errors $\hat{\rho}(\{\epsilon_t(l,\alpha)\}_{t=1}^m, k)$. Lower: The sample auto correlation of the squared errors $\hat{\rho}(\{\epsilon_t^2(l,\alpha)\}_{t=1}^m, k)$.

In the auto correlation plot we have also plotted confidence bands for the auto correlation, assuming independence, given by $\pm 1.96/\sqrt{n}$, see [8], pp. 19-20.

Note that there is no additional dependence present in the auto correlations of the squared error, as compared with the auto correlations of the error terms themselves, which is another indication of independenceides, besides the fact that the auto correlations themselves indicates independence.

What we are left with after extracting the noise from the deperiodized electricity prices is the seasonal component. See Figure 18 below.

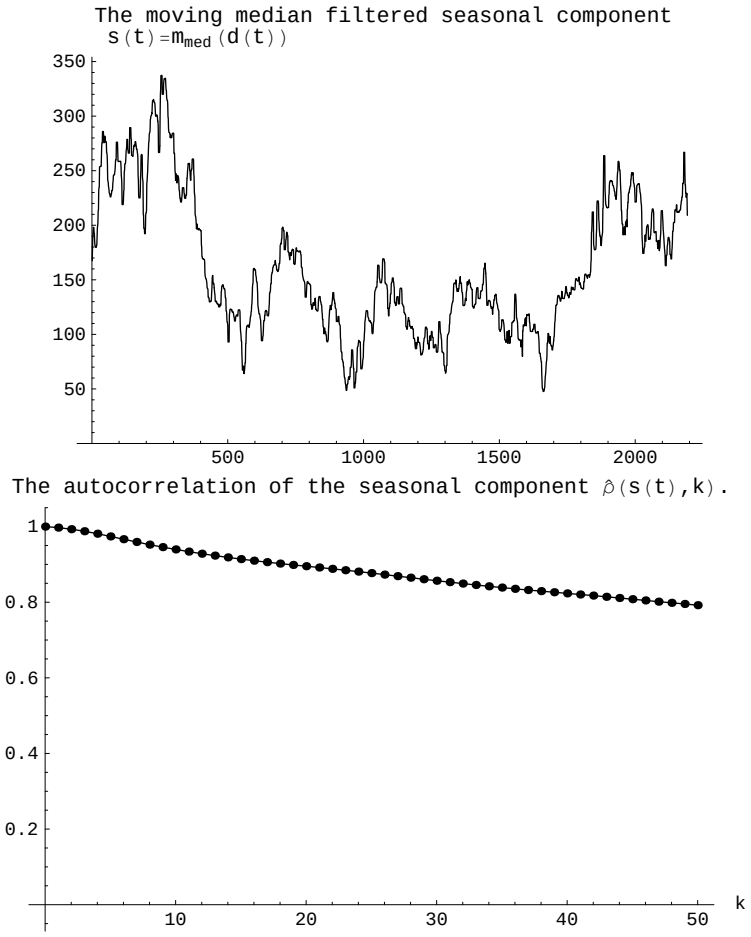


Figure 18: Upper: The exponential moving median filtered seasonal component $s_t = \mathrm{m}_{\mathrm{med}}(d_t)$. Lower: The sample auto correlation fucntion of the seasonal component $\hat{\rho}(\{s_t\}_{t=1}^n, k)$.

The original electricity prices $\{x_t\}_{t=1}^n$ can be broken down as

$$x_t = p_t + s_t + y_t,$$

where $\{p_t\}_{t=1}^n$ is the periodic component, $\{s_t\}_{t=1}^n$ is the seasonal component, and $\{y_t\}_{t=1}^n$ is the noise component.

Hence, we have managed to filter out two different noise time series, both of which satisfy our requirements, by an exponential smoothing and a moving median filtration, respectively. Which one should we choose?

The choice is actually quite easy: If one starts by looking at the independence tests, the moving median filtration showed stronger indications of independence.

Another factor that supports the moving median filtered noise is found from the pure looks of the two filtered components, which is where the robustness of the moving median filtration comes into the picture: The exponentially filtered seasonal component contains a lot of the spikes from the raw electricity prices, whereas the moving median filtered seasonal component shows a more smooth behaviour.

So what does that indicate? The answer is that the moving median filtration suceeded better in putting the jumps and spikes into the noise component, rather than into the seasonal component, than did the exponential filtration.

8 Empirical critical values

8.1 IID data

Sometimes when one wants to make inference, the test statistics used either have no tabulated quantiles nor computer package implementations, or have no known distributions at all. In these situations one might simulate critical quantiles for the test statistic in question.

In this thesis a lot of goodness-of-fit tests are made, using the KS test statistic. The way this test is used here, it assumes that our data IID observations from a parametric distribution F. So we fit the parameters of F to our data, for instance using maximum likelihood (ML) estimation, and perform our KS test by evaluating the KS distance between the fitted distribution $\hat{F}$ and the empirical distribution F_n of the data.

To get critical valued for our KS test we do as follows:

1. Estimate the parameters for the parametric distribution F, to recive a fitted distribution $\hat{F}$.

2. Calculate the KS distance between the fitted distribution $\hat{F}$ and the empirical distribution F_n of the data.

3. Simulate 1000 samples of size n, the size of the original data set, from the distribution $\hat{F}$.

4. For each of the 1000 samples, make a fit of the parameters for the parametric distribution F, and calculate the corresponding KS distance, to obtain simulated KS statistics $\{KS_n^{(i)}\}_{i=1}^{1000}$.

5. Find the empirical 0.95-quantile of $\{KS_n^{(i)}\}_{i=1}^{1000}$. This value is the simulated critical value to be used in the KS test.

It should be observed that, as critical KS test values asymptotically do not depend on the distribution that is tested, for our rather large data set, all simulated critical values should be close to the asymptotic distribution free critical value 0.03, or just below it, because of the flexibility offered by the parameters.

To find critical test values for the Anderson-Darling goodness-of-fit test introduced in Definition 9.1 below, we employed an obvious modification of the above simulation scheme.

We also employed a version of (Steps 3-5 of) the procedure above, to find simulated critical values for our LB tests. But it turned out that these simulated critical values differed in a neglible way from those of the asymptotic χ^2 distributions of the test statistics.

8.2 Non-independent data

Some times our data $\{x_t\}_{t=1}^n$ are observations of a stationary Markov process X with a parametric transition PDF

$$f_t(y;\Theta|x) = \mathbf{P}\big\{X(t+s) \in dy \big| X(s) = x\big\},$$

As such data are not independet, the ML estimation procedure has to be adapted from the elmentary IID context, to our Markov data.

For observations $\{x_t\}_{t=1}^n$ of a stationary Markov process X with transition PDF $f_t(\cdot;\Theta|\cdot)$, the likelihood is given by

$$L(\Theta) = f_{X(1),\ldots,X(n)}(x_1,\ldots,x_n;\Theta) = f_{X(1)}(x_1;\Theta)\prod_{i=2}^{n} f_1(x_i;\Theta|x_{i-1}).$$

As Markov process data are not IID, their goodness-of-fit can not be evaluated by a KS test of the fit of the marginal distribution. However, the likelihood could be used to test goodness-of-fit as well, as is now described:

1. Assume that the data set $\{x_1,\ldots,x_n\}$ is an observation of the random variable $(X(1),\ldots,X(n))$ with a parametric distribution $F_{X(1),\ldots,X(n)}(x_1,\ldots,x_n;\Theta)$.

2. Estimate the parameter(s) Θ by the ML method, and note the value of the corresponding likelihood $L(\Theta)$.

3. Simulate 1000 samples of size n, coming from the fitted distribution.

4. For each of the simulated sample, re-estimate the parameter(s) by the ML method, and note the corresponding likelihoods $\{L(\Theta)^{(i)}\}_{i=1}^{1000}$.

5. Find the empirical 0.05 quantile of $\{L(\Theta)^{(i)}\}_{i=1}^{1000}$. This value is the simulated critical value to be used in the goodness-of-fit test, so that if the likelihood $L(\Theta)$ of the fit to the original data is lower than this critical value, then the hypothesis that the data has the distribution $F_{X(1),\ldots,X(n)}(x_1,\ldots,x_n;\Theta)$ is rejected.

9 Fitting OU processes to the noise

Recall that we in Section 7.3 by a moving median filtration obtained a filtered data set $\{y_t\}_{t=1}^n$ that could be modeled as an OU process, as the coefficient $\alpha = 0.17$ gave us an independent data set $\{e(t)\}_{t=1}^n = \{y_t - \alpha\, y_{t-1}\}_{t=1}^n$. See Figure 15 above.

We have tried to fit VG, GIG, IG, GH, Meixner, Gaussian and AIDD distributions to this noise $\{e(t)\}_{t=1}^n$.

We do not report the detailed results fit of the fits of the VG, GIG, IG and Gaussian distributions, because we had numerical problems that could not be resolved with the ML fit of the VG and GIG distributions, while the IG and Gaussian distributions gave very poor fits. We did also tried the method of moments for the VG fit, but found that the moments of our data could not be fitted to VG moments for any parameter values.

One regularly uses the KS test to test goodness-of-fit. However, if one is interested in analyzing extreme values, an additional *Anderson-Darling* (AD) test is motivated, because the latter test is more sensitive for deviations in the tails betseen the tested distributions, than is the KS test.

Definition 9.1 (Andersson-Darling (AD)) *Given data $x_1, \ldots, x_n$ with* empirical distribution *function F_n, the* AD goodness-of-fit test *of a fitted CDF $\hat{F}$ is based on the AD test statistic* AD_n*, also called the* the AD distance *between F_n and $\hat{F}$, given by*

$$\mathrm{AD}_n = \sup_{x \in \mathbb{R}} \frac{\left|\hat{F}(x) - F_n(x)\right|}{\sqrt{\hat{F}(x)\big(1 - \hat{F}(x)\big)}}.$$

9.1 Fitting Meixner OU processes

We fitted the Meixner distribution to $\{e(t)\}_{t=1}^n$ using the ML method, and calculated the KS and AD goodness-of-fit test statistics. To test if these statistics were significantly too large, indicating a poor fit, we used simulated critical values as described in Section 8.

The ML parameter estimates are given together with the KS and AD tests in Table 5 below.

α	β	δ	μ
114.3976	0.6407	0.0403	-0.3276

	Observed test statistic value	simulated Meixner critical value
KS	0.0270959	0.0295183
AD	0.680677	0.10796

Table 5: Parameter values of the fitted Meixner distribution and the KS and AD goodness-of-fit tests.

The KS test indicates a good fit. However, the AD test, on the other hand, informs us that the Meixner distribution do not seem well-suited to model the

extreme values present in the data. This means that the judgement whether it is a good fit or not depends on what one wants to model. It is of course also possible to model the tails separately.

Figure 19 below shows the fitted Meixner PDF together with a histogram of the data $\{e(t)\}_{t=1}^{n}$. Further, a plot the empirical distribution function is plotted together with the fitted Meixner CDF.

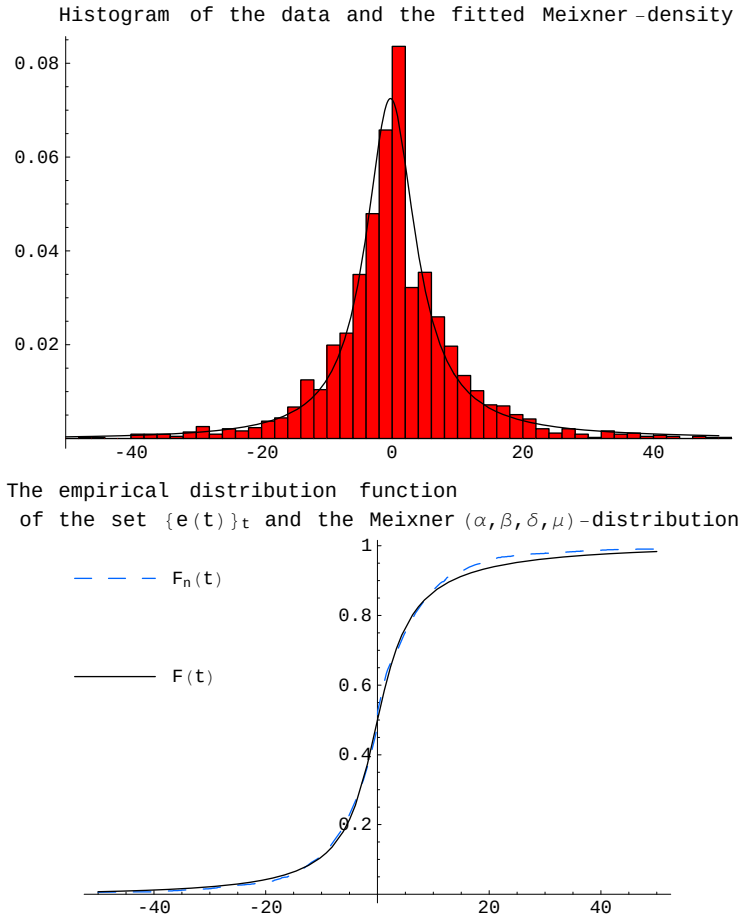


Figure 19: Upper: The fitted Meixner PDF together with a histogram of the data. Lower: The empirical distribution function together with the fitted Meixner CDF.

Following our positive result with the Meixner fit, we are all anxious to see how similar a simulated Meixner OU process trajectory, based on fitted Meixner distribution, looks when compared with the moving median filtered data $\{y_t\}_{t=1}^{n}$. Such a comparison is displayed in Figure 20 below.

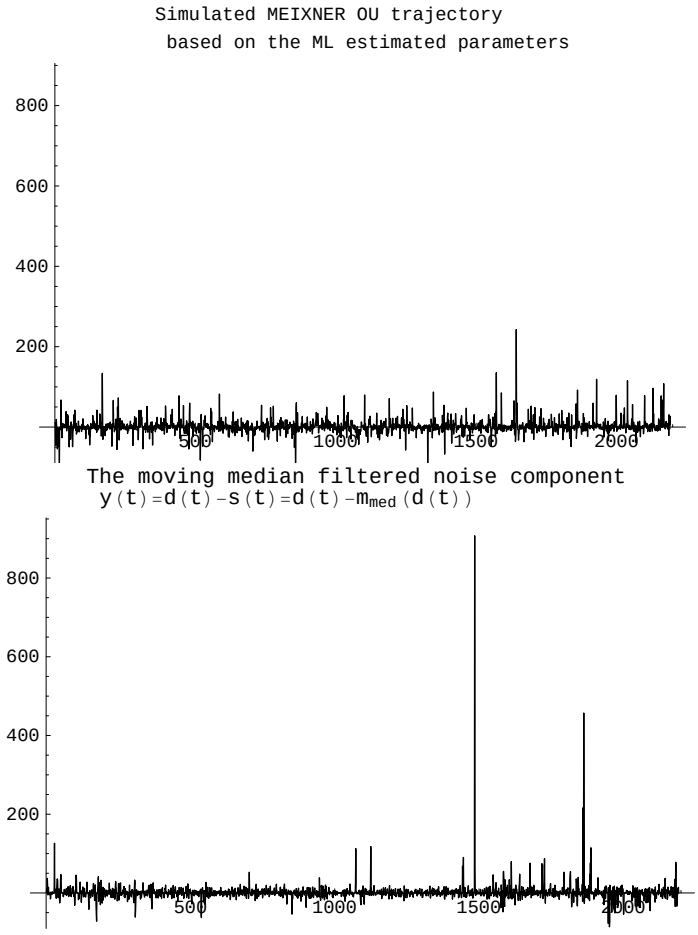


Figure 20: Upper: Simulated Meixner OU trajectory. Lower: The moving median filtered data $\{y_t\}_{t=1}^n$.

9.2 Fitting NIG OU processes

We fitted the NIG distribution to $\{e(t)\}_{t=1}^n$ using the ML method, and performed KS and AD goodness-of-fit tests of the fit. The results are displayed in Table 6 below.

