CHALMERS | GÖTEBORG UNIVERSITY

MASTER'S THESIS

Modeling electricity prices with seasonal long memory time series

DAVID KÄLLBERG

Department of Mathematical Statistics
CHALMERS UNIVERSITY OF TECHNOLOGY
GÖTEBORG UNIVERSITY

Göteborg, Sweden 2008

Thesis for the Degree of Master of Science (30 ECTS)

Modeling electricity prices with seasonal long memory time series

David Källberg

CHALMERS | GÖTEBORG UNIVERSITY

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

Abstract

In this thesis we model the system price from the Nordic power exchange with the seasonal autoregressive fractionally integrated moving average (SARFIMA) process. The innovation process is a generalized autoregressive conditionally heteroskedastic (GARCH) process, or an exponential GARCH (EGARCH) process. The autoregressive conditional mean function is made periodic in order to capture the weakly pattern. All parameters are estimated simultaneously by the method of approximate maximum likelihood. The model effectively captures the dependencies in the data.

Acknowledgements

First, I would like to thank my supervisor Patrik Albin for his encouragement, advice and for helping me with important things. I would also like to thank Alain Angeralides at Östkraft Elhandel for his guidance, inspiration and generosity. I would like to thank my good friend Fredrik Lindgren, because his presence has meant a lot to me during my time in Gothenburg and at Chalmers. Also to be mentioned is the lovely Clara. I am very grateful for everything.

Contents

1 Introduction

During the last two and a half decades, many countries world wide have decided to deregulate and liberalize their power sectors. The general motivation is a belief that introduction of market forces can stimulate efficiency when it comes to investment and technical development. The energy sector has traditionally been regarded as a natural monopoly, but because of transmission improvements and development of new generation technologies, the restructuring has become practically possible.

A reform in Chile in 1982 paved the way for the process. The idea was to separate generation and distribution companies. In 1986, the privatization began in large scale and led to the creation of a wholesale power trading platform.

Great Britian followed the Chilean example and reorganized their power sector in 1990. The power market only included England and Wales initially, but from 2005 Scotland as well. In 1992, the Nordic market opened in Norway and later on in Sweden 1996, Finland 1998 and Denmark 1999. In 1994, markets in Australia (Victoria and South Wales) began their operation. This was 1998 followed by the launch of the Australian National Electricity Market (NEM). About the same time, New Zealand restructured their power sector with the official opening of a market in 1996. A number of markets in North America began operating in the late 90s', among them New York. In 1998, California followed and Texas and Alberta in Canada three years later. The trend of an increasing number of liberalized markets is most notable in Europe, but also visible world wide.

1.1 Nord Pool

Nord Pool is the name of the Nordic power market. It was the first multinational power exchange in the world. Nord Pool consists of the following parts:

- The physical market *Elspot*. In order to participate in this market, a connection that makes it possible to deliver to or take power from the main grid is required. Of the total power consumed in the Nordic region, about 40 % is traded on Elspot.
- The hour-ahead market *Elbas*, that is operational in Sweden, Finland and the eastern part of Denmark. Here, Elspot participants can adjust imbalances in their positions up to 2 hours prior to delivery.
- The financial market *Eltermin*. Here, power derivatives such as forwards, futures and options are being traded.

The participants in Nord Pool include generators, suppliers, retailers, traders, financial institutions and large consumers.

Nord Pool is often referred to as a successful market. The main reason for this is its fragmented structure, with over 350 generation companies. For example, the market share of Nord Pool's largest player Vattenfall is only about 20%. Another explanation is the storage ability and production flexibility that comes from the large amount of hydropower available in the Nordic region.

1.2 Price setting

Despite its name, Elspot is not really a spot market, but rather a day-ahead market. Here, one-hour-long physical power contracts are traded, i.e., the price is fixed for each hour. Every day before 12 pm, the participants informs the market administrator how much electric power and to what price, they want to buy or sell a specific hour the following day. Volume histograms $V_{\text{sell}}(p)$ and $V_{\text{buy}}(p)$ are then created for each hour, denoting the total amount of electricity that the participants wants to sell and buy as a function of the price p. Now, if $V_{\text{buy}}(P) = V_{\text{sell}}(P)$ for some price P, then P is defined as the *spot price*, also called the *system price*. Byers with a bid $p \geq P$ and sellers asking a price $p \leq P$ will get a transaction at the system price. These are the only cases where transactions will take place. However, if the system price does not exist, then there will be no transactions at all.

Now, the system price could result in grid congestions, so called *bottlenecks*. In this case, Nord Pool computes area prices that differ from the system price. The idea behind this price adjustment is to avoid that the limited capacity of the transmission grid is exceeded.

2 Stylized facts about electricity prices

We will now review some of the characteristic features of electricity spot prices

2.1 Seasonality

Electricity cannot be economically stored. There is therefore a need to keep production and consumption in realtime balance. As a result of this, seasonal and cyclical fluctuations in demand are reflected in the spot price. There is an intra-day variability, that comes from the behavior of the consumers. Also, the business week causes an intra-week variation, mainly due to weekend effects. Furthermore, the temperature dependence of demand results in a within-year seasonality.

Most of the seasonal variations in the price comes from the demand side. However, at some markets the supply is also affected by seasonality. For example, in the Nordic region, large amounts of hydropower are traded and water reservoir levels are dependent on snow melting and precipitation. Hence, temperature and weather affects both supply and demand and therefore the price.

2.2 Mean reversion

Under extreme market conditions, large price fluctuations can in the short run be observed in the power market. However, in the long run, supply will be adjusted and the price will move towards the cost of producing electricity. Some mean reversion can also be explained by the temperature dependent demand curve.

2.3 Volatility

Electricity spot prices are very volatile. If we use the classical notion of volatility, the standard deviation of log-returns, the volatility of the system price from Nord Pool that we model in this thesis is about 7%. This can be compared with the volatility of very volatile stocks which is under 4%. Volatility clustering is also present in most electricity price series.

2.4 Jumps and spikes

A spike is a large price jump followed by a quick return to the same level. Some typical spikes in our price data can be seen in Figure 1 below. The spike behavior is often explained by the non-storability of electricity, since shocks in supply or demand cannot be smoothened. However, the size of some jumps are so large that another explanation have been suggested: The spikes are caused by strategies of bidders. For many participants, a sufficient supply of electricity is so important that they are willing to pay very high prices to maintain it. Accordingly, agents place bids at the highest allowed level, and since the suppliers know about this strategy, they respond in such a way that they maximize their profit and the price gets very high.

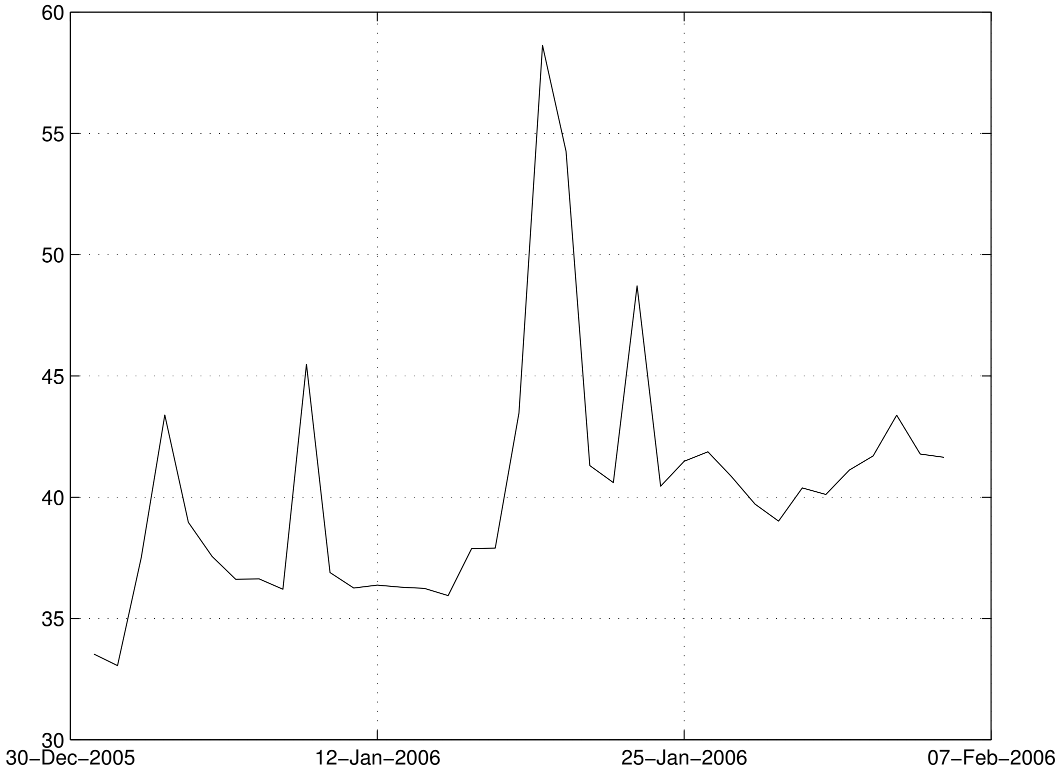


Figure 1: Some spikes in the price data.

3 Time series

In this section we introduce the time series we use to model our price data. In particular, we make use of classical time series analysis, where dependencies are measured via autocovariances or autocorrelations. The autocovariance function (ACVF) and autocorrelation function (ACF) of a stationary stochastic process $(X_t)_{t\in\mathbb{Z}}$ are given by

$$\gamma_X(h) = \mathrm{Cov}(X_0, X_h) \quad \text{and} \quad \rho_X(h) = \frac{\mathrm{Cov}(X_h, X_0)}{\mathrm{Var}(X_0)},$$

respectively. Their sample counterparts are given by the estimators

$$\gamma_{n,X}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(X_t - \overline{X})(X_{t+|h|} - \overline{X}) \quad \text{and} \quad \rho_{n,X}(h) = \frac{\gamma_{n,X}(h)}{\gamma_{n,X}(0)},$$

respectively, where $\overline{X}$ is the sample mean.

The most simple test of the whiteness of a sample $\{X_1, \ldots, X_n\}$ is to plot the sample ACF. Most of the values should fall inside the horizontal lines $\pm 1.96/\sqrt{n}$, which constitute the 95% asymptotic confidence band for a sample of standard Gaussian white noise.

Another test of whiteness is based on the Ljung-Box statistic, given by

$$Q(h) = n(n+2)\sum_{j=1}^{h}\frac{\rho^2_{n,X}(h)}{n-j}.$$

The distribution of $Q(h)$ can be approximated with the χ^2 distribution with h degrees of freedom. Therefore, we reject the hypothesis that $(X_1, \ldots, X_n)$ comes from a white noise sequence if $Q(h) > \chi^2_{1-\alpha}$, where α is some appropriate level. The dependence can then be further analyzed by calculating $Q(h)$ for e.g., the absolute and squared sample.

3.1 ARMA processes

An important class of time series is autoregressive moving average (ARMA) processes. A zero mean stochastic process (X_t) is said to be an ARMA(p,q) process with if it is stationary and satisfies the difference equation

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q} + Z_t, \tag{1}$$

where $\{\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q\}$ are real numbers and (Z_t) is a white noise sequence called *innovations* or *noise*. Furthermore, we say that a process (X_t) with mean $\mu \in \mathbb{R}$ is ARMA(p,q) if $(X_t - \mu)$ is a zero mean ARMA(p,q) process.

3.2 Seasonal long memory processes

The ACF of an ARMA process (X_t) at lag h converges fast to zero when $h \to \infty$, meaning that there exist some $d > 1$ such that

$$d^h \rho_X(h) \to 0 \quad \text{as } h \to \infty.$$

This is why ARMA processes often are refereed to as *short memory processes*. In this section we introduce a class of processes with a more slowly decaying autocorrelation.

The large amount of hydro power traded on the Nordic power market is the reason for the presence of long memory in the electricity price series; one of the first discoveries of long memory was in connection to river flows.

The backward-shift operator is denoted by $\mathcal{B}^n$ for $n \geq 0$, i.e., $\mathcal{B}^n X_t = X_{t-n}$. For $D > -1$ and $s \in \mathbb{N}$ we define the *seasonal difference operator* $\nabla_s^D(\mathcal{B})$ by the binomial expansion

$$\nabla_s^D(\mathcal{B}) := (1-\mathcal{B}^s)^D = \sum_{k=0}^{\infty} \binom{D}{k} (-\mathcal{B}^s)^k = 1 - D\mathcal{B}^s - \frac{D(1-D)}{2!}\mathcal{B}^{2s} - \dots,$$

where

$$\binom{D}{k} = \frac{\Gamma(1+D)}{\Gamma(1+k)\Gamma(1+D-k)}$$

and $\Gamma(\cdot)$ is the gamma function

$$\Gamma(x) := \begin{cases} \int_0^\infty t^{x-1}e^{-t}dt & \text{for} \quad x \in (0,\infty), \\ \infty & \text{for} \quad x \in (-\mathbb{N}), \\ \Gamma(1+x)/x & \text{for} \quad x \in (-\infty, 0) \setminus (-\mathbb{N}). \end{cases}$$

Now we are ready to introduce seasonal autoregressive fractionally integrated moving average (SARFIMA) processes. Only a subclass of these is discussed, because the general form does not come into play in the empirical analysis.