μ	β	α	δ
-1.14773	0.0199287	0.0339394	5.18225

	Observed test statistic value	simulated NIG critical values
KS	0.029483	0.0300413
AD	0.216949	0.118424

Table 6: ML parameter estimates for the NIG distribution and the goodness-of-fit tests.

We have a good fit, according to the KS test, while the AD test statistic is slightly higher than the empirical critical value, indicating a less than perfect fit of the tails. However, as the NIG AD test statistic is considerably smaller than

that for the Meixner fit, it seems that NIG models the extreme values of the data better than Meixner.

Figure 21 below shows the fitted NIG PDF together with a histogram of the data $\{e(t)\}_{t=1}^{n}$. Further, the empirical distribution function is plotted together with the fitted NIG CDF.

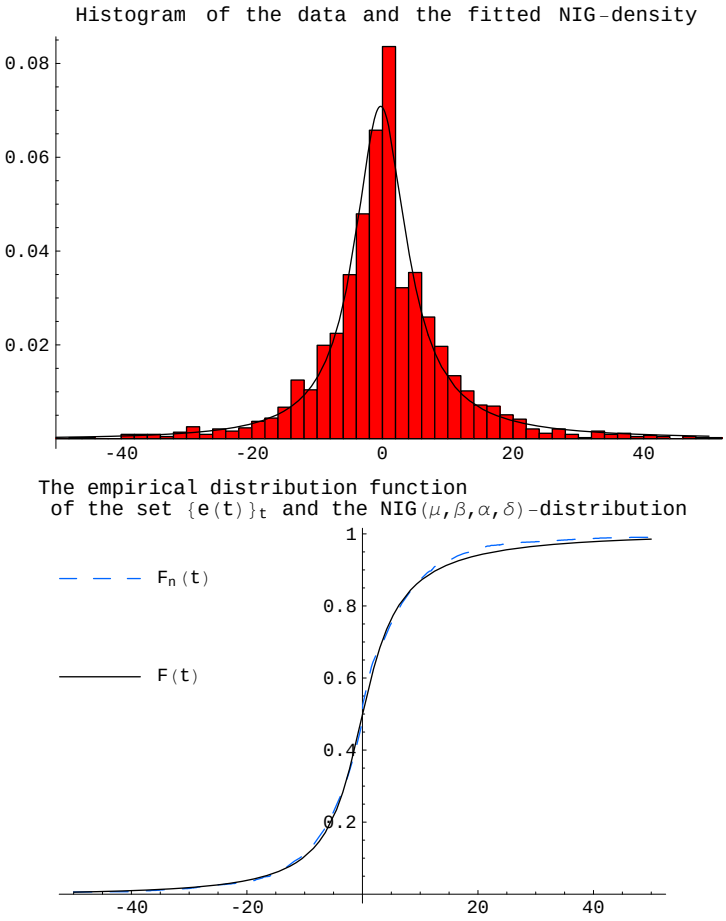


Figure 21: Upper: The fitted NIG PDF together with a histogram of $\{e(t)\}_{t=1}^{n}$. Lower: The empirical distribution function together with the fitted NIG CDF.

Figure 22 below shows a simulated NIG OU process trajectory, based on the fitted NIG distribution, together with the moving median filtered data $\{y_t\}_{t=1}^{n}$.

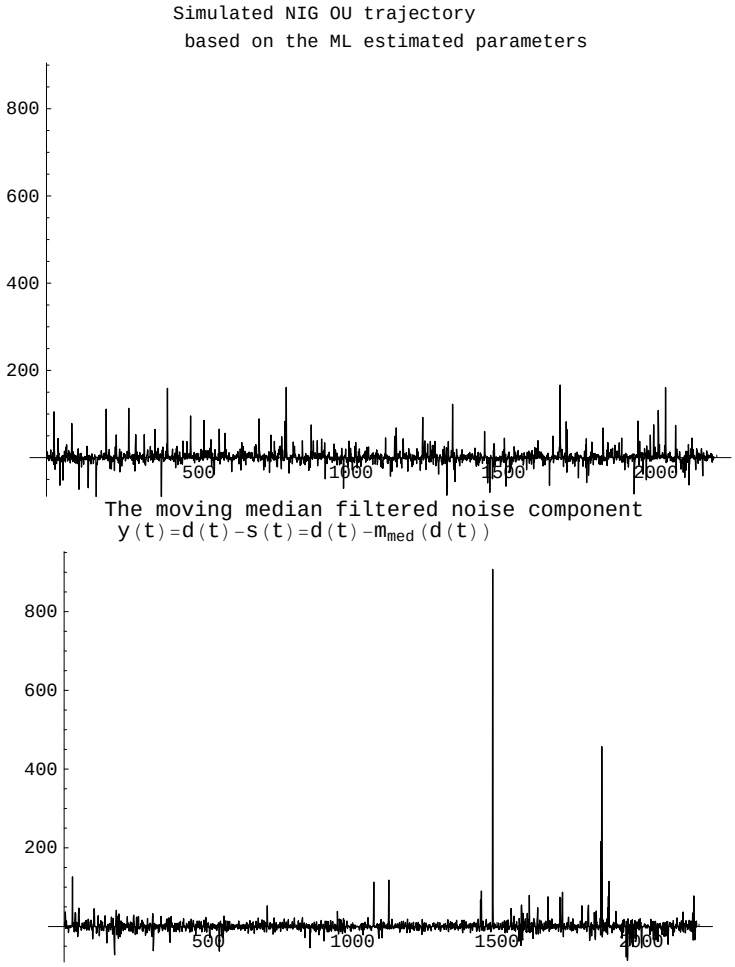


Figure 22: Upper: Simulated NIG OU trajectory. Lower: The moving median filtered data $\{y_t\}_{t=1}^n$.

9.3 Fitting GH OU processes

We fitted the GH distribution to $\{e(t)\}_{t=1}^n$ using the ML method, and calculated the KS distance for the fit. The results are displayed in Table 7 below.

α	β	δ	v	μ	KS
0.0055689	0.0043228	7.07693556	-0.8621878	-0.2520776	0.0363786

Table 7: Parameter values of the fitted GH distribution and the KS distance.

We had difficulties in simulating samples from a GH distribution with the fitted parameters, and could therefore not find the critical value of the KS test by simulations. However, as that critical value should be slightly less than or equal to 0.03, it is clear that the KS distance 0.036 indicates a poor fit of the GH distribution to the data.

As the GH distribution contains the NIG distribution as a special case, a correct fit of the GH distribution should always be as least as good as a NIG

fit. Thus our failure with the fit of the GH distribution has to be blamed on numerical problems finding the true global maximum of the likelihood for the very complicated GH distribution, with its 5 parameters.

Figure 23 below shows the fitted GH PDF together with a histogram of the data $\{e(t)\}_{t=1}^{n}$. Further, the empirical distribution function is plotted together with the fitted GH CDF.

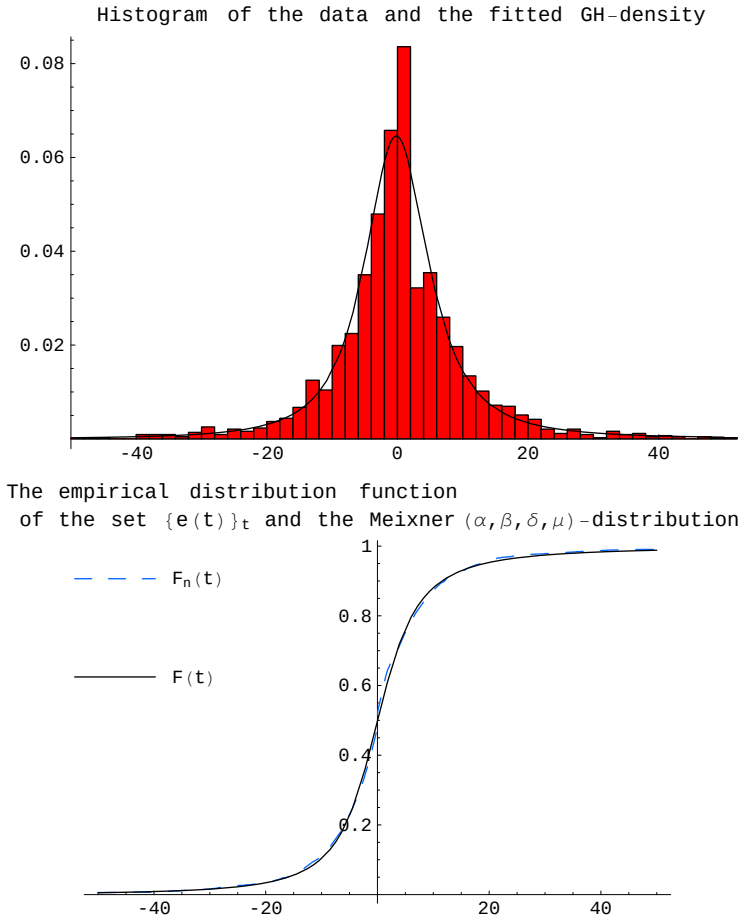


Figure 23: Upper: The fitted GH PDF together with a histogram of $\{e(t)\}_{t=1}^{n}$. Lower: The empirical distribution function together with the fitted GH CDF.

As a result of our problems with simulating random numbers from the fitted GH distribution, we cannot give a plot of a GH OU process trajectory, based on the fitted GH distribution.

9.4 Fitting AIDD OU processes

We now describe the results of the AIDD fit: Recall that the AIDD CHF looks as follows

$$\phi_{\text{AIDD}}(u;\sigma,\mu,a_1,\ldots,a_n,b_1,\ldots,b_n) = \exp\left\{i\mu u - \frac{\sigma^2 u^2}{2} + \sum_{k=1}^{n} a_k\left(\mathrm{e}^{iub_k} - 1\right)\right\}.$$

The more steps n of added rescaled Poisson distributed random variables that is carried out, the better the fit will be, at least in theory. However, because of the many parameters involved, one cannot take for granted that fitting procedures like the ML method will be unproblematic.

The AIDD estimation procedure runs as follows:

Step 1 In the first step, $n = 1$, we found estimates of the parameters $\{\sigma^{(1)}, \mu^{(1)}, a_1^{(1)}, b_1^{(1)}\}$ by the method of moments.

Step 2 In the second step, $n = 2$, we found estimates $\{\sigma^{(2)}, \mu^{(2)}, a_1^{(2)}, a_2^{(2)}, b_2^{(2)}\}$ of the parameters by keeping $b_1 = b_1^{(1)}$ from Step 1, and using the ML method to estimate the other parameters.

Step 3 In the third step, $n = 3$, we found estimates $\{\sigma^{(3)}, \mu^{(3)}, a_1^{(3)}, a_2^{(3)}, a_3^{(3)}, b_3^{(3)}\}$ of the parameters by keeping $b_1 = b_1^{(1)}$ and $b_2 = b_2^{(2)}$ from Step 2, and using the ML method to estimate the other parameters.

Step .. This procedure is carried on step-by-step until a decent fit is achieved.

After nine steps we got the KS distance 0.083, which we felt satisfied with. Naturally one can keep on taking more steps until a much better fit is found. Table 8 below shows the results of the fit.

a_1	a_2	a_3	a_4	a_5
$1.829 \cdot 10^{-4}$	$7.559 \cdot 10^{-6}$	65847	$5.216 \cdot 10^{-4}$	$8.616 \cdot 10^{-4}$
b_1	b_2	b_3	b_4	b_5
0.641	$3.821 \cdot 10^{-2}$	$3.676 \cdot 10^{-7}$	$4.524 \cdot 10^{-1}$	-0.09

a_6	a_7	a_8	a_9	σ	
$6.969 \cdot 10^{-1}$	$1.412 \cdot 10^{-4}$	$2.604 \cdot 10^{-6}$	$3.213 \cdot 10^{-4}$	0.0100	
b_6	b_7	b_8	b_9	μ	KS
$3.326 \cdot 10^{-6}$	$-7.147 \cdot 10^{-4}$	0.888	0.893	-0.0241	0.083

Table 8: The estimated AIDD parameters and the KS distance.

Naturally this model is somewhat intractable because of the many parameters. Therfore it argubly is to prefer to try to find distributions with fewer parameters, that give good fits, as we have done with the Meixner and NIG distributions.

Figure 24 below shows the fitted AIDD PDF together with a histogram of the data $\{e(t)\}_{t=1}^n$.

Figure 25 below shows a simulated AIDD OU process trajectory, based on the fitted AIDD distribution, together with the moving median filtered data $\{y_t\}_{t=1}^n$.

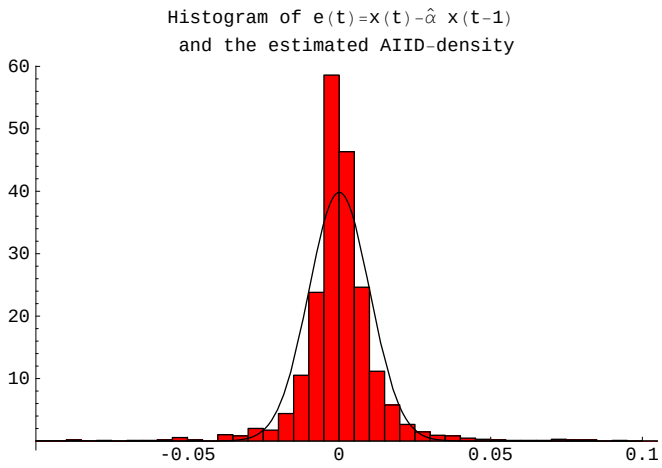


Figure 24: The fitted AIDD PDF together with a histogram of $\{e(t)\}_{t=1}^{n}$.

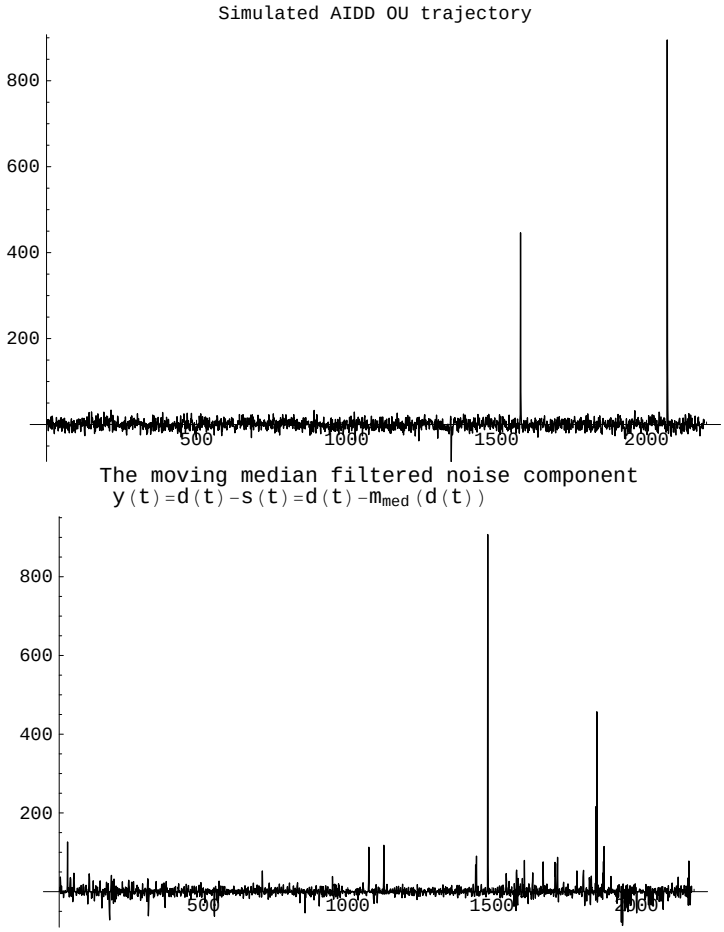


Figure 25: Upper: Simulated AIDD OU trajectory. Lower: The moving median filtered data $\{y_t\}_{t=1}^{n}$.