Definition. *A zero mean stochastic process $(X_t)_{t\in\mathbb{Z}}$ is a seasonal autoregressive fractionally integrated moving average process with period s and degree of seasonal differencing D, denoted by $SARFIMA(0,D,0)_s$, if it satisfies*

$$\nabla_s^D X_t = Z_t \quad \textit{for } t \in \mathbb{Z},$$

where $(Z_t)_{t\in\mathbb{Z}}$ is a white noise sequence.

Now we give som properties of SARFIMA$(0,D,0)_s$ processes: The following result can be found in [3].

Theorem. *Let $(X_t)_{t\in\mathbb{Z}}$ be a SARFIMA$(0,D,0)_s$ process with $-0.5 < D < 0.5$. Then,*

1. *$(X_t)_{t\in\mathbb{Z}}$ is invertible with infinite autoregressive representation given by*

$$\Pi(\mathcal{B}^s)X_t = \sum_{k=0}^{\infty} \pi_k X_{t-sk} = Z_t \quad \textit{and} \quad \pi_k = \frac{\Gamma(k-D)}{\Gamma(-D)\Gamma(k+1)},$$

where $\pi_k \sim k^{-D-1}/\Gamma(-D)$ as $k \to \infty$.

2. *$(X_t)_{t\in\mathbb{Z}}$ is stationary with infinite moving average representation given by*

$$X_t = \Psi(\mathcal{B}^s)Z_t = \sum_{k=0}^{\infty} \psi_k Z_{t-sk} \quad \textit{and} \quad \psi_k = \frac{\Gamma(k+D)}{\Gamma(D)\Gamma(k+1)},$$

where $\phi_k \sim k^{D-1}/\Gamma(D)$ as $k \to \infty$.

3. $(X_t)_{t\in\mathbb{Z}}$ *has autocorrelation function of order* $k \geq 0$, *given by*

$$\rho_X(sk+\xi) = \begin{cases} \Gamma(1-D)\Gamma(k+D)/(\Gamma(D)\Gamma(k-D+1)) & \text{for} \quad \xi = 0, \\ 0 & \text{for} \quad \xi \in \{1,\ldots,s-1\}, \end{cases} \tag{2}$$

where $\rho_X(sk) \sim \Gamma(1-d)k^{2d-1}/\Gamma(D)$ *as* $k \to \infty$.

When $D > 0$ we say that $(X_t)_{t\in\mathbb{Z}}$ have *seasonal long memory*.

3.3 GARCH processes

Let us now take a look at the white noise sequence $(Z_t)_{t\in\mathbb{Z}}$ in the preceding section. This can be modeled as an i.i.d. sequence. However, often when it comes to financial time series, such error processes exhibit conditional heteroskedasticity. This means that the variance conditioned on the past, $\mathrm{Var}(Z_t|Z_{t-1}, Z_{t-2}, \ldots)$, changes over time. One of the most widely used models to capture this property is the generalized autoregressive conditional heteroskedastic (GARCH) model of Bollerslev [2]: A stochastic process $(X_t)_{t\in\mathbb{Z}}$ is a GARCH(p,q) process if it satisfies

$$X_t = \sigma_t Z_t \quad \text{for } t \in \mathbb{Z}, \tag{3}$$

where $(Z_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence of random variables with $\mathbb{E}Z_0 = 0$ and $\mathrm{Var}(Z_0) = 1$ and $(\sigma_t)_{t\in\mathbb{Z}}$ is a nonnegative process satisfying the recursive equation

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 : \tag{4}$$

Here the constants $\{\omega, \alpha_1, \ldots, \beta_1, \ldots\}$ are required to be nonnegative in order to ensure that σ_t remains nonnegative.

The structure $p = q = 1$ is sufficient for our purposes. In this case, a weakly stationary solution of (3)-(4) exists if $\omega > 0$ and $\alpha + \beta < 1$.

The simple structure of GARCH(p,q) processes imposes some limitations on them. Firstly, the nonnegativity constraints can cause troubles when it comes to estimations. Secondly, is reasonable from an economics point of view that positive shocks in a market influence the conditional variance in a different way than negative. Negative changes in e.g., stock returns, tend to be associated with increases in volatility and vice versa. This so called *leverage effect* cannot be captured by the symmetry in equation (4). Because of these drawbacks, we next introduce another GARCH-type process.

3.4 Exponential GARCH

The exponential GARCH (EGARCH) process of Nelson [7] is in its simplest form given by

$$X_t = \sigma_t Z_t \quad \text{for } t \in \mathbb{Z}, \tag{5}$$

where $(\sigma_t)_{t\in\mathbb{Z}}$ satisfies

$$\log \sigma_t^2 = \omega + \alpha|Z_{t-1}| + \gamma Z_{t-1} + \beta \log \sigma_{t-1}^2 \quad \text{for } t \in \mathbb{Z} \tag{6}$$

with parameter values $\omega, \alpha, \gamma \in \mathbb{R}$ and $|\beta| < 1$. Note that when $\gamma > 0$, the volatility responds asymmetrically to positive and negative values of (Z_t), making the EGARCH process able to capture leverage effects. Moreover, it is easy to see that the sequence $(\log \sigma_t^2)_{t\in\mathbb{Z}}$ constitutes an ARMA(1,0) process with mean $\mu = (\omega + \alpha\mathbb{E}|Z_0|)/(1-\beta)$ and white noise sequence $(\gamma Z_{t-1} + \alpha(|Z_{t-1}| - \mathbb{E}|Z_0|))$. It follows from the theory of ARMA processes (see [4]), that the unique stationary solution of (6) is given by

$$\log \sigma_t^2 = \alpha(1-\beta)^{-1} + \sum_{k=0}^{\infty} \beta^k (\gamma Z_{t-1-k} + \alpha|Z_{t-1-k}|).$$

4 Distributions

4.1 Conditional distributions

The use of conditional maximum likelihood to estimate our model require that we specify a conditional density. Besides the standard normal, we make use of two other densities:

- The Student-t distribution is often better than then Gaussian distribution to capture the observed kurtosis in financial data. Its density function, normalized to have unit variance, is given by

$$f_\nu(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma(\frac{\nu}{2})} \times \frac{1}{(1+x^2/(\nu-2))^{(\nu+1)/2}},$$

  where $\nu > 2$ is the shape parameter.

- The generalized error distribution (GED) was suggested by Nelson [7]. In the case of unit variance, its density function can be expressed as

$$f_\nu(x) = \frac{\nu}{\lambda_\nu 2^{1+1/\nu}\Gamma(1/\nu)} \, \mathrm{e}^{-\frac{1}{2}|x/\lambda_\nu|^\nu} \quad \text{with} \quad \lambda_\nu = \sqrt{\frac{\Gamma(\frac{1}{\nu})}{2^{2/\nu}\Gamma(\frac{3}{\nu})}},$$

  where $\nu > 0$ is the tail-thickness parameter. When $\nu < 2$, the tails are heavier than those of a normal distribution and when $\nu > 2$, they are lighter. For $\nu = 2$ the standard normal occurs and when $\nu = 1$ the Laplace distribution results.

4.2 Heavy-tailed and skewed distributions

Sometimes it is useful to investigate some property of a data set by fitting different distributions to it. In this section we introduce two classes of distributions that often are used for this purpose.

4.2.1 Generalized hyperbolic distributions

The Student-t distribution is a member of a more general class of distributions, namely the generalized hyperbolic (GH) distributions. They are defined as having density function

$$\begin{aligned} & f_{\mathrm{GH}}(x;\mu,\delta,\alpha,\beta,\lambda) \\ = & \frac{(\alpha^2-\beta^2)^{\lambda/2}\,(\delta^2+(x-\mu)^2)^{(\lambda-\frac{1}{2})/2}}{\sqrt{2\pi}\alpha^{\lambda-1/2}\delta^\lambda K_\lambda\left(\delta\sqrt{\alpha^2-\beta^2}\right)} K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)\,\mathrm{e}^{\beta(x-\mu)} \end{aligned}$$

with parameters

$$\begin{array}{lllll} \delta \geq 0 & \text{and} & |\beta| < \alpha & \text{if} & \lambda > 0, \\ \delta > 0 & \text{and} & |\beta| < \alpha & \text{if} & \lambda = 0, \\ \delta > 0 & \text{and} & |\beta| \leq \alpha & \text{if} & \lambda < 0, \end{array}$$

where K_λ is the modified Bessel function of the third kind. Here the parameters $\mu \in \mathbb{R}$ and δ describe location and scale, α determines the shape and β the skewness. The parameter λ influence the amount of mass contained in the tails.

The tail behavior of the GH distribution is classified as semi-heavy, which means that the tails are much heavier than those of Gaussian laws, but lighter than the tails of non-Gaussian stable laws. The behavior is described by the following asymptotic relation.

$$f_{\mathrm{GH}}(x;\mu,\delta,\alpha,\beta,\lambda) \sim |x|^{\lambda-1}\,\mathrm{e}^{(\mp\alpha+\beta)x} \quad \text{as } x \to \pm\infty,$$

up to a multiplicative constant .

We make use of the following subclasses of GH.

- For $\lambda = 1$, we obtain the hyperbolic (HYP) distributions.
- For $\lambda = \frac{1}{2}$, we obtain the normal-inverse Gaussian (NIG) distributions. A convenient feature of NIG distributions is that they are closed under convolutions, so that sums of independent NIG random variables are NIG distributed. Most other GH distributions does not share this property.

4.2.2 Stable distributions

Stable distributions are heavy-tailed (except when $\alpha = 2$, see below). The following result can be found in [9] and introduce stable random variables in terms of four alternative but equivalent definitions.

Theorem (Stable random variable). *The following conditions are equivalent and fully characterizes a stable random variable: A random variable X is said to have a stable distribution if*

1. *for any $a, b > 0$, there exist $c > 0$ and $d \in \mathbb{R}$ such that*

$$aX_1 + bX_2 \stackrel{d}{=} cX + d,$$

where X_1 and X_2 are independent copies of X and $\stackrel{d}{=}$ denotes equality in distribution;

2. *for any $n \geq 2$, there exist $c_n > 0$ and $d_n \in \mathbb{R}$ such that*

$$X_1 + X_2 + \ldots + X_n \stackrel{d}{=} c_n X + d_n;$$

3. *X belongs to a domain of attraction, i.e., there exist a sequence of independent and identically distributed random variables $Y_1, Y_2, \ldots$ and sequences numbers $c_n > 0$ and real numbers $d_n \in \mathbb{R}$, such that*

$$\frac{Y_1 + Y_2 + \ldots + Y_n}{c_n} + d_n \stackrel{d}{\Rightarrow} X,$$

where $\stackrel{d}{\Rightarrow}$ denotes convergence in distribution.

4. *there exist parameters $\alpha \in (0,2]$, $\sigma \geq 0$, $\beta \in [-1,1]$ and $\mu \in \mathbb{R}$ such that X has characteristic function*

$$\mathbb{E}\{\mathrm{e}^{itX}\} = \begin{cases} \exp\left\{-\sigma^\alpha |t|^\alpha \left(1 - i\beta\,\mathrm{sign}(t)\tan\frac{\pi\alpha}{2}\right) + i\mu t\right\} & \text{if } \alpha \neq 1 \\ \exp\left\{-\sigma|t|\left(1 + i\beta\frac{2}{\pi}\,\mathrm{sign}(t)\ln|t|\right) + i\mu t\right\} & \text{if } \alpha = 1 \end{cases} . \qquad (7)$$

The most common way to specify a stable distribution is through its characteristic function (7) and we use $X \sim S_\alpha(\beta, \sigma, \mu)$ to denote that specification. Here, α is called the index of stability or tail exponent and determines the rate at which the tails taper off. The Gaussian distribution results when $\alpha = 2$. For $\alpha < 2$, the variance is infinite and when $\alpha > 1$, the mean of the distribution exists. The parameters β, σ and μ are the skewness, scale and location parameters, respectively.

Regarding the tail-behavior of $X \sim S_\alpha(\beta, \sigma, \mu)$ in the non-Gaussian case when $\alpha < 2$, it can be shown (see [9]) that

$$\begin{cases} \lim_{x\to\infty} x^\alpha \mathbb{P}(X > x) &= C_\alpha \frac{1}{2}(1+\beta)\,\sigma^\alpha, \\ \lim_{x\to\infty} x^\alpha \mathbb{P}(X < -x) &= C_\alpha \frac{1}{2}(1-\beta)\,\sigma^\alpha, \end{cases}$$

where

$$C_\alpha = 1 \Bigg/ \left(\int_0^\infty x^{-\alpha} \sin x \, dx\right) = \begin{cases} \dfrac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)} & \text{if } \alpha \neq 1 \\ \dfrac{2}{\pi} & \text{if } \alpha = 1 \end{cases}.$$

Except for the three special cases α=0.5, 1, 2, the stable density and cumulative distribution function do not have closed form expressions. This is a major drawback, because e.g., maximum likelihood methodology becomes infeasible, or at least difficult and computationally demanding. We make use of the quantile method described by McCulloch [5] to estimate the stable parameters.

4.3 Goodness of fit

In order to quantitatively assess goodness-of-fit, we make use of test statistics based on the vertical distance between the fitted distribution function F and the empirical distribution function (edf) F_n of the sample $(X_1, \ldots, X_n)$, given by

$$F_n(x) = \frac{1}{n} \sum_{m=1}^{n} \mathbf{1}_{\{X_m \le x\}},$$

where $\mathbf{1}_{\{.\}}$ is the indicator function. The most basic of these statistics is the Kolmogorov distance (KD), given by

$$\mathrm{KD} = \sup_{x\in\mathbb{R}} |F_n(x) - F(x)|.$$

We also make use of the Anderson-Darling (AD) test statistic, a weighted version of KD which puts more weight in the tails of the distribution. It is given by

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

Critical values of KD and AD are not known for most distributions F. It is possible to estimate such values via bootstrap resampling. Our purpose is only to compare fits, so we do not do this, but a description of the procedure can be found in [1].

5 Modeling

In this section we consider a time series model for the daily average system price from Nord Pool.

5.1 The data

The price period from 2003-01-29 to 2006-04-28 (about 1200 days) is used to estimate the model. These numbers were provided through Nord Pool's market data service [8]. The raw price and its first differences are shown in Figure 2 below and some descriptive statistics are reported in Table 1. The presence of volatility clustering is quite obvious from looking at the first differences in Figure 2. Furthermore, Table 2 below gives a first indication of day of the week effects, even though these statistics are unreliable due to the small sample size and high volatility.

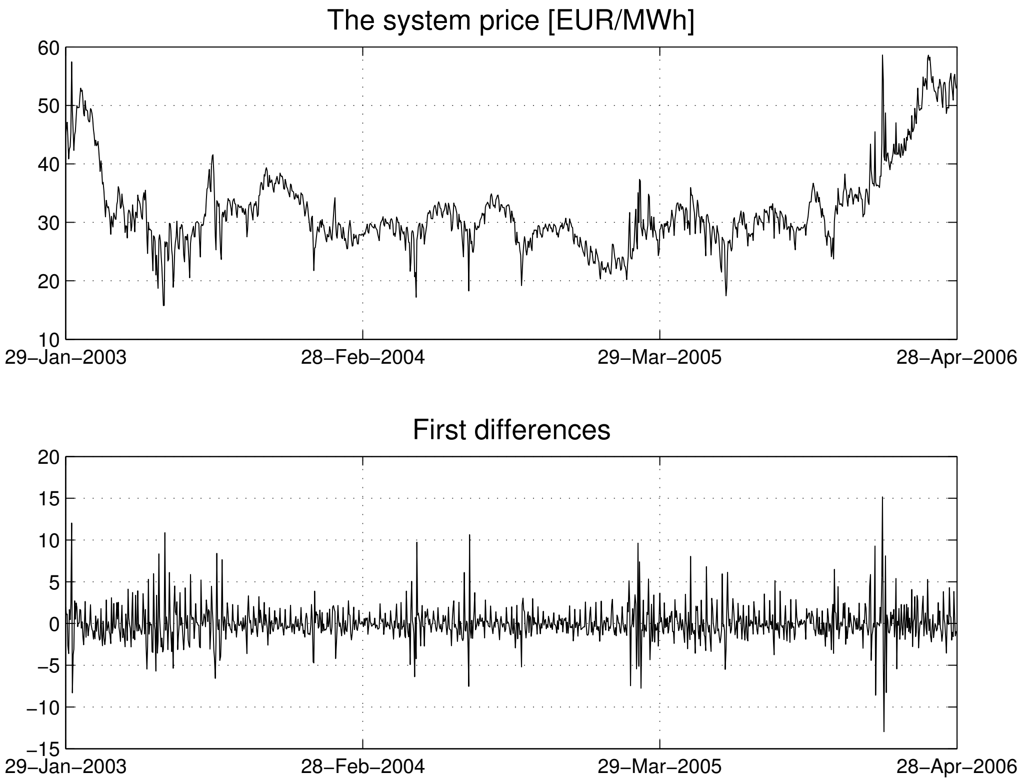


Figure 2: Raw data and first differences.

Table 1: Descriptive statistics of the data.

	Raw data	1st differences
Max	58.63	15.16
Min	15.76	-12.95
Mean	32.1516	0.0067
Variance	53.0133	4.3932
Skewness	1.4273	0.8925
Kurtosis	5.1399	11.1231

Table 2: Mean and standard deviation for first differences.

	Mean	St. Dev.
Sun-Mon	2.7722	2.0195
Mon-Tue	0.0506	1.8521
Tue-Wed	0.0081	1.4438
Wed-Thu	-0.0225	1.8966
Thu-Fri	-0.6323	1.7343
Fri-Sat	-1.3359	1.5845
Sat-Sun	-0.7930	1.2813

The quantile plot in Figure 3 below tells us that we can expect the data to have heavier tails than the normal distribution.

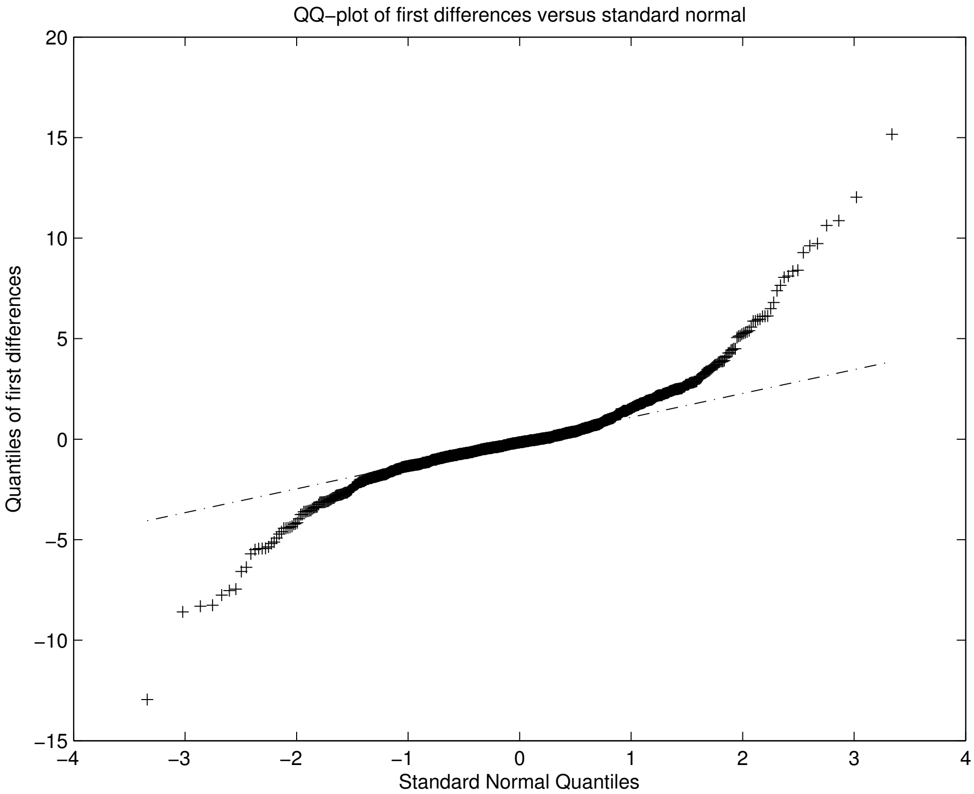


Figure 3: Normal quantile plot of the first differences.

5.2 Preprocessing

The statistics and procedures we use to analyze the data, e.g., the sample ACF and seasonal decomposition, are sensitive to outliers. To identify and remove these, we borrow a procedure from [6] that goes as follows: The time series is divided into seven, one for each day of the week. Next, every value that deviates more than three standard deviations from the mean is replaced by the arithmetic average of its two adjacent values. We do not want this manipulation of the raw data to influence the analysis in any significant manner, so the procedure is only applied once. The resulting 18 outliers does not occur on a particular day of the week.

The annual seasonality in our data is modest, so we do not incorporate it in the time series modeling. Instead, it is filtered out as a part of the preprocessing. In particular, we decompose the outlier treated data (P_t) according to

$$P_t = p_t + S_t,$$

where S_t is the sinusoid

$$S_t = A + Bt + C \sin\left(\frac{t - D}{365}\right)$$

The coefficients $\{A, B, C, D\}$ are calculated using nonlinear least squares. The result is reported in Table 3 below. Clearly, the contribution from the annual seasonality to the variability of the data is small. The estimated seasonal component and preprocessed data can be seen in Figure 4.

Table 3: Estimated parameters of the seasonal component with 95% confidence intervals and coefficient of determination R^2.

		C.I.
A	30.05	(29.27,30.82)
B	0.003192	(0.002064,0.004319)
C	2.807	(2.273,3.341)
D	-58.56	(-70.14,-46-97)
R^2	0.1073	

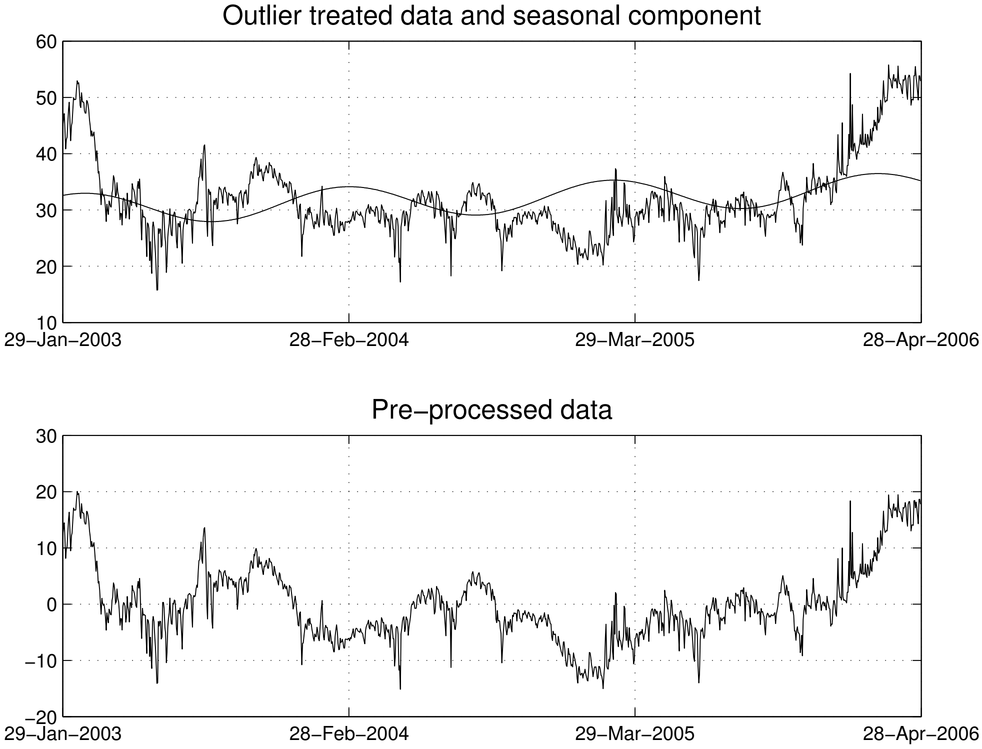


Figure 4: Annual seasonality and preprocessed data.

5.3 The model

In this section we consider the following combination of the time series introduced in Section 3 as a model for the preprocessed data (p_t).

$$\begin{aligned} \nabla_s^D X_t &= \epsilon_t, \\ X_t &= p_t - \mu_t, \end{aligned} \tag{8}$$

where $s = 7$ and (ϵ_t) is a GARCH(1,1) or an EGARCH process with conditional standard deviation process (σ_t). The conditional mean equation is given by

$$\mu_t = \phi_1 p_{t-1} + \ldots + \phi_r p_{t-r} + \sum_{k=1}^{s} \delta_k W_{k,t},$$

where the constants $\{\delta_1, \ldots, \delta_s\}$ represent the weakly pattern and where $(W_{k,t})$ is the daily dummy of day k. We sometimes use the notation δ_{Day} in the sequel to be more specific.