10 Seasonal component and periodicity

10.1 Seasonal component

Recall that we had filtered the electricity spot prices and thereby achieved a component called the seasonal component.

A factor that affects the electricity spot price a lot is the amount of water in the reservoirs. Thus, the plot of the typical amount of water in the reservoirs during a year in Figure 26 below is of great interest for our study of electricity spot prices.

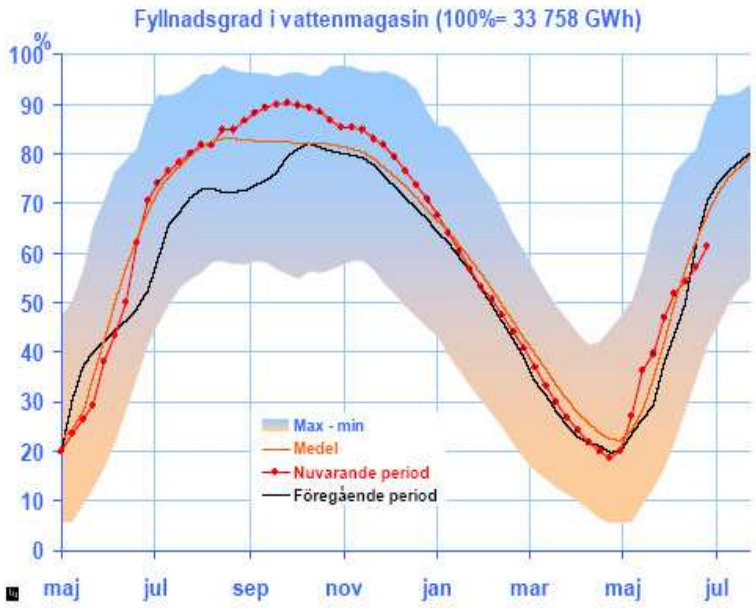


Figure 26: Water in the reservoirs.

Figures 27-29 below show the seasonal component, for each of the years 1996-2001 separately, in order to illustrate the behaviour of the price related to the amount of water in the reservoirs.

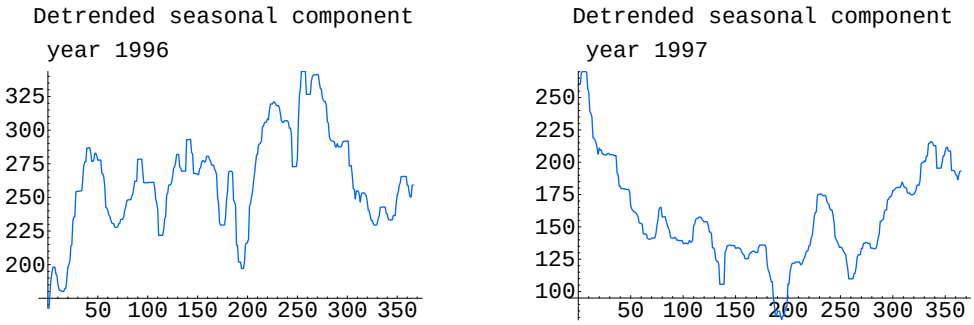


Figure 27: Upper: Seasonal component of 1996. Lower: Seasonal component of 1997.

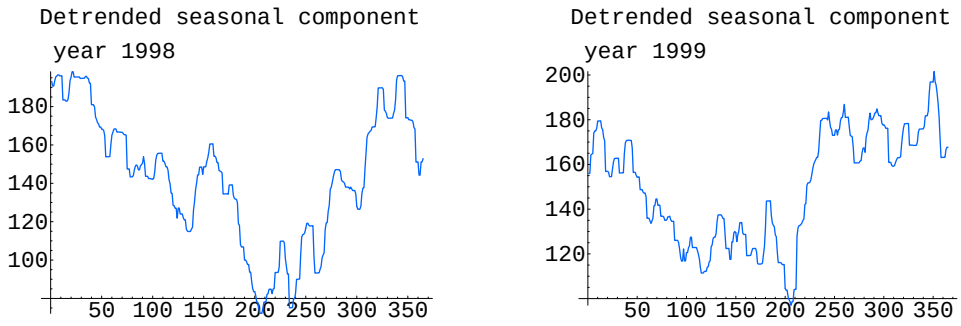


Figure 28: Upper: Seasonal component of 1998. Lower: Seasonal component of 1999.

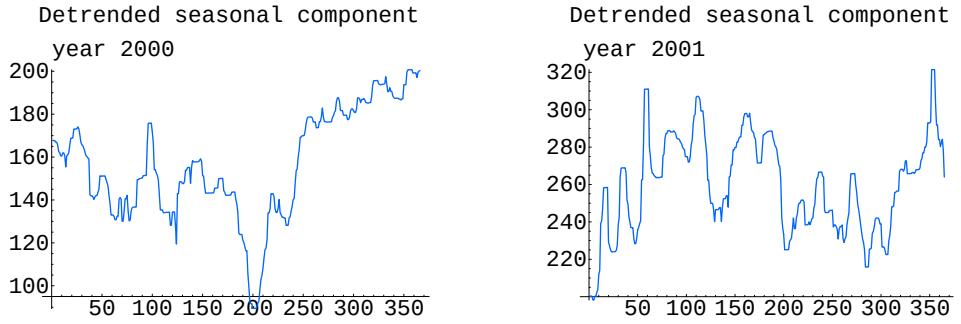


Figure 29: Upper: Seasonal component of 2000. Lower: Seasonal component of 2001.

Before starting to explain the models used for fits to the seasonal component, we will start by explaining which data set we have chosen to model. To that en we first present the seasonal component data set in Figure 30 below.

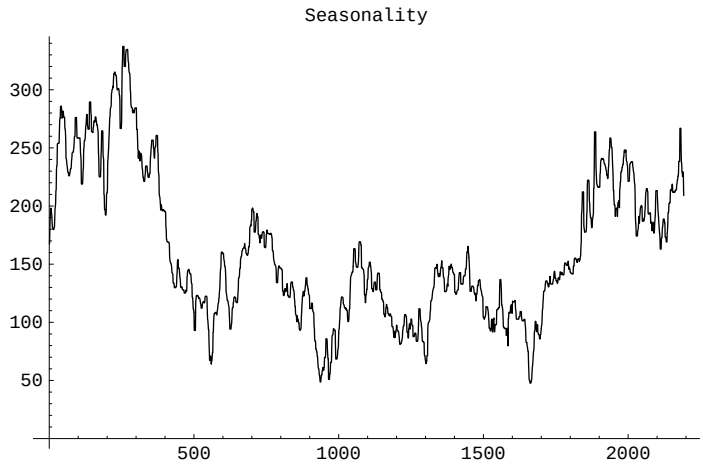


Figure 30: The seasonal component for the years 1996-2001.

We first remove linear trend from the seasonal component. The resulting detrended data set is shown in Figure 31 below.

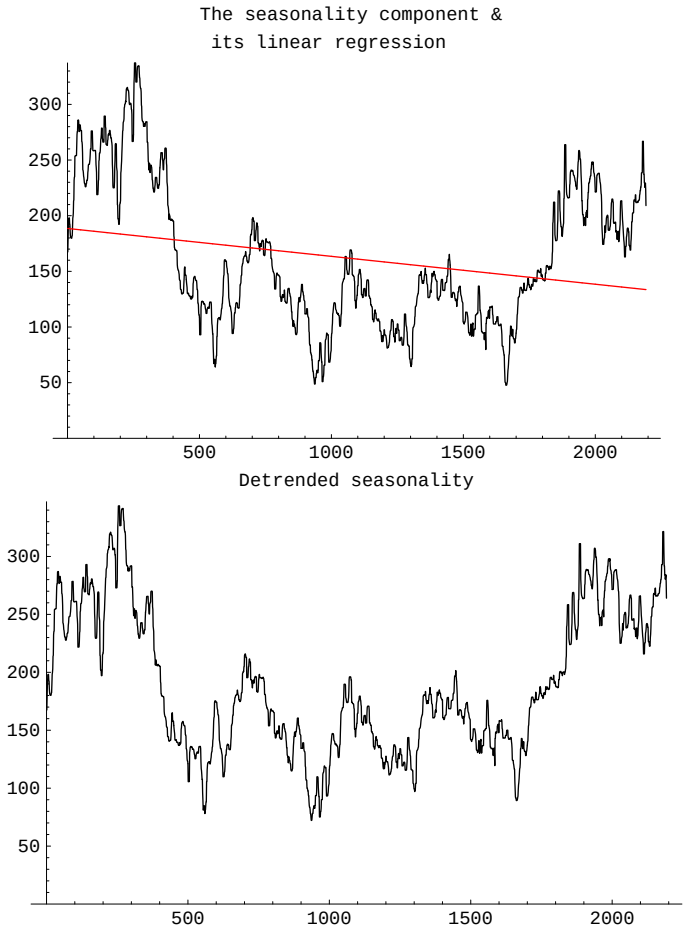


Figure 31: Upper: The seasonal component and its linear regression.
Lower: The detrended seasonal component.

A statistical test showed that the slope of linear trend -0.025 differs significantly from the zero trend.

As a second step we used Fourier series to filter out seasonal periodicities the detrended data set. To that end we the Fourier series

$$\sum_{k=1}^{8}\left[a(k)\cos\left(\frac{tk2\pi}{365}\right)+b(k)\sin\left(\frac{tk2\pi}{365}\right)\right],$$

where

$$\begin{aligned} a(k) &= \frac{2}{n}\int_0^n \cos\left(\frac{tk2\pi}{365}\right)\mathrm{metroid}(t)\,dt, \\ b(k) &= \frac{2}{n}\int_0^n \sin\left(\frac{tk2\pi}{365}\right)\mathrm{metroid}(t)\,dt. \end{aligned}$$

Here n is the length of our data set (the detrended seasonal component), which is equal to 2192. Further, metroid is the function which linearly interpolates the data points in our time series, $\{x_t\}_{t=1}^n$, which is defined as

$$\mathrm{metroid}(t) = (\lfloor t\rfloor + 1 - t)\,x_{\lfloor t\rfloor+1} + (t-\lfloor t\rfloor)\,x_{\lfloor t\rfloor+2}.$$

We filtered the detrended seasonal component using the Fourier series as described above, and found the deterministic periods to be as shown in Figure 32 below.

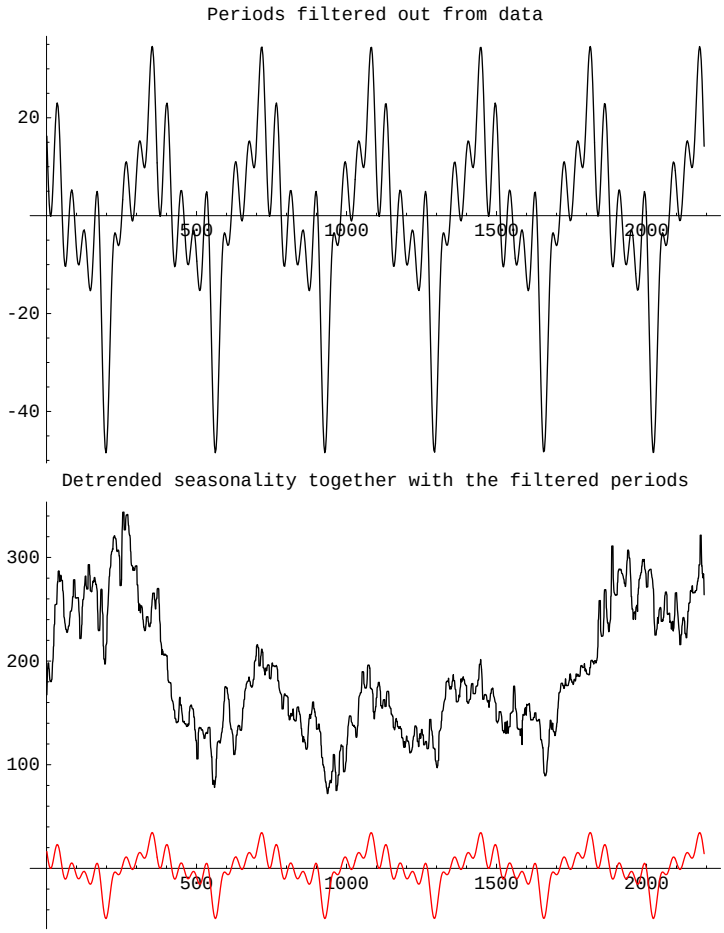


Figure 32: Upper: Periods extracted from the detrended seasonal component.
Lower: Extracted periods together with the detrended seasonal component.

To get the final result of the filtering procedure, we subtract the deterministic periods we found from the detrended seasonal component. The detrended deperiodized seasonal component obtained in this manner is shown in Figure 33 below.

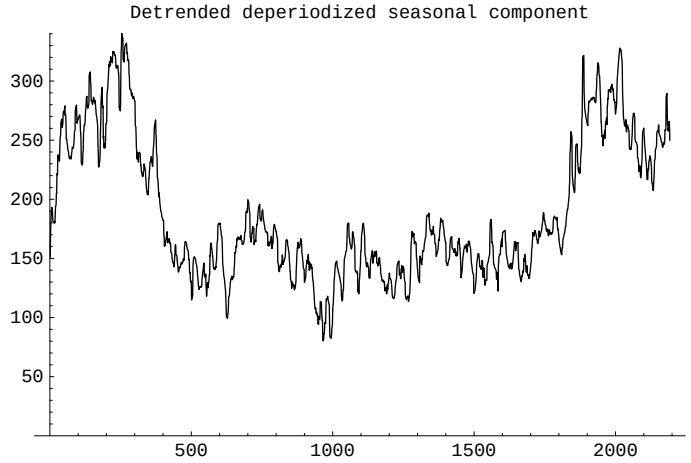


Figure 33: Detrended deperiodized seasonality.

In addition, we mean adjust the detrended deperiodized seasonal component. Since we had no access to historical data on water levels, nor to data on historical temperatures, we were forced to remove the influence of these two factors in some way. The reason for this is that we know that they are two major influences on electricity prices. This can easily be seen by looking at the years 1996 and 2001. The mean levels of the electricity prices of these years are much higher than those of the years 1997-2000. The reason for this is that the years 1996 and 2001 are so called dry years, meaning that the levels of water in the reservoirs of these years were much lower than the water levels of the years 1997-2000.

The temperatures during a year are strongly correlated with the levels of water in the reservoirs. This phenomenon, in turn, is strongly correlated to the supply and demand of eletricity. We selected to deal with this problem by removing the mean level of each year, of the detrended deperiodized seasonal component, so the data would lie on the same level. The data mean adjusted in this fashion is shown in Figure 34 below.

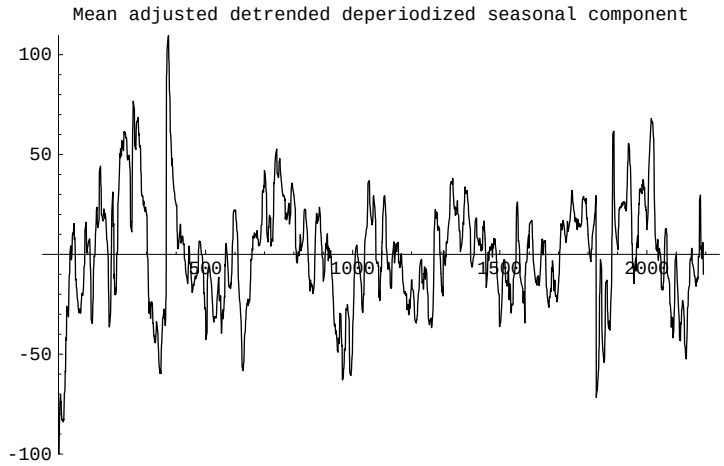


Figure 34: Mean adjusted detrended deperiodized seasonal component.

It is this data set which we will model from here on. For easiness, we will refer to the mean adjusted detrended deperiodized seasonal component as our data set.