5.3.1 Approximate maximum likelihood estimation

In order to fit the model (8) by the maximum likelihood method, we first specify a density function $g_\psi(x)$, where the extra parameter vector ψ also has to be estimated. To estimate $\theta = \{\phi_1, \ldots, \delta_1, \ldots, \omega, \alpha, \gamma, \beta, D\}$ and ψ we then proceed as follows: The infinite autoregressive polynomial in the left hand side of (8) is truncated and approximations $\hat{\epsilon}_t(\theta)$ and $\hat{\sigma}_t(\theta)$ are extracted. These are plugged into the approximate log-likelihood, given by

$$\mathcal{L}(\Theta) = \sum_{t=r+1}^{T} \log \hat{f}_t(\Theta),$$

where $\Theta = \{\theta, \psi\}$ and

$$\hat{f}_t(\Theta) = \frac{1}{\hat{\sigma}_t(\theta)} g_\psi \left(\frac{\hat{\epsilon}_t(\theta)}{\hat{\sigma}_t(\theta)} \right).$$

The estimate of Θ is now given by

$$\hat{\Theta} = \arg\max \mathcal{L}(\Theta).$$

In particular, as $g_\psi(x)$ we use the standard normal, Student-t and GED densities.

Because of the primitive optimization routines we use to find $\hat{\Theta}$ (Quasi-Newton line search and the simplex method), clever initial values are important. We perform the following steps to find these in the GARCH(p,q) case.

1. Initial values for the conditional mean parameters $\{\phi_1, \ldots, \delta_1, \ldots\}$ are found by linear regression.

2. The sample ACF of the residuals from the regression is then calculated to find D_0 from equation (2).

3. Next, initial values for the GARCH process are found by the MATLAB command `ugarch`.

4. Finally, in the case with a conditional density parameter, the initial value for this is found by the maximum likelihood method.

In the case with the EGARCH process, step 3 is replaced by a linear regression with the conditional variance and residuals from the fitted SARFIMA-GARCH model.

5.4 Numerical results

In this section we report the estimation results of the different versions of (8) that we consider. All of these are re-estimated with insignificant parameters constrained to zero, until only significant ones remain. It holds for all versions that the weekly pattern is only due to weekend effects, meaning that the only significant parameters are δ_{Mon}, δ_{Sat} and δ_{Sun}. However, it is not obvious that δ_{Fri} is zero, but because it is useful to keep the number of parameters down, this is not included. Of the autoregressive parameters, ϕ_1 is of course significant, and then the choice between ϕ_3 and ϕ_5 is somewhat arbitrary, but both cannot be included.

The reported standard errors are calculated from second order numerical derivatives, based on finite differences, and should therefore be interpreted with caution.

5.4.1 SARFIMA

The assumption $\sigma_t = \sigma$ and use of the normal density, gives the estimates reported in Table 4 below. By inspection of the residuals and corresponding quantile plot in Figure 5, we conclude that we deal with fairly heavy tailed data. To further investigate this, we fit the residuals to the hyperbolic, normal-inverse Gaussian and stable distribution. The result can be seen in Table 5. Led by the goodness-of-fit statistics and cdf plots in Figure 6, we conclude that best distribution is the stable, closely followed by the normal-inverse Gaussian.

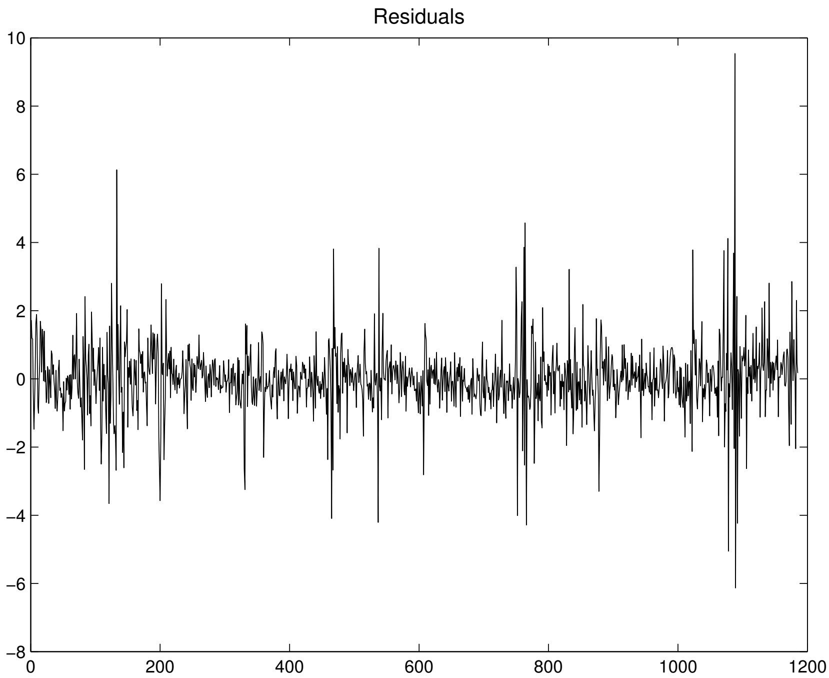
We do not report Ljung-Box values for this case, but the sample ACF plots in Figure 7 gives an indication of linear dependence in the squared residuals, motivating an extension of the model with a GARCH-type process.

Table 4: Parameter estimates for the SARFIMA-normal model.

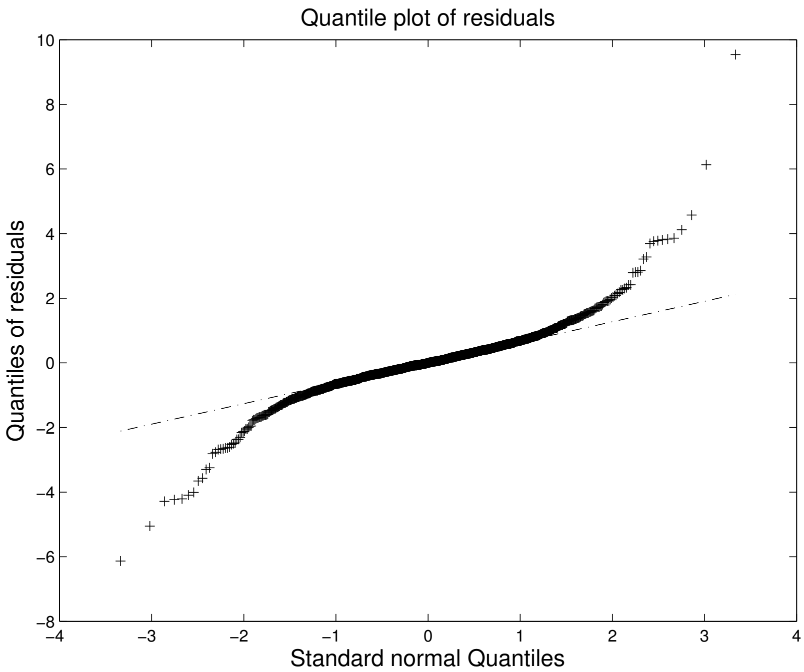
	ϕ_1	ϕ_3	δ_{Mon}	δ_{Sat}	δ_{Sun}	ω	D
Est.	0.8475	0.1203	2.4056	-1.4058	-1.0494	2.6219	0.1346
S.E.	0.0210	0.0207	0.1592	0.1534	0.1571	0.1077	0.0262

Table 5: Parameter estimates and goodness-of-fit statistics.

	α	β	σ, δ	μ	KD	AD
Normal	2	0	1.6192	0.0075	6.4877	105.8438
NIG	0.3902	0.0092	0.9691	-0.0107	3.5226	34.4141
HYP	0.9658	0.0190	0.1169	-0.0290	3.8752	40.8419
Stable	1.4638	0.0247	0.7121	0.0057	3.6681	33.6867

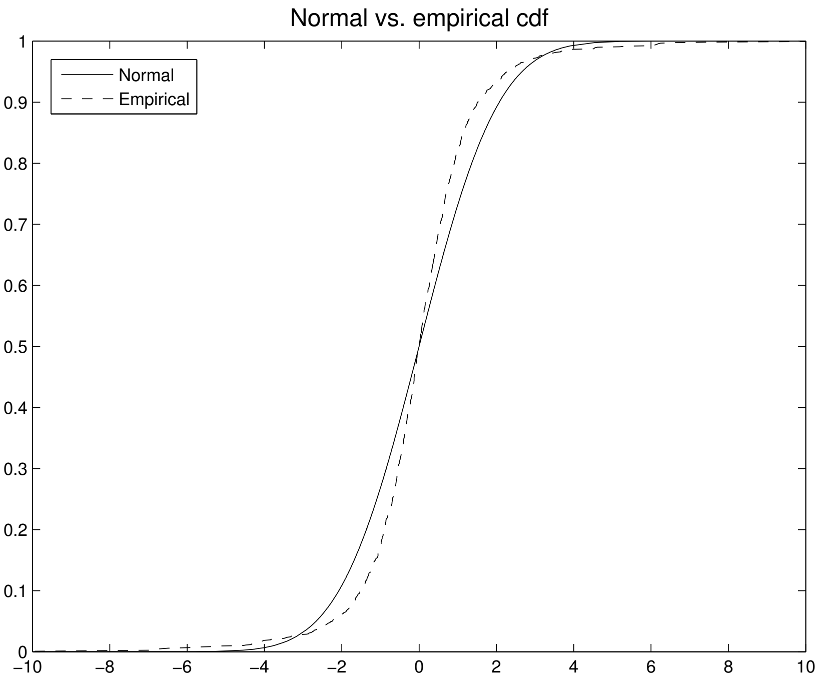


(a) Residuals

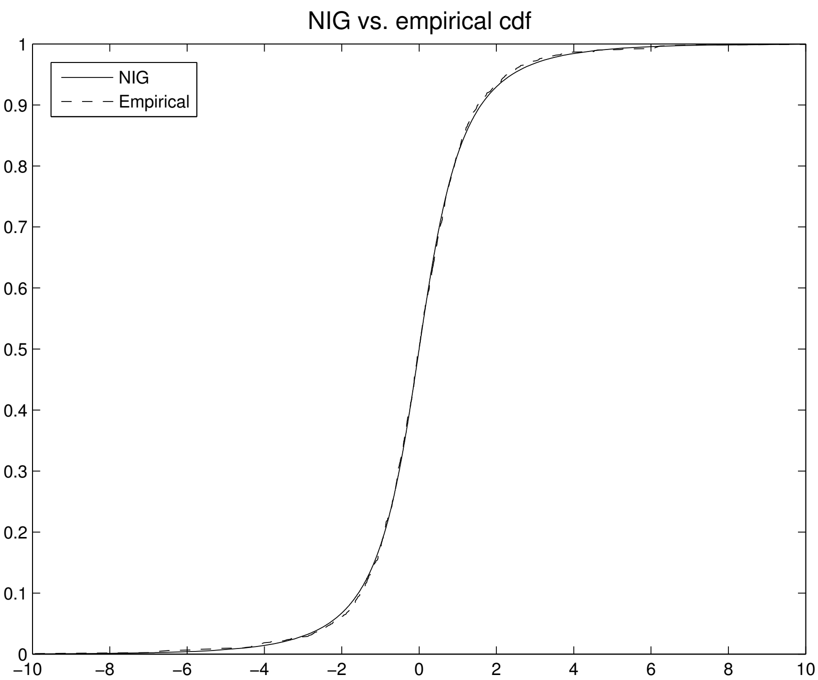


(b) Quantile plot

Figure 5: Residuals and their normal quantile plot in the SARFIMA model.