We will be modeling the data using the *Lévy market model* $X(t) = X(0)\,\mathrm{e}^{L(t)}$ where $\{L(t)\}_{t\geq 0}$ is a Lévy process, which is a family of exponential models used regularly in financial modeling. Further, we be modeling the data using two different diffusion processes, namely the *CIR* process and the *Vasicek interest rate model*

10.2 Periodicity

Recall that we had recognized weekly periods in the auto correlation function of the electricity spot prices and thereby used filtration methods for extracting them. We called this deterministic component *Periodicity*.

Figure 35 shows a plot of the periodicity of the electricity spot prices.

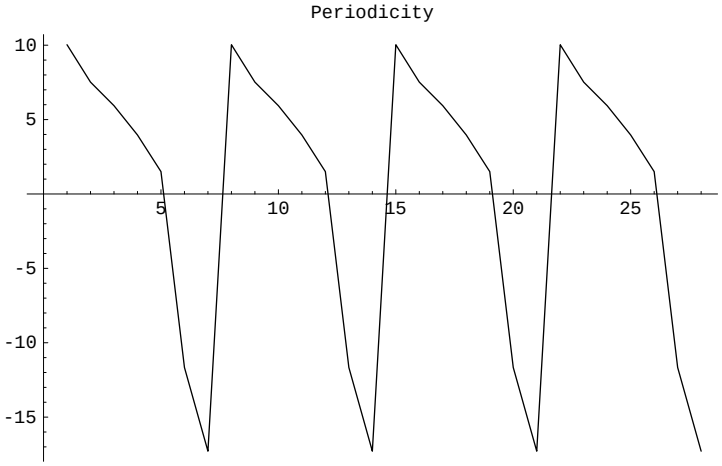


Figure 35: Periodicity, 4 weeks.

For modelling this deterministic component, the *Periodicity*, we use the previously defined *metroid* function together with Fourier series which we described earlier. In our specific case, we included $k = 7$ components in the Fourier series

$$\sum_{k=0}^{6}\left[c(k)\cos\left(\frac{tk2\pi}{7}\right) + d(k)\sin\left(\frac{tk2\pi}{7}\right)\right].$$

11 Lévy market model

For the Black-Scholes type exponential Lévy model $X(t) = X(0)\,\mathrm{e}^{L(t)}$, we have

$$\varepsilon_t \equiv \log(X(t)) - \log(X(t-1)) = L(t) - L(t-1) \stackrel{D}{=} L(1).$$

If our observed data $\{x_t\}_{t=1}^n$ follows this model, the observed log-returns (log-increments) $r_t = \log(x_t) - \log(x_{t-1})$ should thus be IID.

The mean adjusted detrended deperiodized seasonal component has some negative values. Therefore we changed their starting point to that of the original data, the seasonal component, so that the log-returns are taken on the mean adjusted detrended deperiodized seasonal component starting as the seasonal components.

Figure 36 below shows the observed log-returns.

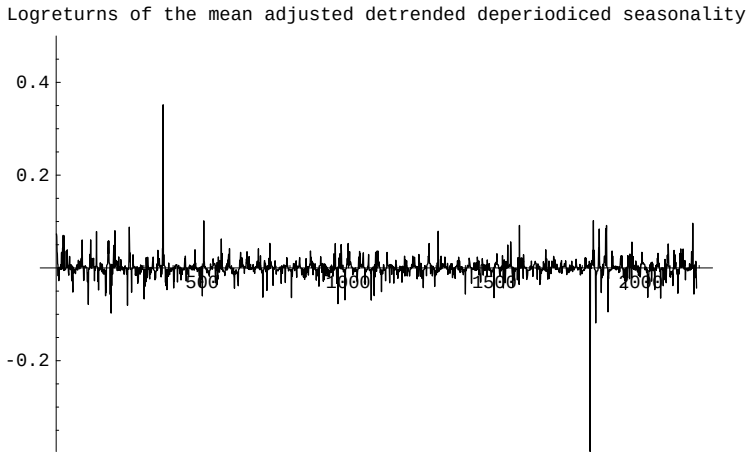


Figure 36: Log-returns of the data.

Figures 37 and 38 below show a histogram of the observed log-returns, and a normal probability plot of them, respectively.

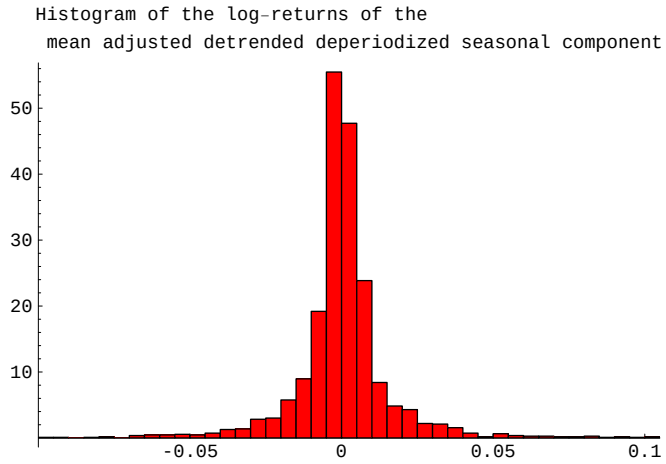


Figure 37: Histogram of the log-returns.

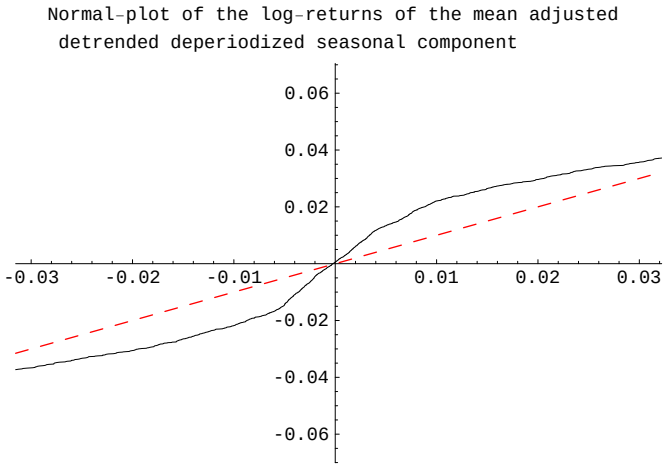


Figure 38: Normal probability plot of the log-returns.

Table 9 lists the first 4 moments of the observed log-returns.

Mean	0.000217616
Variance	0.000395041
Standard Deviation	0.0198756
Skewness	122.526
Kurtosis	-0.94944

Table 9: Stylized facts of the log-returns.

We will try to fit appropriately distributed theoretical log-increments $\{\varepsilon_t\}_{t=1}^n$ to our observed log-returns $\{r_t\}_{t=1}^n$.

After the successful procedure of getting a good fit for the log-returns, we want to get back to the original data set. So, after simulating $\{\varepsilon_t\}_{t=1}^n$, we might compare the simulated Lévy model $X(t) = X(t-1)\,\mathrm{e}^{\varepsilon_t}$ and $X(0) = x_0$ with the observed data $\{x_t\}_{t=1}^n$.

We did try several different Lévy processes to achieve a good fit of the log-returns. In the final stage we transform the data back to its original form.

Below we report on the fitting of the Meixner distribution, the NIG distribution, and the GH distribution to the observed log-returns.

We also tried to fit a Gaussian distriburion, which gave a very poor fit, and had unsuccesful attempts to fit VG and GIG distribution: For both these latter distributions the ML method gave numerical problems that we could not resolve. We tried the method of moments for the VG distribution, but it turned out that VG moments could not be fitted to the moments of the dataset.

11.1 Fitting Meixner distribution to log-returns

The Meixner distribution gave good fit to the observed log-returns: Table 10 below shows the results of the ML parameter estimation for the Meiner distribution, together with a KS test.

α	β	δ	μ
0.1188618	0.1592844	0.0462980	-0.0002550

	Observed test statistic value	Meixner empirical critical value
KS	0.0277297	0.0282861

Table 10: ML estimates of parameters of the Meixner distribution and the KS test.

As one can see from the KS test, we have a good fit.

Figure 39 below shows the PDF of the ML method fit of the Meixner distribution together with a histogram of the observed log-returns. Further, a plot of the empirical distribution function of the log-returns is given together with the fitted Meixner CDF.

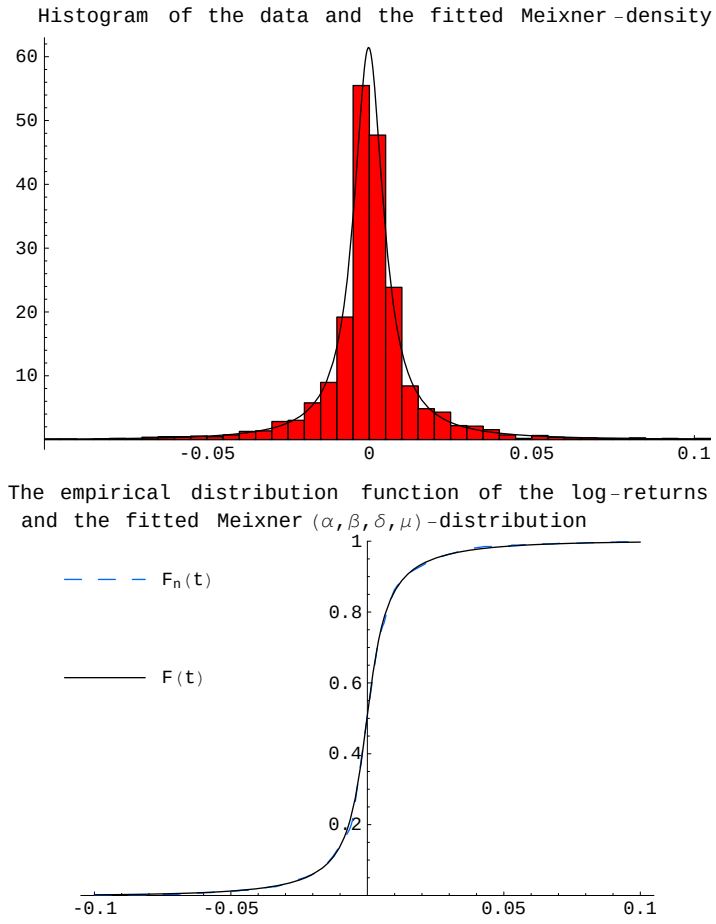


Figure 39: Upper: The fitted Meixner PDF together with a histogram of the log-returns. Lower: Empirical distribution function together with the fitted Meixner CDF.

We can see that the Meixner distribution gives a good fit to the log-returns.

Figure 40 below shows simulated Meixner noise from the fitted model, together with the observed log-returns.

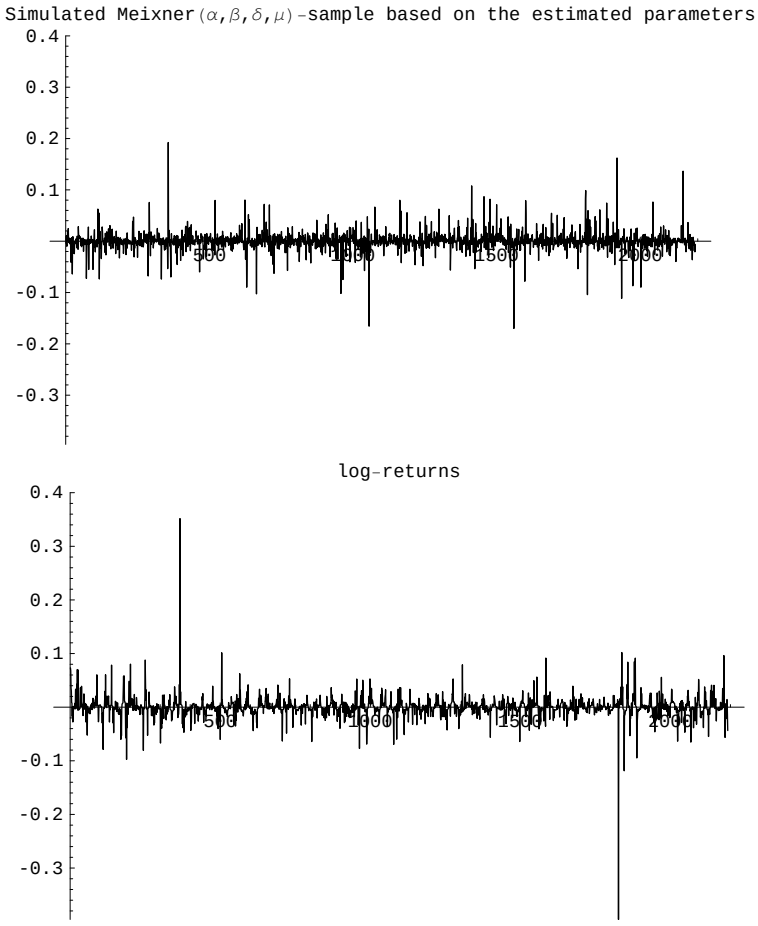


Figure 40: Upper: Simulated Meixner noise. Lower: Log-returns.

11.2 Fitting NIG distribution to log-returns

The NIG distribution gave a good fit to the observed log-returns: Table 11 below shows the results of the ML parameter estimation for the NIG distribution, together with a KS test.

μ	β	α	δ
-0.0002470	1.3498047	16.546860	0.0056761

	Observed test statistic value	NIG empirical critical value
KS	0.0280766	0.029992

Table 11: Parameter values of the fitted NIG distribution and value of the goodness-of-fit test statistic.

Figure 41 below shows the PDF of the ML fitted NIG distribution together with a histogram of the log-returns. Further, a plot of the empirical distribution function of the log-returns is given together with the fitted NIG CDF.

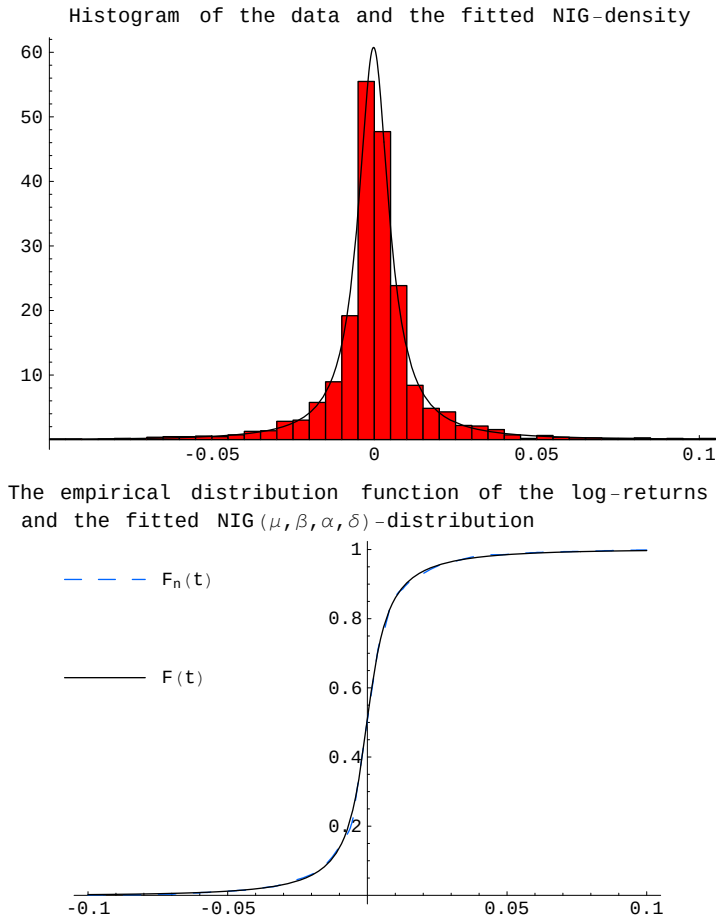
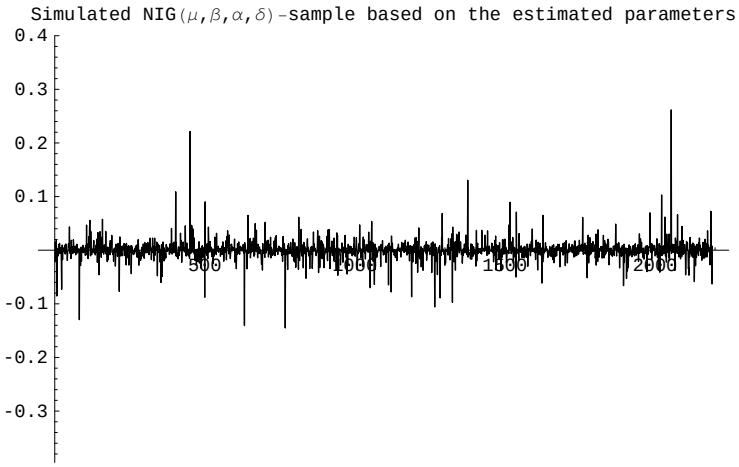


Figure 41: Upper: The fitted NIG PDF together with a histogram of the log-returns. Lower: Empirical distribution together with the fitted NIG CDF.