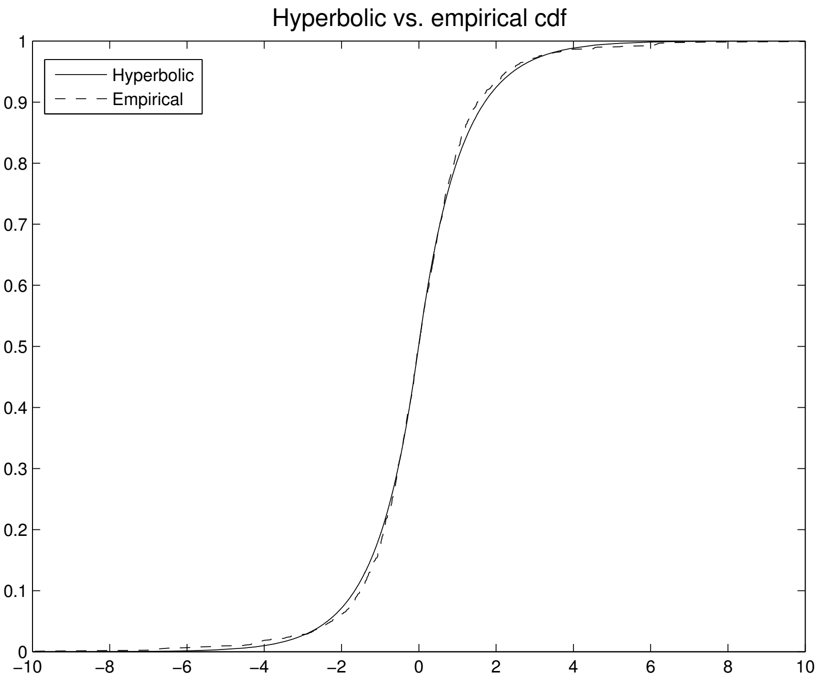


(a) Normal cdf together with empirical cdf

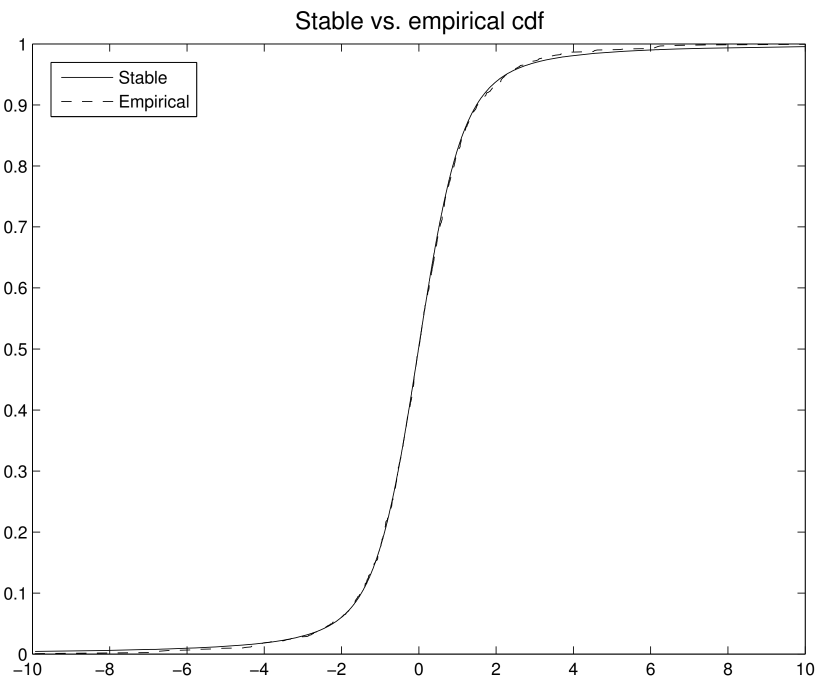


(b) NIG cdf together with empirical cdf

Figure 6: Cumulative distribution function vs. empirical distribution function.



(c) Hyperbolic cdf together with empirical cdf



(d) Stable cdf together with empirical cdf

Figure 6: Cumulative distribution function vs. empirical distribution function (continued).

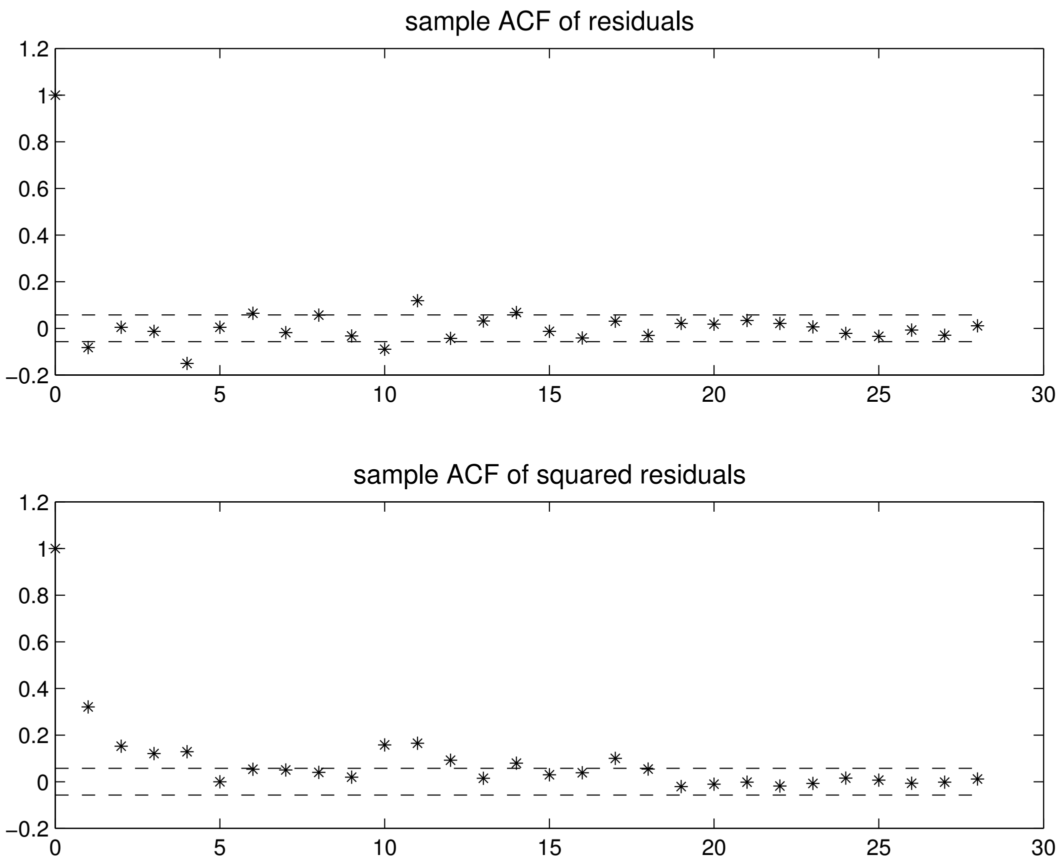


Figure 7: Sample ACF of the residuals(upper) and squared residuals(lower). The horizontal lines are the 95% confidence bands for standard Gaussian white noise.

5.4.2 SARFIMA-GARCH(1,1)

The parameter estimates are reported in Table 6. We see that the GARCH parameters are all significant. Note however that the stationary assumption $\alpha + \beta < 1$ is violated in the normal case.

The sample ACF plots in Figure 9 and the whiteness of the residuals in Figure 8 implies that the model is able to capture the dependencies in the data. The Ljung-Box values in Table 7 confirm this, at least for the residuals and squared residuals. In the same table, the goodness-of-fit statistics suggests that the most accurate conditional distribution specification for this model is given by the Student-t density. However, according to the histograms plots in Figure 10, the GED density performs better in the center of the distribution.

It can be seen in Figure 11 that the conditional standard deviation corresponds very well with the price changes in the preprocessed time series.

Table 6: The SARFIMA-GARCH(1,1) parameter estimates.

	Normal	Student-t	GED
ϕ_1	0.9099	0.9146	0.9252
	(0.0168)	(0.0156)	(0.0167)
ϕ_5	0.0792	0.0765	0.0690
	(0.0168)	(0.0154)	(0.0253)
δ_{Mon}	1.7694	1.8380	1.8944
	(0.1071)	(0.0943)	(0.0964)
δ_{Sat}	-0.9991	-1.0136	-1.0122
	(0.0938)	(0.0868)	(0.0126)
δ_{Sun}	-0.7728	-0.6718	-0.6042
	(0.1020)	(0.0877)	(0.1667)
ω	0.0933	0.1714	0.1370
	(0.0221)	(0.0615)	(0.0417)
α	0.3040	0.3325	0.2971
	(0.0362)	(0.0799)	(0.0569)
β	0.7065	0.6603	0.6798
	(0.0267)	(0.0654)	(0.0489)
D	0.1336	0.1024	0.0811
	(0.0260)	(0.0220)	(0.0127)
ν	—	3.7435	1.0459
	(—)	(0.4656)	(0.0622)

Standard errors are given in parenthesis below each value.

Table 7: SARFIMA-GARCH(1,1): Ljung-Box values for the residuals, squared residuals and absolute residuals $(Q(\cdot)^* - Q(\cdot)^{***})$, with critical value 41.3377, and goodness-of-fit statistics.

	$Q^*(28)$	$Q^{**}(28)$	$Q^{***}(28)$	KD	AD
Normal	28.2833	21.6472	47.7182	2.2619	8.8909
Student-t	29.1721	22.6492	48.0293	1.0691	1.5328
GED	32.1133	21.9076	47.1917	1.2661	2.4842

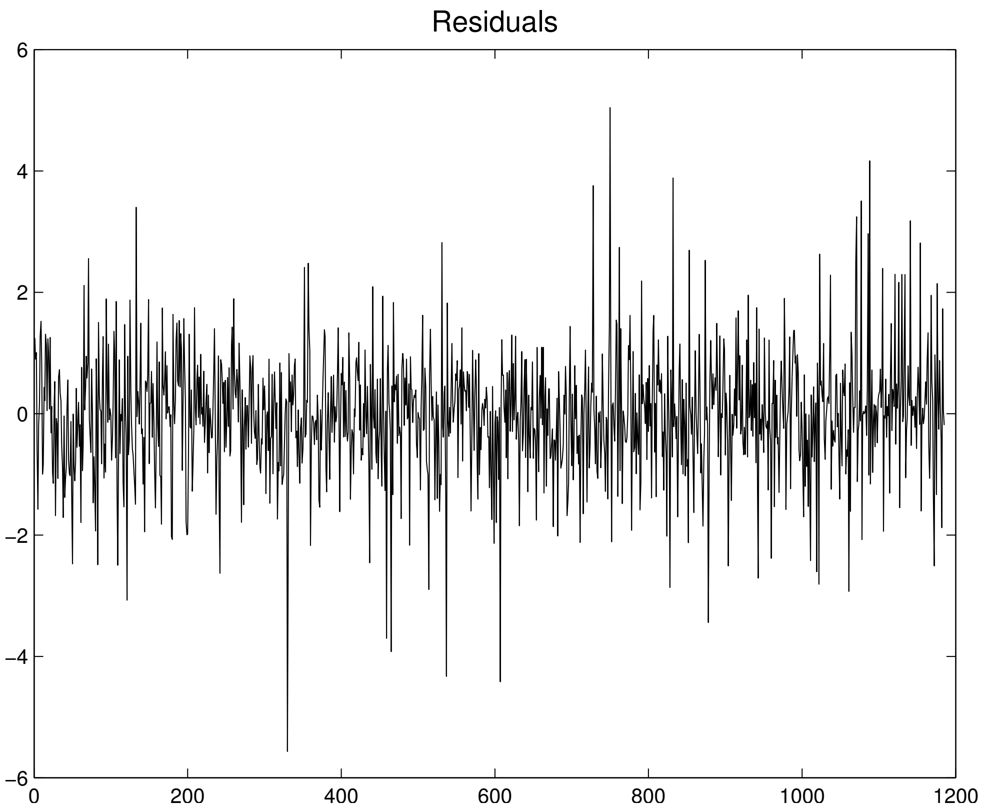


Figure 8: Residuals of the SARFIMA-GARCH(1,1)-n model.

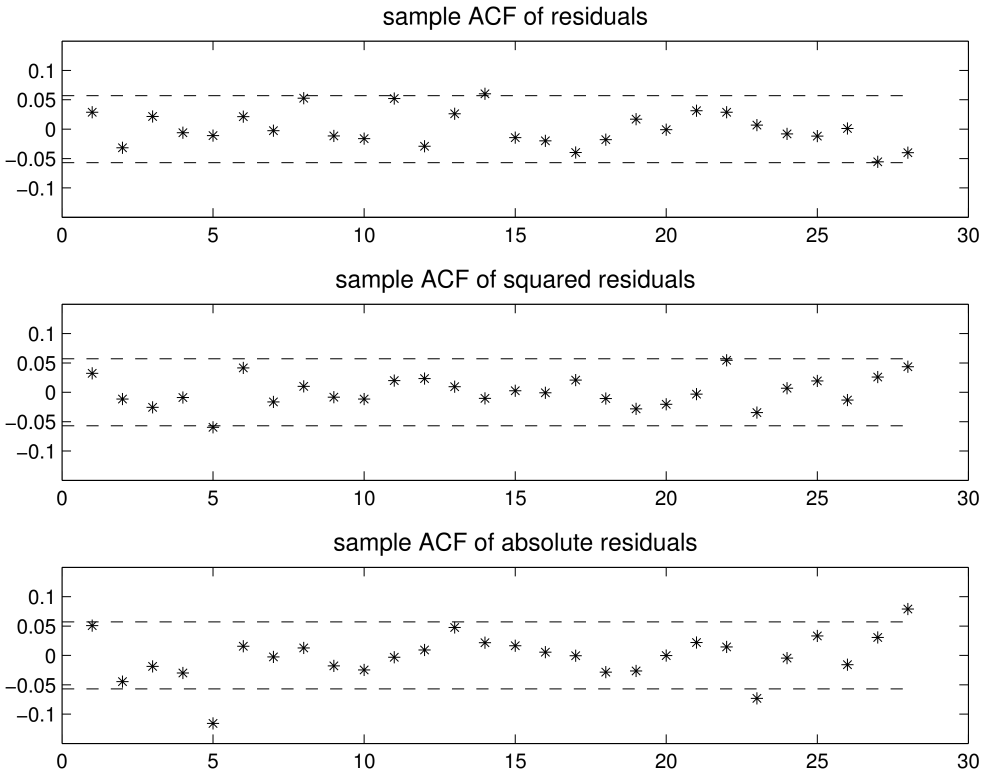
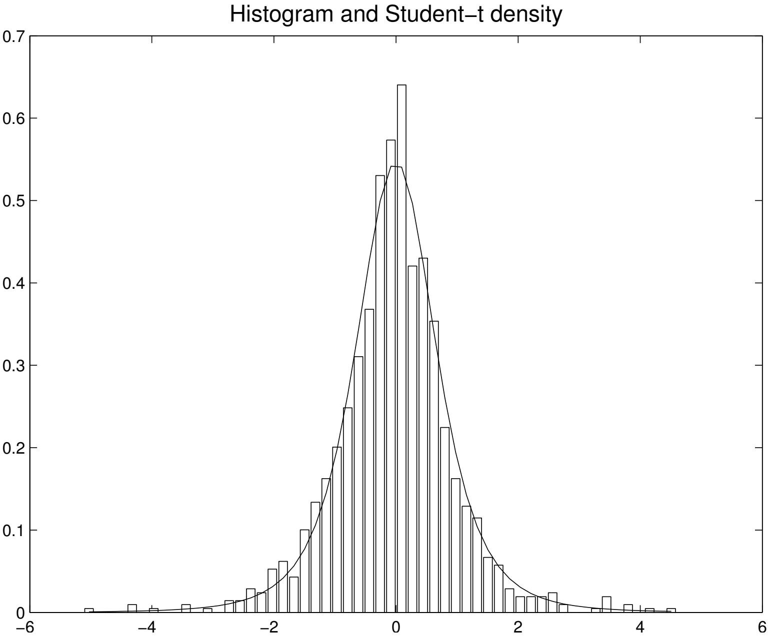
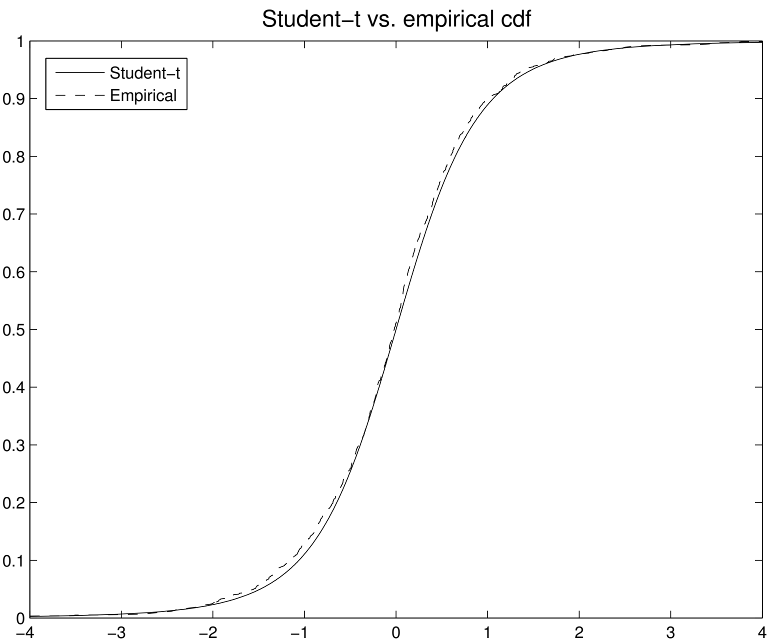


Figure 9: Sample ACF for residuals, squared residuals and absolute residuals for the SARFIMA-GARCH(1,1)-n model.

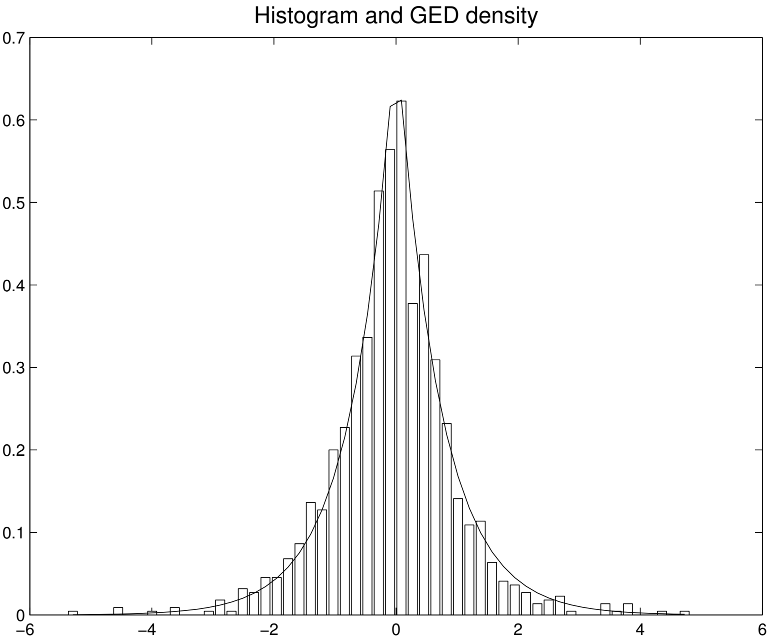


(a) Histogram and Student-t density

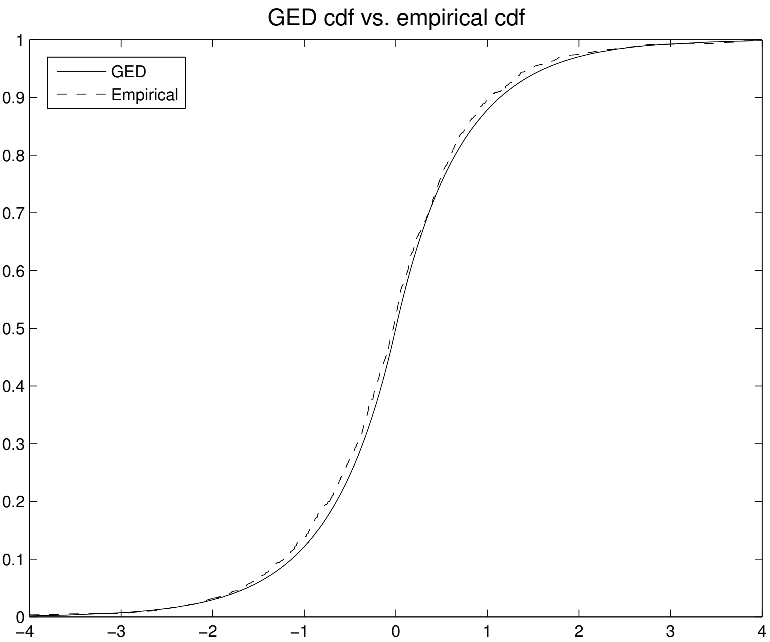


(b) Empirical cdf and Student-t cdf

Figure 10: Histograms, densities and cdf plots for the SARFIMA-GARCH(1,1) model.

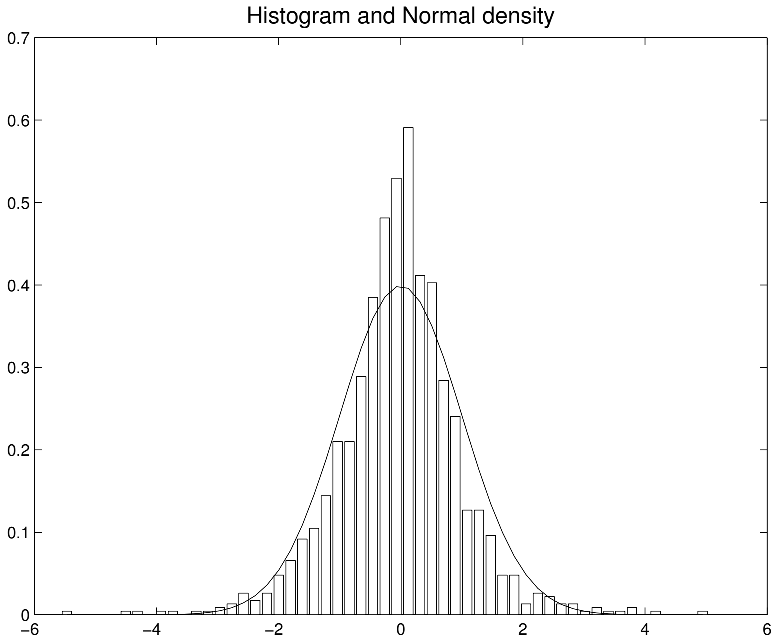


(c) Histogram and GED density

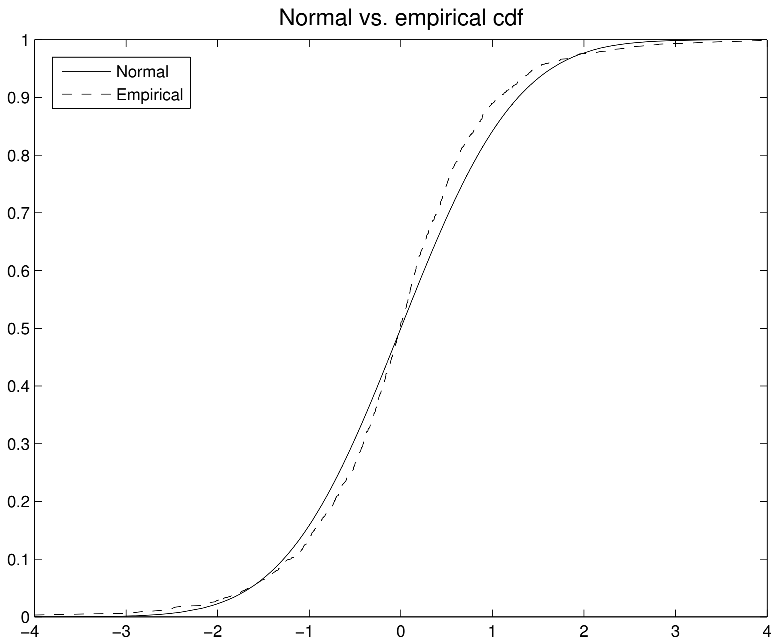


(d) Empirical cdf and GED cdf

Figure 10: Histograms, densities and cdf plots for the SARFIMA-GARCH(1,1) model (continued).



(e) Histogram and normal density



(f) Empirical cdf and normal cdf

Figure 10: Histograms, densities and cdf plots for the SARFIMA-GARCH(1,1) model (continued).

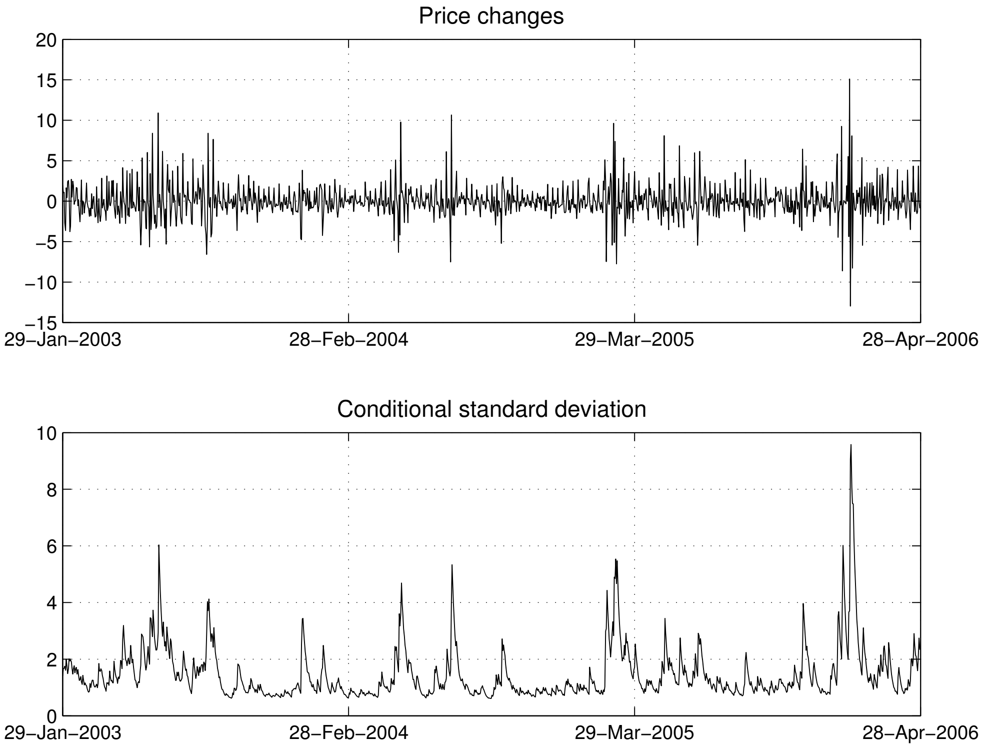


Figure 11: Price changes (upper) and conditional standard deviation in the SARFIMA-GARCH(1,1)-n model (lower).