We see that the NIG distribution gives a really good fit to our data set.

Figure 42 below shows a plot of simulated NIG noise based on the fitted NIG model, together with the observed log-returns.



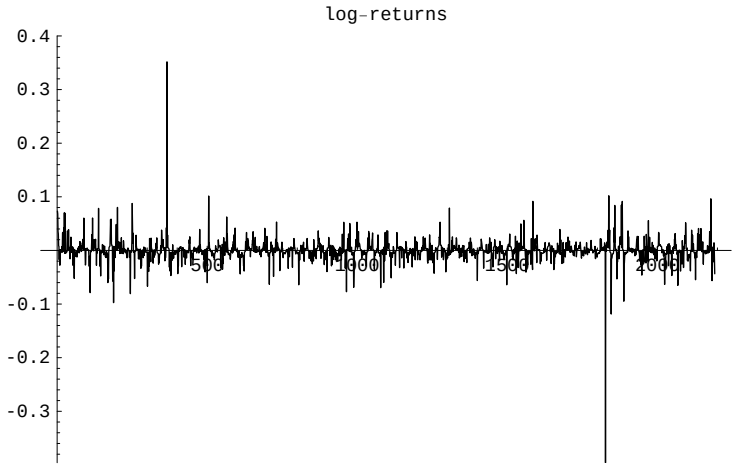


Figure 42: Upper: Simulated NIG noise. Lower: Log-returns.

11.3 Fitting GH distribution to log-returns

The GH distribution gave a good fit as well: Table 12 below shows the results of the ML parameter estimation for the GH distribution, together with a KS test.

α	β	δ	v	μ
17.588743	1.2799168	0.0055485	-0.478612	-0.000248

	Observed test statistic value	GH empirical critical value
KS	0.02822	0.02959

Table 12: Parameter values of the fitted GH distribution and value of the goodness-of-fit test statistic.

Figures 43 and 44 below show a plot of the PDF of the fitted GH distribution, together with a histogram of the log-returns, and a plot of the empirical distribution function of the log-returns, together with the fitted GH CDF, respectively.

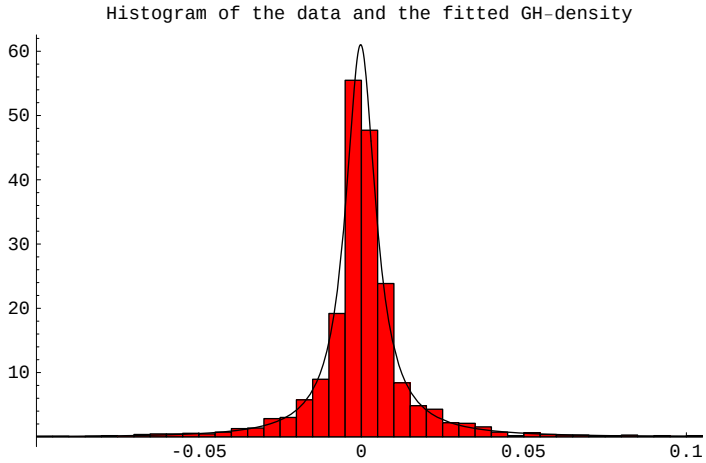


Figure 43: The ML fitted GH PDF and a histogram of the observed log-returns.

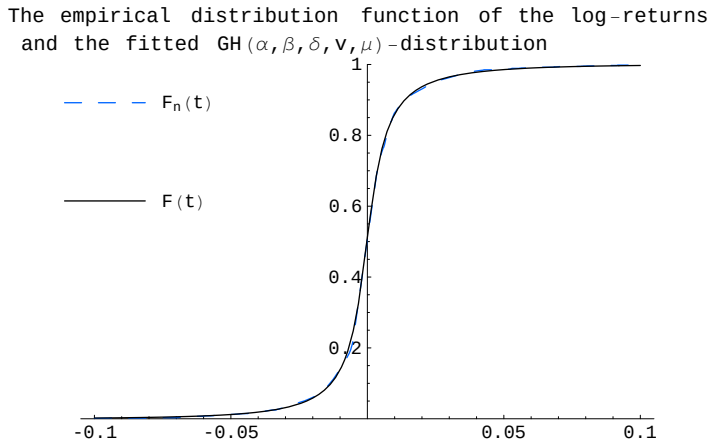


Figure 44: Empirical distribution function together with the fitted GH CDF.

Figure 45 below shows a plot of simulated GH noise based on the fitted GH model, together with the observed log-returns.

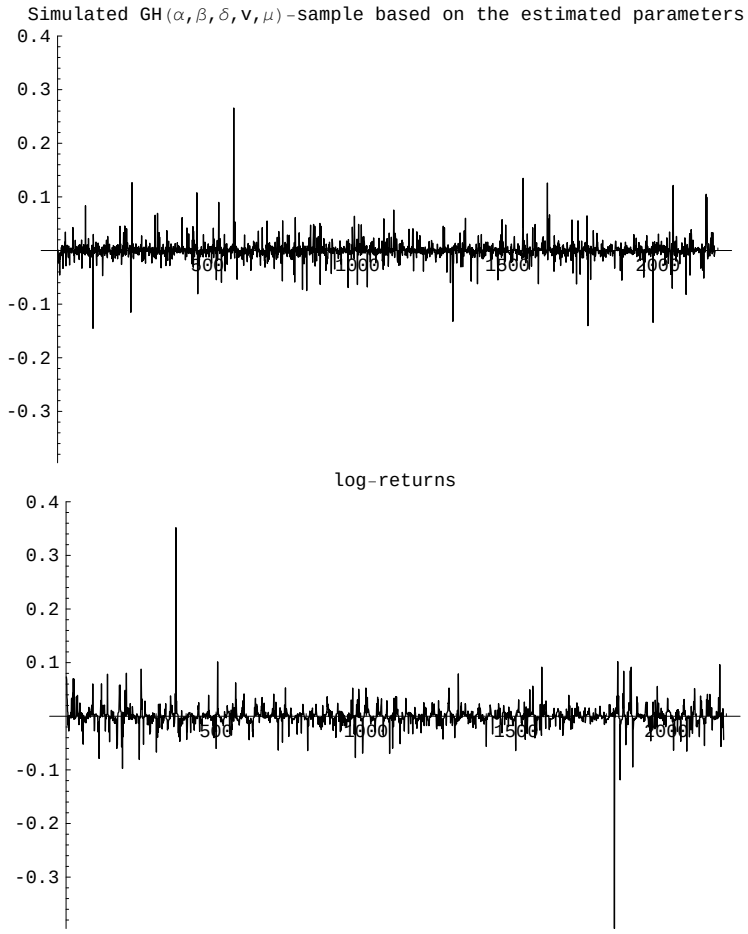


Figure 45: Upper: Simulated GH noise. Lower: Log-returns.

11.4 Independence of log-returns

Thus far we have assumed that the log-returns are independent and identically distributed. However, we have not yet checked the assumption of independence. This we will do now: We start by plotting the empirical auto correlation function of the observed log-returns, as well (as is customaty in mathematical finance) the auto correlation of the squared log-returns. See Figure 46 below.

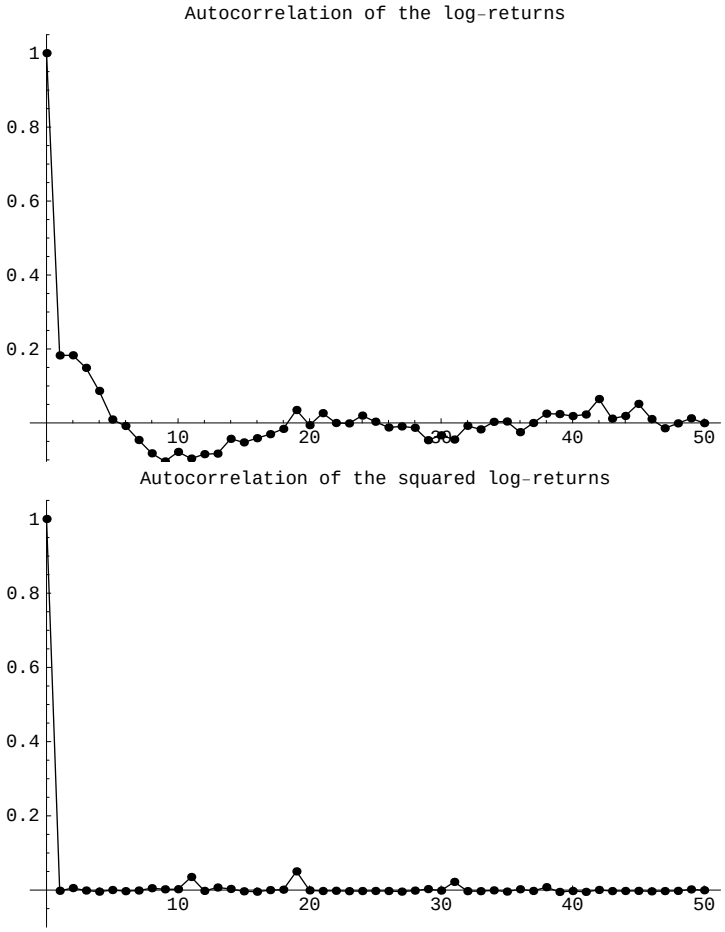


Figure 46: Upper: auto correlation of the log-returns. Lower: Auto correlation of the squared log-returns.

Further, we give a table of the LB test statistic, calculated for the observed log-returns, at lags $M = 1, \ldots, 25$. These test statistics are compared to the χ^2_M distribution for $M = 1, \ldots, 25$. See Table 13 below.

M	1	2	3	4	5
$Q_{LB}(M)$	73.1209	146.54	195.232	211.697	211.886
$\chi^2_{0.95,M}$	3.84146	5.99153	7.81473	9.48773	11.0705
M	6	7	8	9	10
$Q_{LB}(M)$	212.035	216.729	231.592	255.506	269.068
$\chi^2_{0.95,M}$	12.5916	14.0671	15.5073	16.919	18.307
M	11	12	13	14	15
$Q_{LB}(M)$	289.558	305.299	320.507	324.599	330.728
$\chi^2_{0.95,M}$	19.6751	21.0261	22.362	23.6848	24.9958
M	16	17	18	19	20
$Q_{LB}(M)$	334.496	336.538	337.113	339.812	339.893
$\chi^2_{0.95,M}$	26.2962	27.5871	28.8693	30.1435	31.4104
M	21	22	23	24	25
$Q_{LB}(M)$	341.436	341.436	341.439	342.308	342.332
$\chi^2_{0.95,M}$	32.6706	33.9244	35.1725	36.415	37.6525

Table 13: Quantiles of the $\chi^2_{0.95,M}$ distribution together with the LB test statistic $Q_{LB}(M)$ of the log-returns, for lags $M = 1, \ldots, 25$.

We can see in the comparison, for every lag, that the log-returns are in fact strongly dependent, rather than independent. A further illustration of this is given in Figure 47 below.

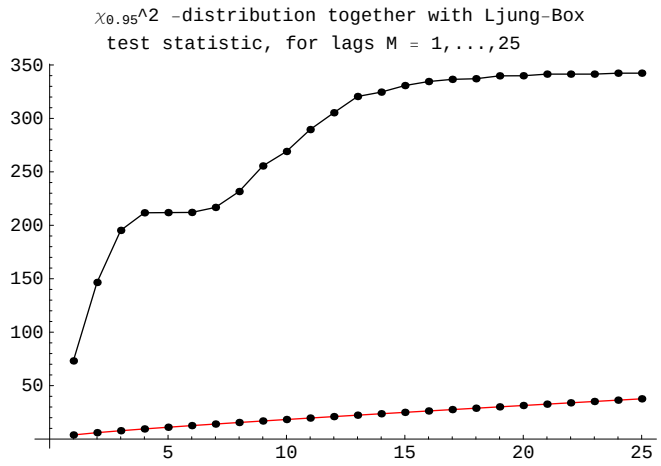


Figure 47: Upper: The LB test statistic of the log-returns, $Q_{LB}(M)$, $M = 1, \ldots, 25$. Lower: $\chi^2_{0.95,M}$-quantiles, $M = 1, \ldots, 25$.

As the requirement of independence of the log-returns is not fulfilled, we conclude that the Lévy market model is not the model we are looking for.

12 Fitting diffusions to the seasonal component

12.1 CIR

The Cox Ingersoll Ross (CIR) model $\{X(t)\}_{t\geq 0}$ is given by

$$dX(t) = b(a - X(t))dt + \sqrt{X(t)}dB(t),$$

where $a, b > 0$ are parameters and $\{B(t)\}_{t\geq 0}$ is standard Brownian motion.

The moments of a CIR process X are given by

$$\begin{aligned} \mathbf{E}[X(t)] &= \mathrm{e}^{-bt}\, X(0) + a\left(1 - \mathrm{e}^{-bt}\right), \\ \mathbf{Var}[X(t)] &= X(0)\left(\mathrm{e}^{-bt} - \mathrm{e}^{-2bt}\right)/b + a\left(1 - 2\,\mathrm{e}^{-bt} + \mathrm{e}^{-2bt}\right)/(2b). \end{aligned}$$

Further, the CIR process has a stationary distribution

$$F_{\text{stat}}(x; a, b) = 1 - \frac{\Gamma(2ab, 2bx)}{\Gamma(2ab)}$$

and stationary density

$$f_{\text{stat}}(x; a, b) = \frac{4^{ab} b^{2ab}\, \mathrm{e}^{2bx}\, x^{2ab-1}}{x\Gamma(2ab)}.$$

The transition density is given by

$$f_{X(t)|X(0)}(y|x) = \frac{2b}{1 - \mathrm{e}^{-bt}} \exp\left\{-\frac{2b(x + \mathrm{e}^{bt}y)}{\mathrm{e}^{bt} - 1}\right\} \left(\frac{\mathrm{e}^{bt}y}{x}\right)^{ab-1/2} I_{2ab-1}\left(\frac{4b\,\mathrm{e}^{bt/2}\sqrt{xy}}{\mathrm{e}^{bt} - 1}\right),$$

where I is the modified Bessel function of the first kind.

Using the stationary density and the transition density we estimate the parameters using the ML method. Since the process has a stationary distribution, we can use the KS test. The result are diplayed in Table 14 below.

a	b
111.854	0.0980397

	Observed test statistic value
KS	0.0474246

Table 14: Upper: Parameter values of the estimated CIR model.
Lower: The KS goodness-of-fit test statistic.

Figure 48 shows the empirical distribution together with the stationary distribution of the fitted CIR model. Figure 49 shows the mean adjusted seasonal component together with a simulated fitted CIR trajectory.