5.4.3 SARFIMA-EGARCH

The parameter estimates are reported in Table 8. Interestingly, the EGARCH-parameter γ is not significant, so no leverage effect can be detected in our time series.

From the residuals in Figure 12, ACF plots in 13 and the Ljung-Box values in Table 9, we conclude that the model takes care of most dependencies in the data. However, it does seem to perform a bit worse than the SARFIMA-GARCH(1,1) model.

The Student-t density seems to be the most correct conditional density if one is led by the statistics in Table 9, but like in the case with SARFIMA-GARCH(1,1), the GED density seems to perform better in the center of the distribution (see Figure 14).

Figure 15 shows that the conditional standard deviation corresponds well with the price changes.

Table 8: The SARFIMA-EGARCH parameter estimates.

	Normal	Student-t	GED
ϕ_1	0.9215	0.9163	0.9261
	(0.0163)	(0.0150)	(0.0248)
ϕ_5	0.0682	0.0742	0.0679
	(0.0165)	(0.0149)	(0.0293)
δ_{Mon}	1.7174	1.7876	1.8319
	(0.0962)	(0.0910)	(0.0975)
δ_{Sat}	-0.9966	-0.9874	-0.9989
	(0.0887)	(0.0845)	(0.1166)
δ_{Sun}	-0.7593	-0.6693	-0.6192
	(0.0956)	(0.0833)	(0.0084)
ω	-0.3438	-0.3398	-0.3324
	(0.0293)	(0.0473)	(0.0442)
α	0.5234	0.6067	0.5365
	(0.0443)	(0.0917)	(0.0732)
β	0.9342	0.8686	0.8975
	(0.0147)	(0.0397)	(0.0310)
D	0.1378	0.0949	0.0788
	(0.0239)	(0.0206)	(0.0239)
ν	—	3.7069	1.0468
	(—)	(0.4558)	(0.0611)

Standard errors are given in parenthesis below each value.

Table 9: SARFIMA-EGARCH: Ljung-Box values for the residuals, squared residuals and absolute residuals($Q(\cdot)^* - Q(\cdot)^{***}$), with critical value 41.3377, and goodness-of-fit statistics.

	$Q(28)^*$	$Q(28)^{**}$	$Q(28)^{***}$	KD	AD
Normal	28.6899	24.2915	54.4548	8.7273	2.3766
Student-t	32.0373	30.9112	51.9250	1.0269	1.2767
GED	33.8280	27.3486	49.4528	1.1626	1.9642

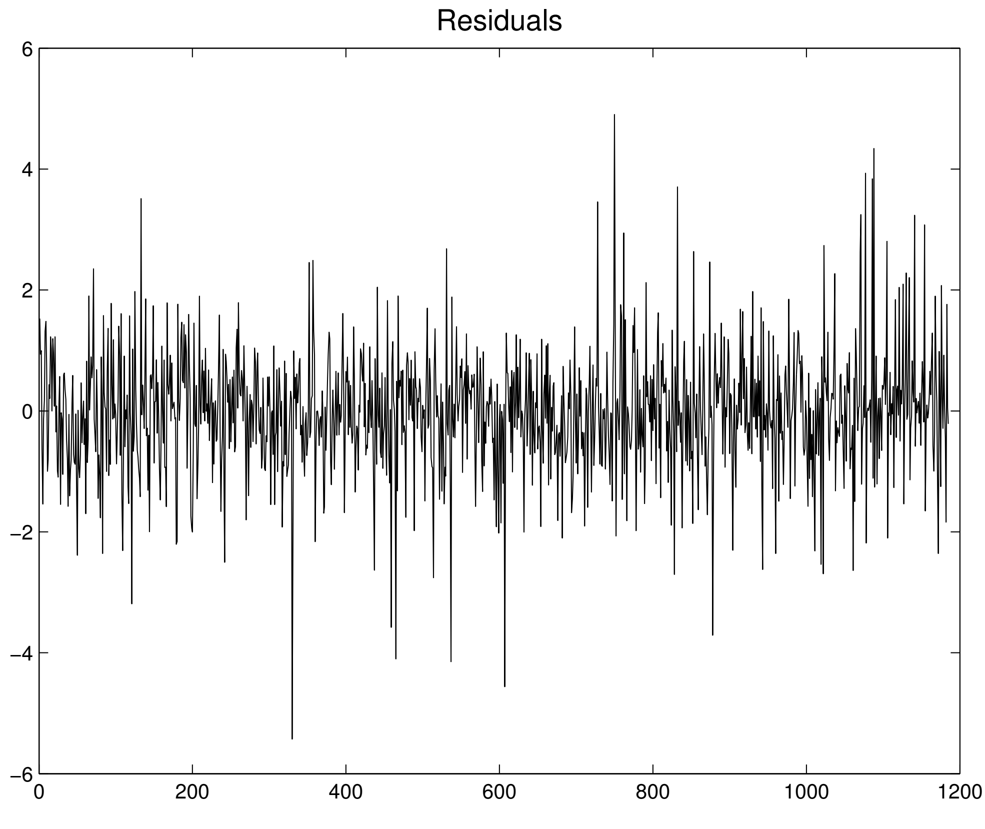


Figure 12: The residuals of the SARFIMA-EGARCH-n model.

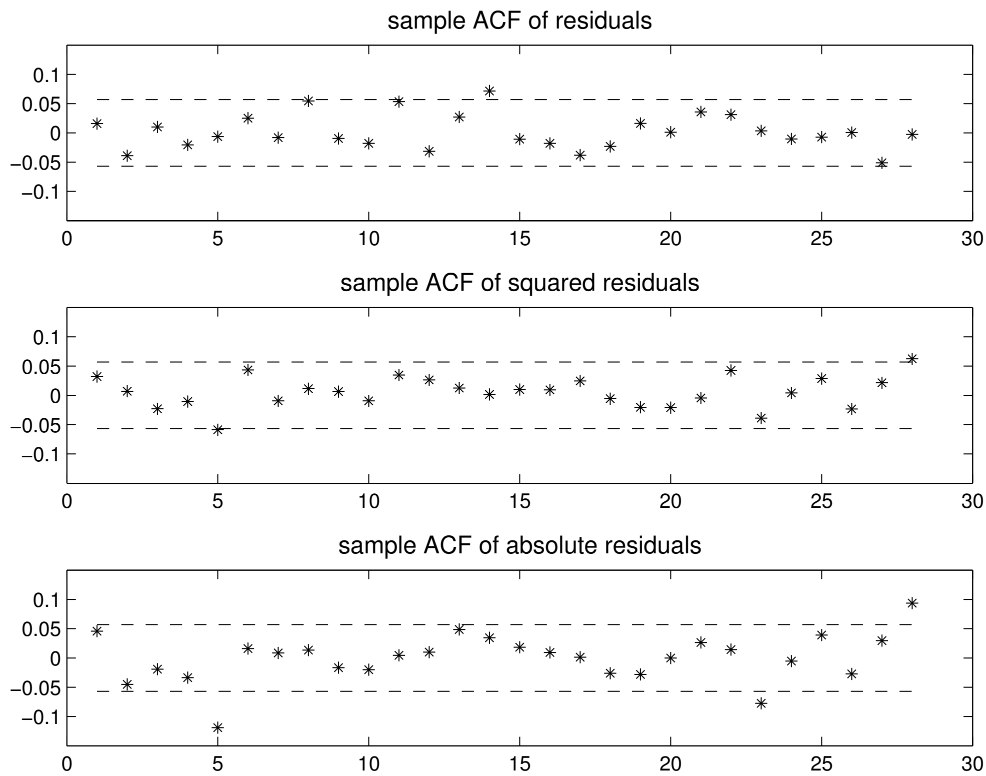
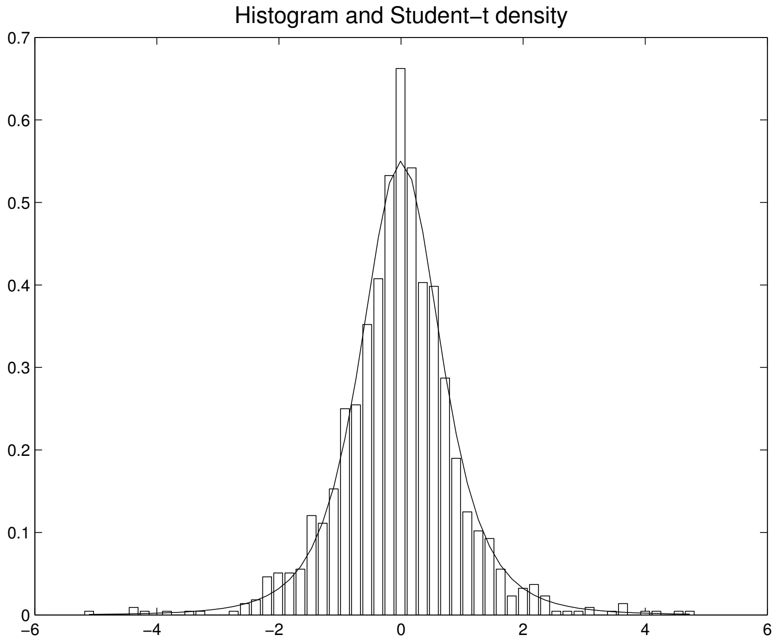
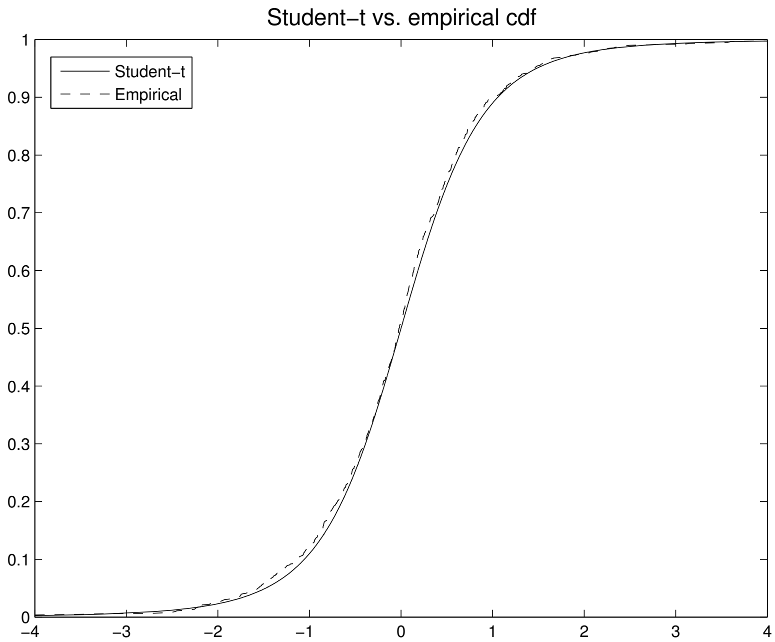


Figure 13: Sample ACF plots for the SARFIMA-EGARCH-n model.