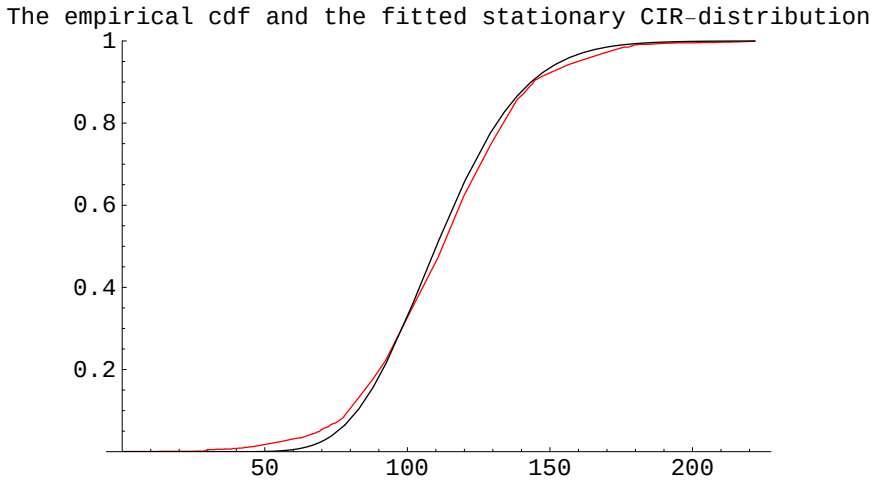


Figure 48: Empirical distribution function and the fitted stationary CIR distribution.

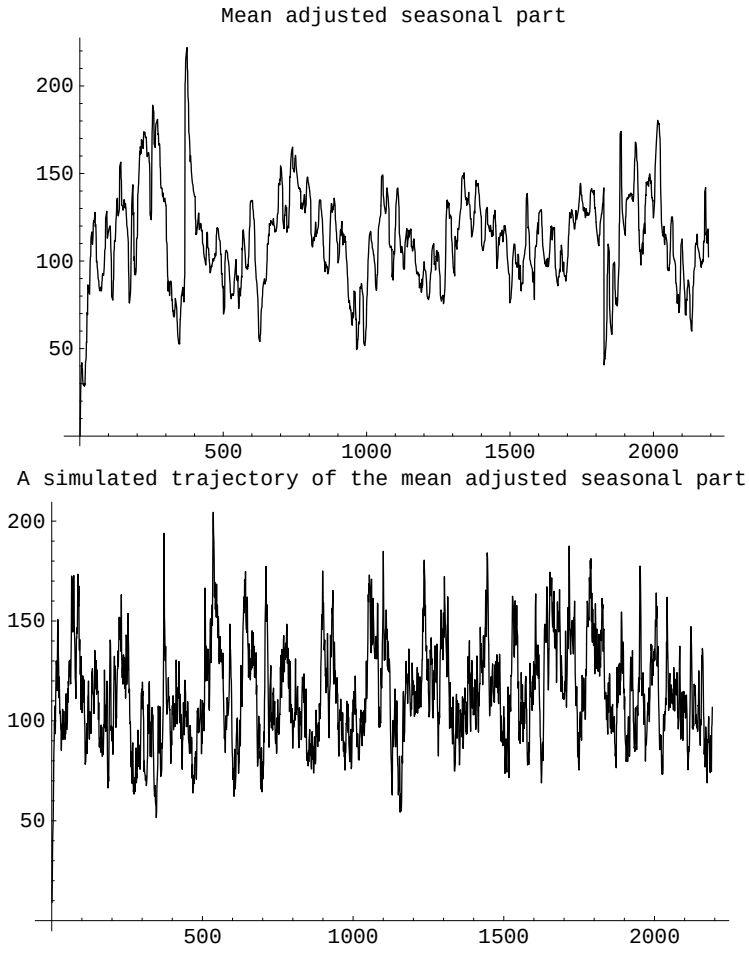


Figure 49: Upper: Mean adjusted seasonal component. Lower: CIR trajectory.

As we can see the KS value is $0.047 \geq 0.03$. The CIR fit is rejected. Although $0.047 - 0.03 = 0.017$ is not big (we could possibly have a good fit for a lower significance level), and let us pretend for a second that the fit is not rejected, this is still not a guarantee that everything is ok. It could be the case that the finite dimensional distributions, of dimension $n \geq 2$, do not give a good fit. Hence, since we saw that the 1-dimensional fit was rejected, we chose to not go any further in our analysis and we rejected the fit of the whole CIR process.

Another problematic feature of the CIR process is that it is very hard to simulate (see [9]). The square root in the sde causes problems since it can create complex values. Usually one has to do simulations with dt being very small, $dt = 1/1000$ will be satisfactory. Hence, if the data time scale consists of time unit steps equal to 1 and one wants to replicate the data consisting of n data points by simulations, one has to sample $1000n$ CIR steps. Now, imagine that you would have to simulate 1000 samples, each consisting of $2192 \cdot 1000$ data points, you would need weeks in order to finish the simulations.

12.2 Vasicek

One diffusion process which, considering its characteristics, is a good candidate for modelling the mean adjusted detrended deperiodized seasonal component is the *Vasicek interest rate model* $\{R(t)\}_{t\geq 0}$, given by

$$dR(t) = (\alpha - \beta R(t))dt + \sigma dB(t),$$

where $\alpha \in \mathbb{R}$ and $\beta, \sigma > 0$ are parameters. The solution to this equation

$$R(t) = \mathrm{e}^{-\beta t} R(0) + \frac{\alpha}{\beta}\left(1 - \mathrm{e}^{-\beta t}\right) + \sigma \mathrm{e}^{-\beta t} \int_0^t \mathrm{e}^{\beta s}\, dB(s).$$

is found by applying Ito's formula to $f(t, X(t))$, where

$$f(t,x) = \mathrm{e}^{-\beta t} R(0) + \frac{\alpha}{\beta}\left(1 - \mathrm{e}^{-\beta t}\right) + \sigma \mathrm{e}^{-\beta t} x$$

and

$$X(t) = \int_0^t \mathrm{e}^{\beta s}\, dB(s).$$

It follows that $R(t)$ is a Gaussian random variable with moments

$$\begin{aligned} \mathbf{E}[R(t)] &= \mathrm{e}^{-\beta t} R(0) + \alpha\left(1 - \mathrm{e}^{-\beta t}\right)/\beta, \\ \mathbf{Var}[R(t)] &= \sigma^2\left(1 - \mathrm{e}^{-2\beta t}\right)/(2\beta). \end{aligned}$$

One desirable property of this model is its so called mean reverting behaviour.

In order to fit the Vasicek process to our data and execute the likelihood based goodness-of-fit test previously described, we need the transition density function

$$f_{R(t)|R(s)}(y|x) = \frac{\sqrt{\beta}}{\sigma\sqrt{\pi}\sqrt{\mathrm{e}^{-2s\beta} + \mathrm{e}^{2t\beta}}} \exp\left\{ t\beta + \frac{[\mathrm{e}^{s\beta}(\alpha + x\beta) - \mathrm{e}^{t\beta}(\alpha + y\beta)]^2}{\beta\sigma^2(\mathrm{e}^{2s\beta} - \mathrm{e}^{-2t\beta})} \right\}.$$

Our parameter estimates are $\hat{\alpha} = -0.048237$, $\hat{\beta} = 0.0237674$, and $\hat{\sigma} = 5.54064$. The value of the log-likelihood function was -6837.28. The corresponding empirical critical value was found to be -6918.41, meaning that the fit was accepted. See Figure 50 below.

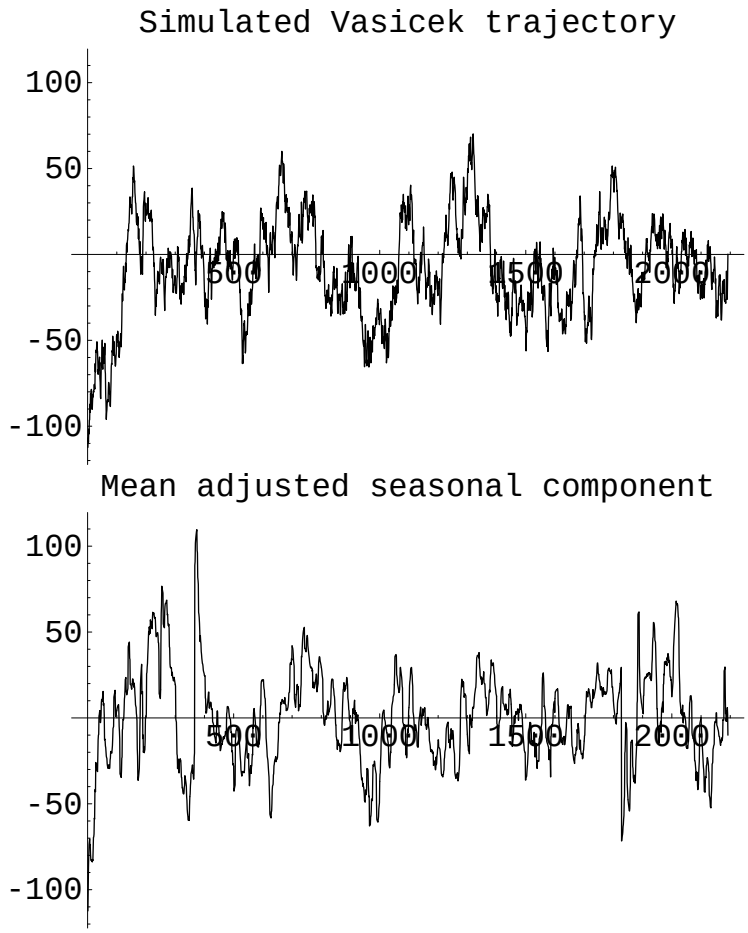


Figure 50: Upper: Simulated trajectory from the Vasicek model.
Lower: The mean adjusted seasonal component.

Note that the Vasicek model is a Gaussian OU process modified to have an exponential type of drift. See [11], p. 150, for a more thorough presentation.

13 Merging of models

According to our modelling, the electricity spot prices follow the model

$$E(t) = D(t) + S(t) + N(t).$$

The model is built up by OU processes used as noise, $N(t)$, a diffusion process used for the seasonal stochastic modelling, $S(t)$, and a deterministic function consisting of all deterministic components extracted, $D(t)$.

The deterministic function $D(t)$ is built up by the extracted deterministic components

$$D(t) = a + bt + \sum_{k=1}^{8}\left[a(k)\cos\left(\frac{tk2\pi}{365}\right) + b(k)\sin\left(\frac{tk\pi}{365}\right)\right] + \sum_{k=0}^{6}\left[c(k)\cos\left(\frac{tk2\pi}{7}\right) + d(k)\sin\left(\frac{tk2\pi}{7}\right)\right].$$

The affine part of D(t) represents the linear trend, where a is the starting value and b is the estimated slope. The second part is the Fourier series which represents the yearly periodicity. Finally the third part is describing the intra-week periodicity, referred to as the Period component.

The seasonal component $S(t)$ is represented by the Vasicek diffusion process

$$dS(t) = (\alpha - \beta)S(t)dt + \sigma dB(t).$$

The OU noise process $N(t)$ is given by

$$dN(t) = -\lambda N(t)dt + dL(t),$$

where either

$$N(t) - \mathrm{e}^{-\lambda}\, N(t-1) \overset{D}{=} \mathrm{Nig}(\alpha, \beta, \delta, \mu), \quad \text{t=1,...,n,}$$

or

$$N(t) - \mathrm{e}^{-\lambda}\, N(t-1) \overset{D}{=} \mathrm{Meixner}(\alpha, \beta, \delta, \mu), \quad \text{t=1,...,n.}$$

13.1 CIR with different OU noises

Just for curiosity we tried to simulate the final model using CIR instead of using Vasicek, before finally simulating the model using Vasicek. It could still be the case that the trajectories look similar to the original spot prices. Hence, $S(t)$ is defined by

$$dS(t) = \beta(\alpha - S(t))dt + \sqrt{S(t)}dB(t).$$

Figure 51 shows 11 trajectories of our model consisting of the seasonal component, $S(t)$, where it is a CIR process, the previously defined deterministic components $D(t)$, and the noise being a NIG OU process. The last graph in the fourth

row and the third column shows the electricity spot prices we are trying to model, for the year 1996.

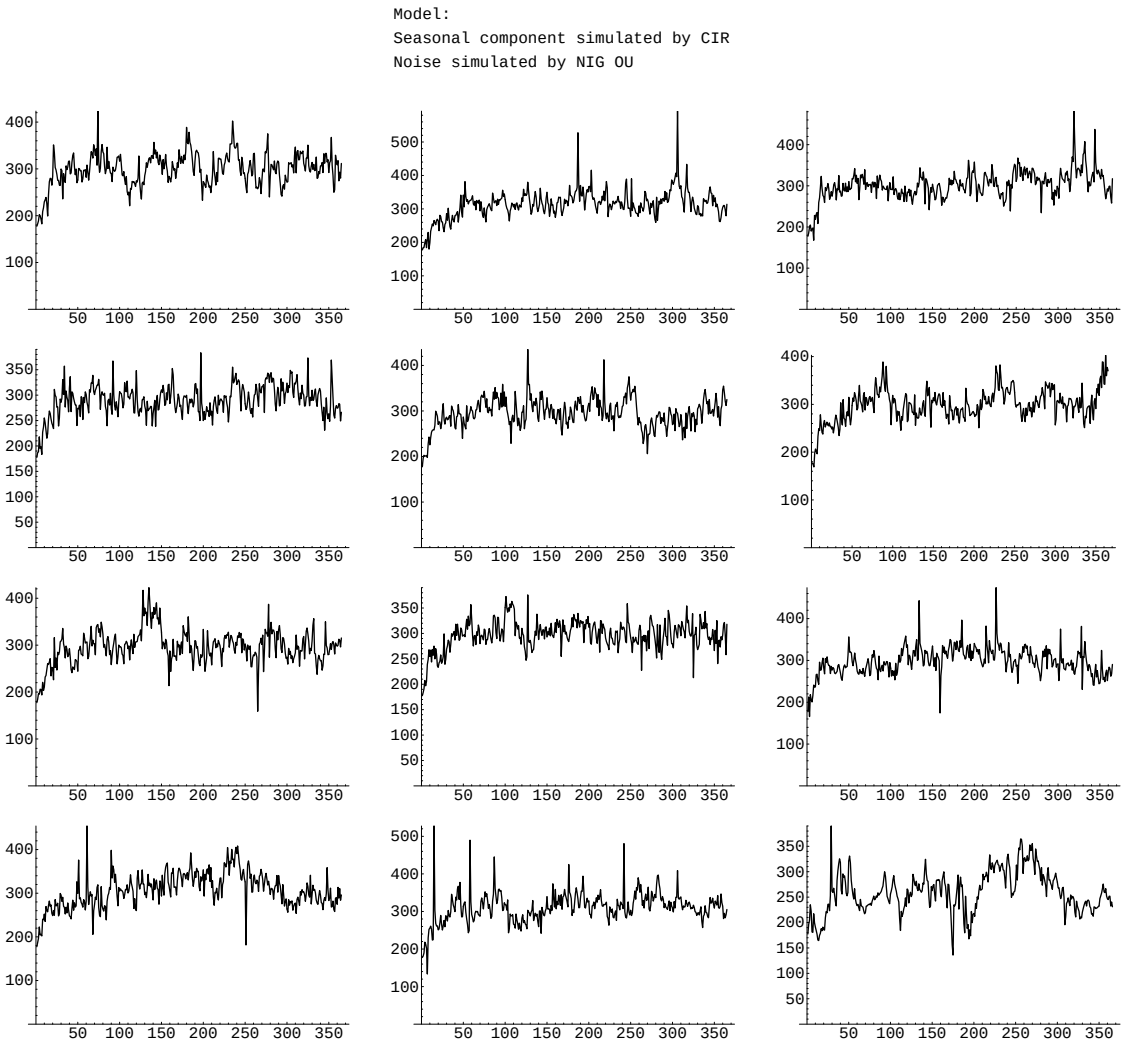


Figure 51: Trajectories of the model built up by a CIR process and a NIG OU process. The electricity spot prices, year 1996, are shown in the 4:th row and the 3:rd column.

Figure 52 below shows 11 trajectories for modelling the year 2001. The last graph in the fourth row and the third column shows the electricity spot prices we are trying to model, for the year 2001.

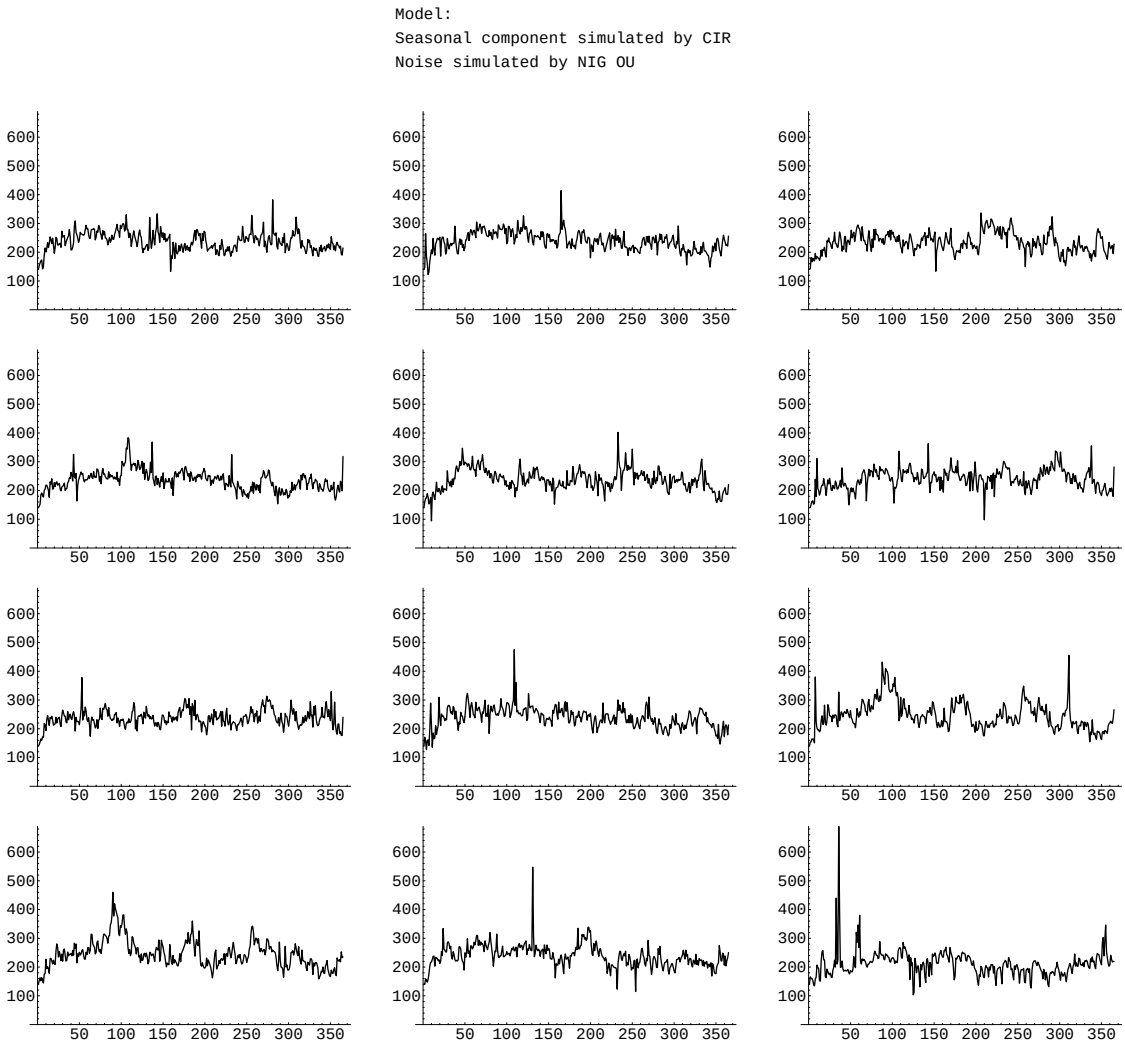


Figure 52: Trajectories of model built by up a CIR process and a NIG OU process. The electricity spot prices, year 2001, are shown in the 4:th row and the 3:rd column.

Figure 53 below shows 11 trajectories of our model consisting of the seasonal component, $S(t)$, where it is a CIR process, the previously defined deterministic components $D(t)$, and the noise being a Meixner OU process. The last graph in the fourth row and the third column shows the electricity spot prices we are trying to model, for the year 1996.

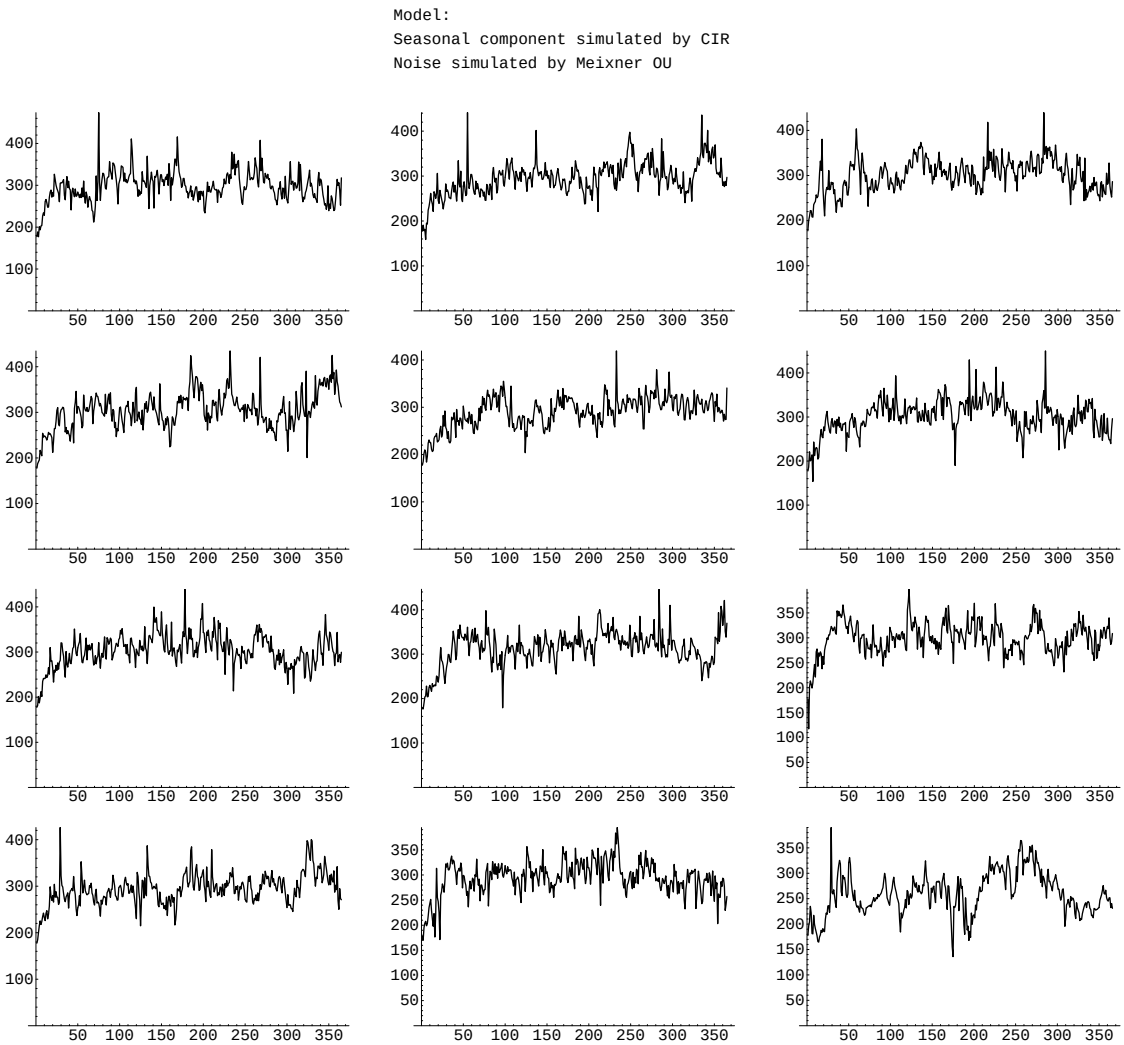


Figure 53: Trajectories of model built up by a CIR process and a Meixner OU process. The electricity spot prices, year 1996, are shown in the 4:th row and the 3:rd column.

Figure 54 below shows 11 trajectories for modelling the year 2001. Again the last graph in the fourth row and the third column shows the electricity spot prices we are trying to model, for the year 2001.

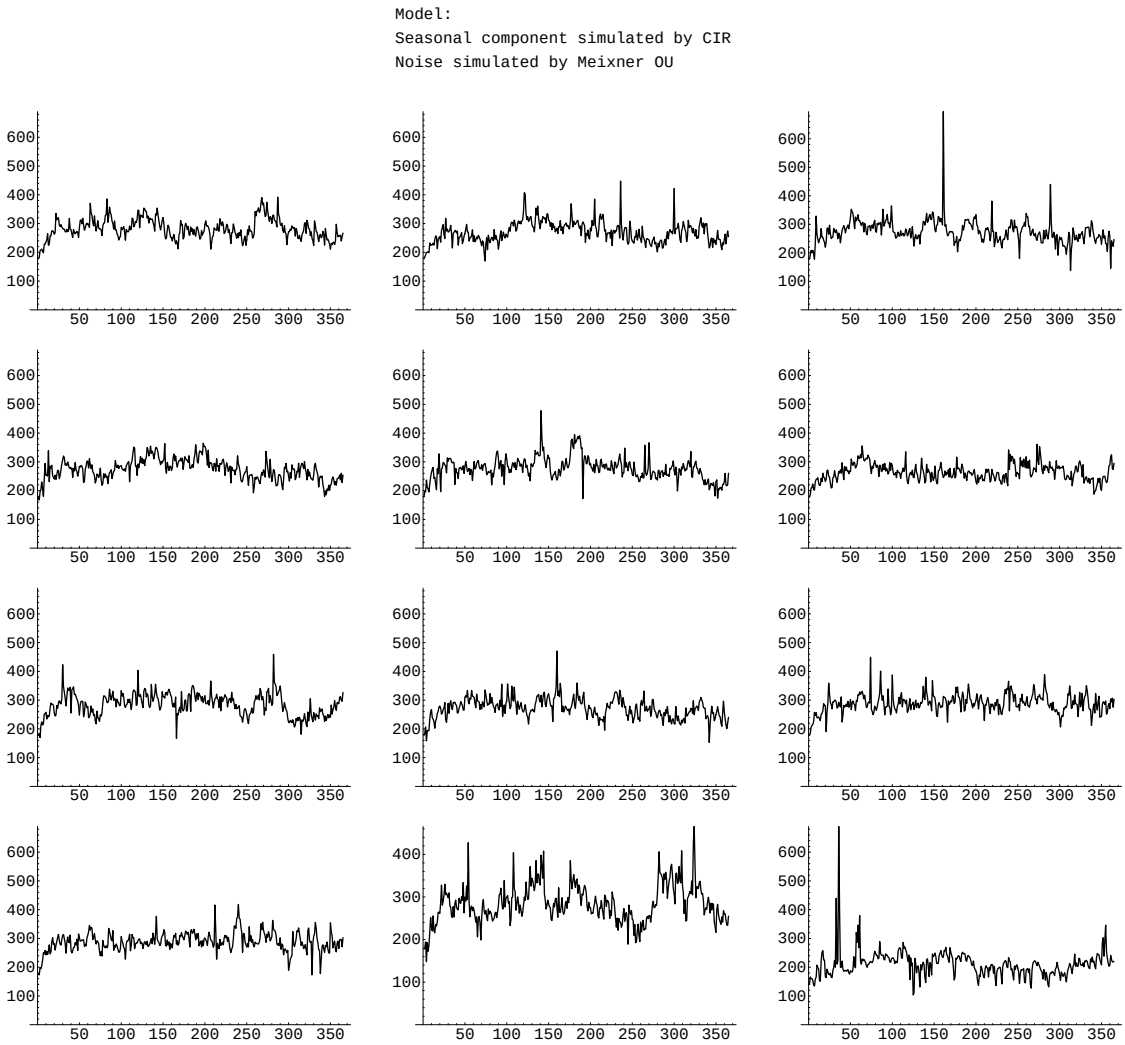


Figure 54: Trajectories of model built up by a CIR process and a Meixner OU process. The electricity spot prices, year 2001, are shown in the 4:th row and the 3:rd column.

As one can see, despite the fact that the CIR process did not pass the goodness-of-fit test, using it for describing $S(t)$, it did a good job. It managed to create trajectories which did not deviate too much from the looks of the real electricity spot prices. It really fulfills the graphical demands one can put on a process trying to fit data.

13.2 Vasicek with different OU noises

Figure 55 below shows 11 trajectories of our model consisting of the seasonal component, $S(t)$, where it is a Vasicek process, the previously defined deterministic components $D(t)$, and the noise being a NIG OU process. The last graph in the fourth row and the third column shows the electricity spot prices we are trying to model, for the year 1996.

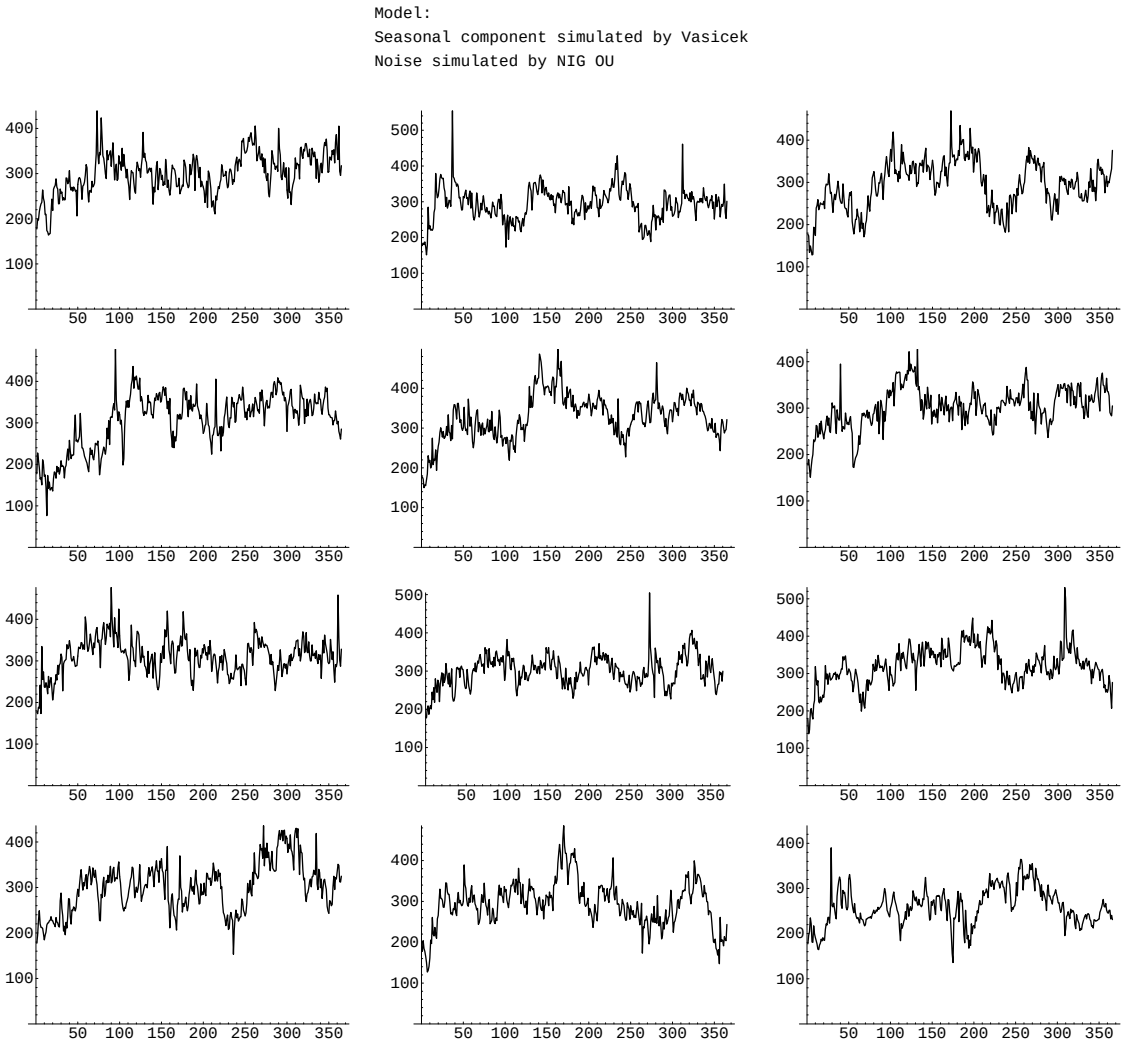


Figure 55: Trajectories of the model built up by a Vacisek process and a NIG OU process. The electricity spot prices, year 1996, are shown in the 4:th row and the 3:rd column.