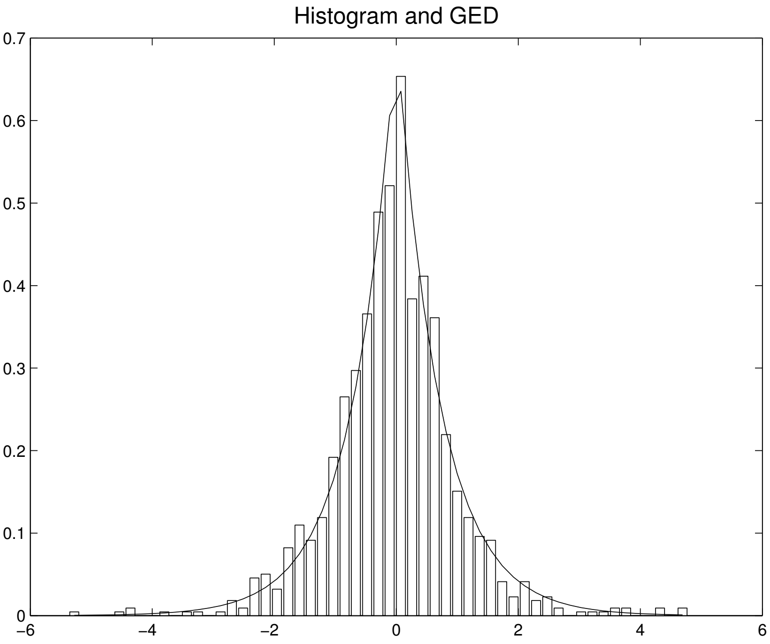


(a) Histogram and Student-t density

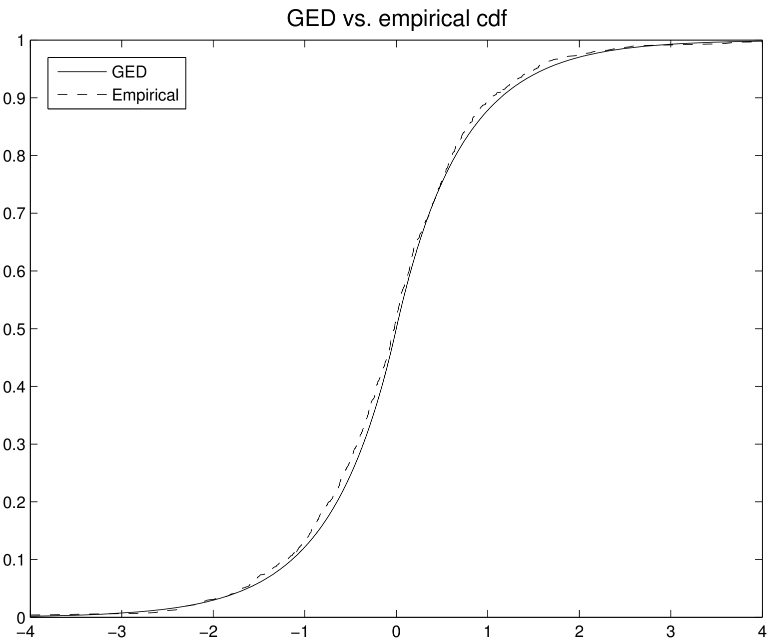


(b) Empirical cdf and Student-t cdf

Figure 14: Histograms, densities and cdf plots for the SARFIMA-EGARCH model.

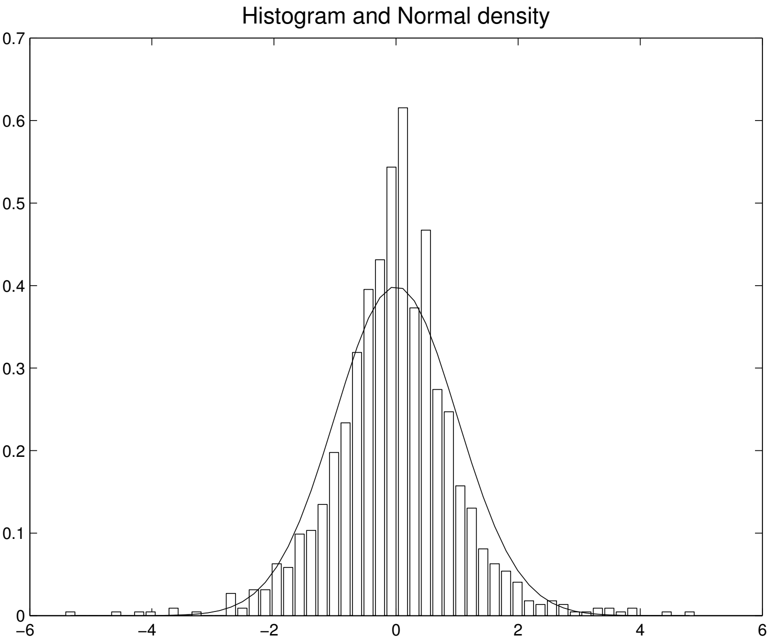


(c) Histogram and GED density

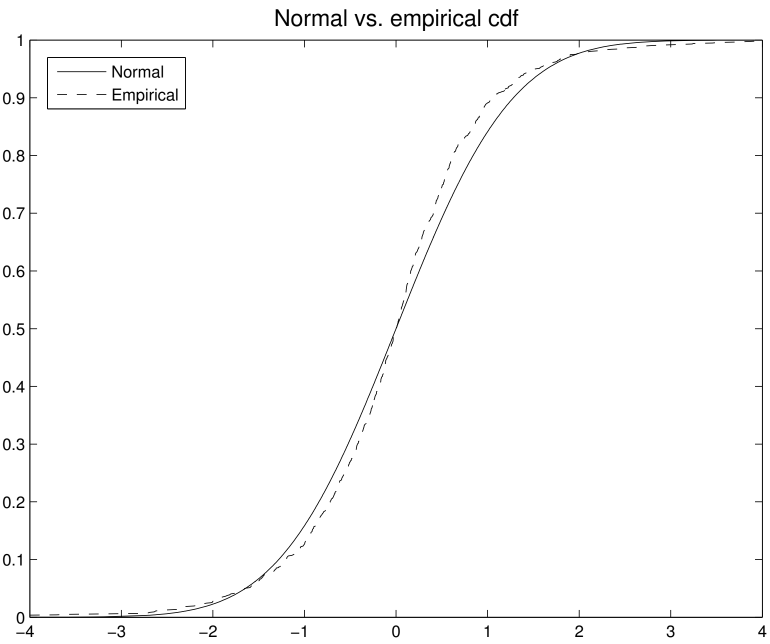


(d) Empirical cdf and GED cdf

Figure 14: Histograms, densities and cdf plots for the SARFIMA-EGARCH model (continued).



(e) Histogram and normal density



(f) Empirical cdf and normal cdf

Figure 14: Histograms, densities and cdf plots for the SARFIMA-EGARCH model (continued).

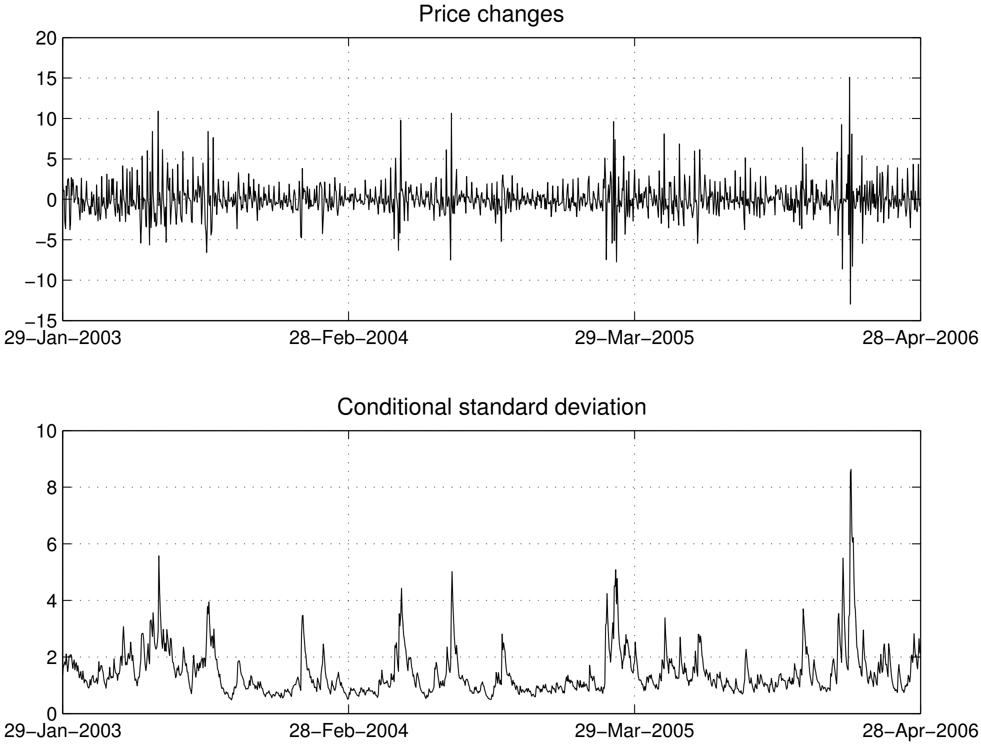


Figure 15: Price changes (upper) and conditional standard deviation for the SARFIMA-EGARCH-n model (lower).

5.4.4 The importance of D

In order to determine the importance of the seasonal difference parameter D, we re-estimate all models with the constraint $D = 0$. The Ljung-Box values for the residuals of the SARFIMA, SARFIMA-GARCH and SARFIMA-EGARCH model in this case are 80.66, 62.36 and 69.47, respectively. Hence, all models perform considerably worse when $D = 0$. Recall that the 95% critical value is 41.34, so we are left with severe dependence in the residuals. The sample ACF plots in Figure 16 below indicates that most of this comes from the seasonal long memory, with significant autocorrelations at lags = 7, 14 and 21, at least for the SARFIMA-GARCH and SARFIMA-EGARCH models.

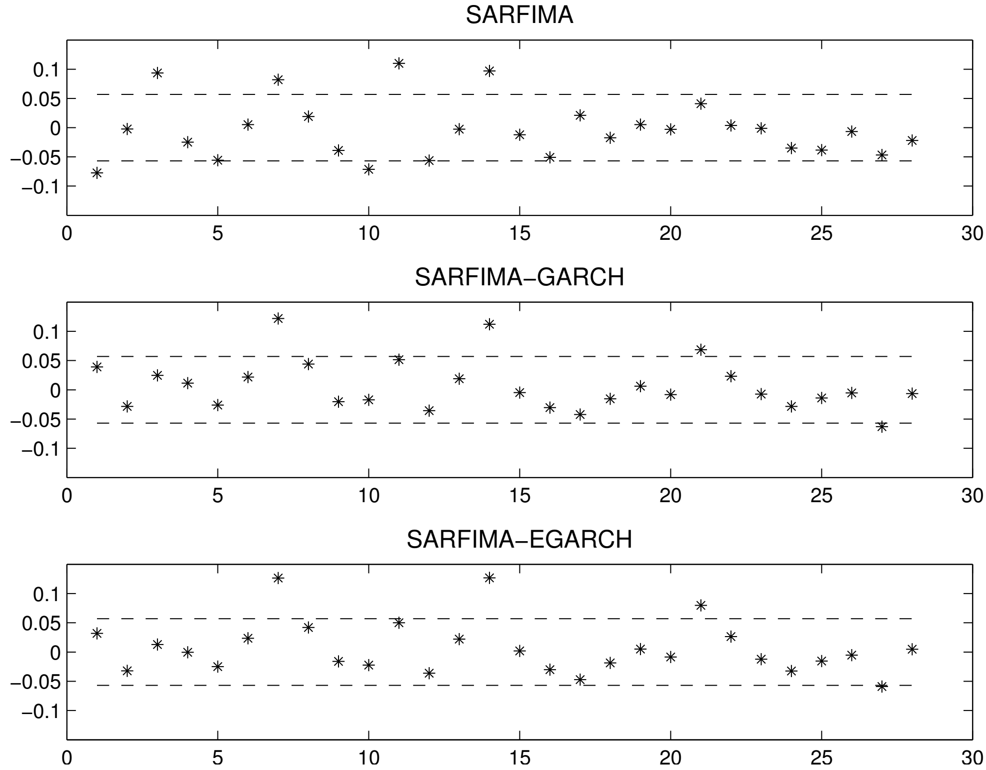


Figure 16: Sample ACF when $D = 0$.

5.5 Conclusions

The SARFIMA-GARCH(1,1) and SARFIMA-EGARCH processes, both with an autoregressive conditional mean equation that takes care of the weekly pattern, effectively captures the dependencies in the system price from the Nordic power exchange. In particular, they capture the time varying standard deviation and seasonal long memory in the price series. Leaving out the GARCH-type process results in heavy-tailed residuals.

References

[1] Angeralides, A. and Cronie, O. (2006). Modelling electricity prices with Ornstein-Uhlenbeck processes. Master's Thesis, Department of Mathematical Sciences, Chalmers University of Technology and Göteborg University.

`http://www.math.chalmers.se/~palbin/CorrectedThesis30Aug.pdf`

[2] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31** 307–327.

[3] Brietzke, E.H.M., Lopes. S.R.C. and Bisognin, C. (2005). A closed formula for the Durbin-Levinson's algorithm in seasonal fractionally integrated processes. *Mathematical and Computer Modeling* **42** 1191–1206.

[4] Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods.* Springer-Verlag.

[5] McCulloch, J.H. (1986). Simple consistent estimators of stable distribution parameters. *Communications in Statistics - Simulation* **15** 1109–1136.

[6] Mugele, J.H., Rachev, S.T. and Trück, S. (2005). Stable modeling of different European power markets. *Investment Management and Financial Innovations* **2** 65–85.

[7] Nelson, D. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59** 347–370.

[8] `www.nordpool.com`

[9] Samorodnitsky, G. and Taqqu, M.S. (1994). *Stable Non-Gaussian Random Processes.* Chapman & Hall.