Figure 56 below shows 11 trajectories for modelling the year 2001. The last graph in the fourth row and the third column shows the electricity spot prices we are trying to model, for the year 2001.

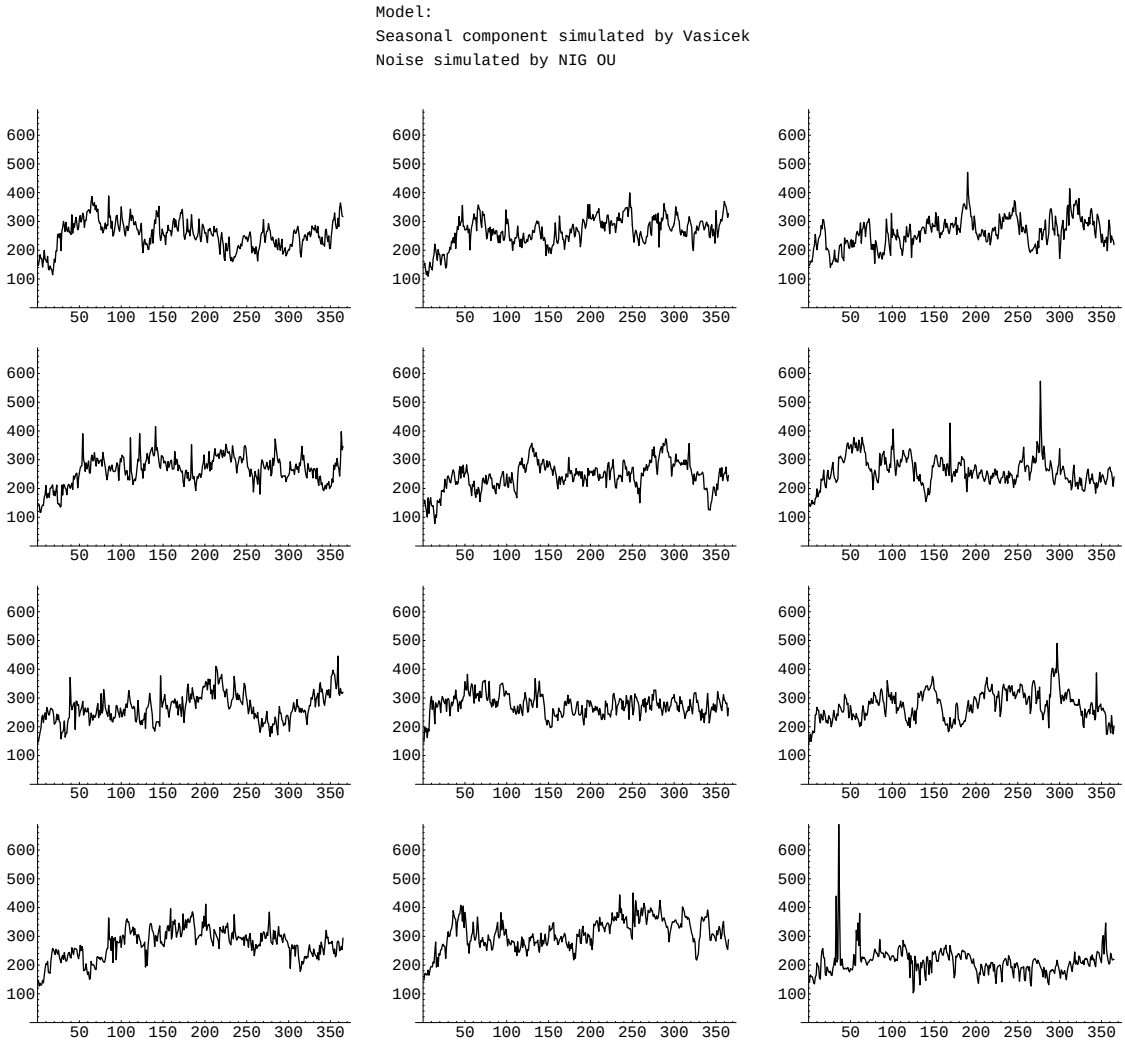


Figure 56: Trajectories of the model built up by a Vacisek process and a NIG OU process. The electricity spot prices, year 2001, are shown in the 4:th row and the 3:rd column.

Figure 57 below shows 11 trajectories of our model consisting of the seasonal component, $S(t)$, where it is a Vasicek process, the previously defined deterministic components $D(t)$, and the noise being a Meixner OU process. The last graph in the fourth row and the third column shows the electricity spot prices we are trying to model, for the year 1996.

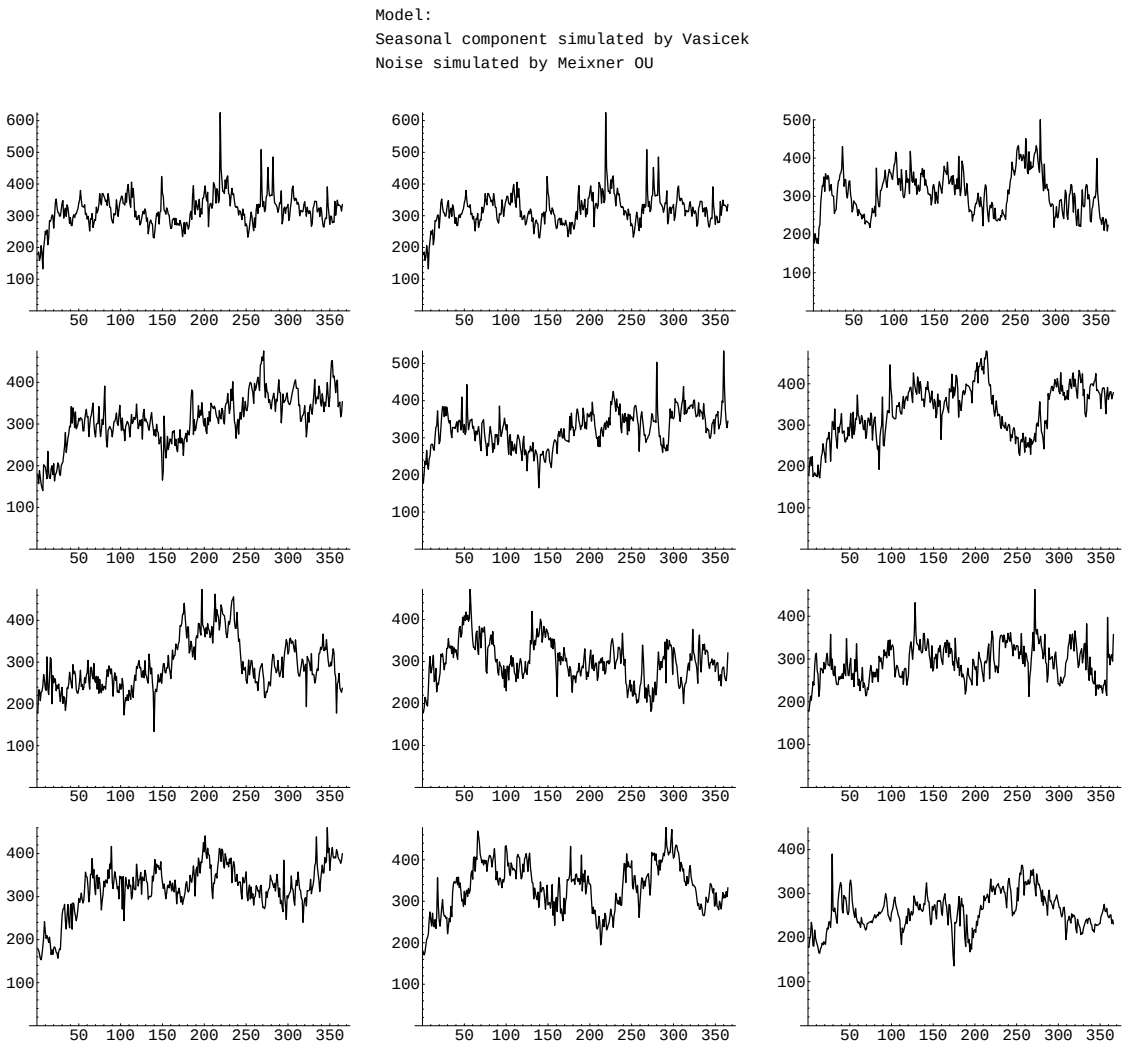


Figure 57: Trajectories of the model built up by a Vacisek process and a Meixner OU process. The electricity spot prices, year 1996, are shown in the 4:th row and the 3:rd column.

Figure 58 below shows 11 trajectories for modelling the year 2001. The last graph in the fourth row and the third column shows the electricity spot prices we are trying to model, for the year 2001.

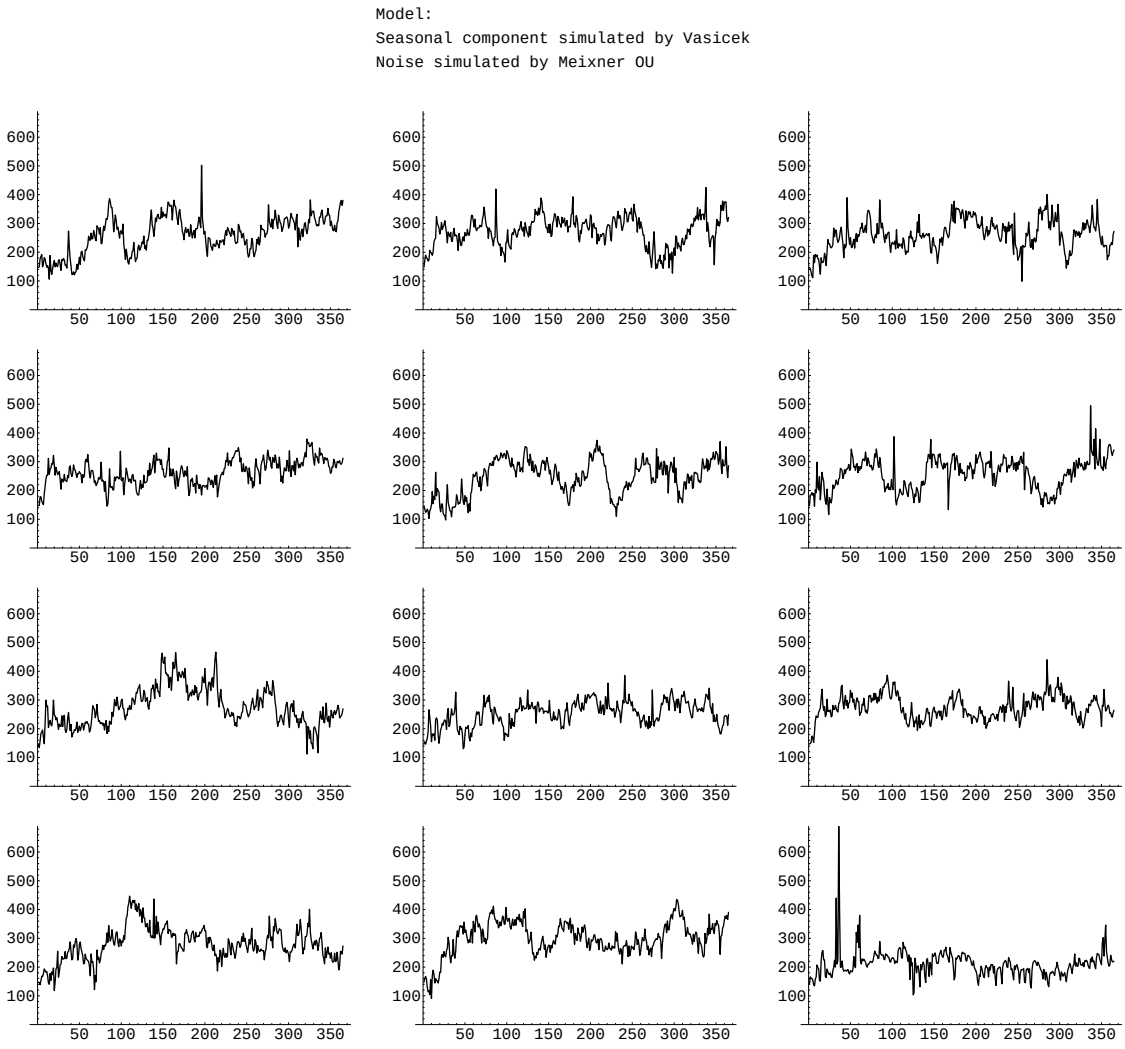


Figure 58: Trajectories of the model built up by a Vacisek process and a Meixner OU process. The electricity spot prices, year 2001, are shown in the 4:th row and the 3:rd column.

14 Afterthoughts

Today it is very crucial to have a good model for predicting electricity spot prices and to analyze risk. We would like to give some points where further research can be made on the hot subject of modelling electricity spot prices. This is not an easy task. This needs a deep understanding of the electricity spot market and the factors that affects electricity spot prices.

- Searching for weather data and try to use them in the spot price modelling.
- Derive expressions for futures and forwards.
- For modelling electricity spot prices from year 2005 and on, so you need to be able to model the emission allowances thus needing their historical data.
- Analyzing and modelling of the extreme observations (spikes).
- Trying to model the noise using other distributions in the OU processes. The distributions we have in mind are for instance mixture models, which could be normal mixture models. The reason why one would want to apply these distributions is to achieve distributions with heavier tails than the semi-heavy tails seen in this thesis.

References

[1] Aas, K. and Haff, I.H. (2006). The generalized hyperbolic skew Student's t-distribution. *Journal of Financial Econometrics Advance Access* **4** 275–309.

[2] Albin, J.M.P. (2002). Fifteen Lectures, From Lévy Processes to Semimartingales. Unpublished manuscript.

[3] Albin, P. (2003). *Stochastic Processes*. Studentlitteratur. [In Swedish.]

[4] Benth, F.E., Groth, M. and Kettler, P.C. (2005). A quasi-Monte Carlo algorithm for the normal inverse Gaussian distribution and valuation of financial derivatives. Preprint, Department of Mathematics, University of Oslo.

[5] Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, 2nd Ed. Springer.

[6] Chatfield, C. (1996). *The Analysis of Time Series, an Introduction.* Chapman & Hall.

[7] Geman, H. (2005). *Commodities and Commodity Derivatives, Modeling and Pricing for Agriculturals,Metals and Energy.* Wiley.

[8] Mikosch, T. (2006). Modeling dependence in discrete time. Unpublished manuscript.

[9] Muszta, A. (2005). Contributions to numerical solution of stochastic differential equations. PhD-thesis, Chalmers University of Technology.

[10] Rose, C. and Smith, M.D. (2002). *Mathematical Statistics with Mathematica.* Springer.

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

[12] `www.vattenportalen.se/fov_problem_vattenkraft.htm` [In Swedish.]

[13] `www.svenskfjarrvarme.se/download/577/Energipolitik-97.pdf` [In Swedish.]