CHALMERS | GÖTEBORG UNIVERSITY

MASTER'S THESIS

ARMA and GARCH-type Modeling Electricity Prices

YAN DONG

Department of Mathematical Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
GÖTEBORG UNIVERSITY
Göteborg Sweden 2012

Thesis for the Degree of Master of Science

ARMA and GARCH-type Modeling Electricity Prices

YAN DONG

CHALMERS | GÖTEBORG UNIVERSITY

Department of Mathematical Sciences
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg, Sweden
Göteborg, May 2012

Matematiskt centrum
Göteborg 2012

Abstract

In this master thesis statistical models for electricity prices during January and February 2012 is developed. The time series is modeled based on autoregressive and moving average (ARMA) model and extreme value theory. The spikes are simulated by Generalized Pareto distribution (GPD). The innovation process is analyzed by autoregressive conditionally heteroskedastic (GARCH) process and exponential GARCH (EGARCH) process. All the parameters are estimated by maximum likelihood method.

Acknowledgments

I would like to thank my supervisor Partik Albin for his encouragement and advice. I also learn a lot from him during the whole master period. He not only helps me in academic level, but also encourage me during my studying overseas. I also would like to thank my examinor Anastassia Baxevani. I have obtained both statistical knowledge and scientific methods from her course: Computer Intensive Statistical Methods. Lastly, I would like to thank Robin Axelsson, an outstanding Swedish master student in Chalmers.

Contents

1 Introduction

Many statistical models for time series are worked through researching mean or variance (volatility) of the process. The ARMA model is used to present the stationary time series based on autoregressive process and moving average of noises. On the other hand, the autoregressive conditional heteroskedasticity (ARCH) model is focus on time varying conditional variance. In practice, however, high ARCH order has to be selected. Bollerslev extended this model to Generalized ARCH (GARCH) model, which can solve this problem. GARCH model describe variance at a certain time with both past values and past variances. Most time series is sufficiently modeled using GARCH(1, 1) that only includes three parameters.

Electricity prices are significantly affected by the demand and supply on the market. Generally, high demand results in high price. Moreover, electricity prices are affected by external factors such as weather, prices of fossil fuels, availability of nuclear power, water reservoir levels, prices of exhaust rights etc. Like most financial time series, electricity price has the characteristic of volatility clustering. Mandelbrot quoted (1963): large changes tend to be followed by large changes, and small changes by small changes. GARCH-type model successfully captures this property. Another main feature of electricity price is fat tails. Extreme values in electricity prices are better analyzed by extreme value theory.

In this project, we aim to explore the properties of electricity price applying ARMA , GARCH-type models and extreme value theory. We will mainly focus on modeling hourly electricity prices from Nord Pool in this winter, Jan-Feb 2012, hoping those external factors might be ignored. We pay attention to periodic components and extreme values. At first, in Section 4.2, the predictable periodicities of 24-hours and 168-hours and extreme values are removed. Then, in Section 4.3, we will fit the time series into ARMA model. In Sections 4.4 and 4.5, to analyze the volatility of data set, we employ GARCH-type models. In this thesis, extreme value theory is also involved in Section 4.6. At last, in Section 4.7, we simulate time series using above models and GP distribution.

2 Electricity Markets

2.1 The Nord Pool

Nord Pool market was created in 1996 as a result of the establishment of common electricity market of Norway and Sweden. Nord Pool Sport runs the largest market for electrical energy in the world, measured in volume traded (TWh) and in market share. It operates in Norway, Denmark, Sweden, Finland and Estonia. More than 70 % of the total consumption of electrical energy in the Nordic market is traded through Nord Pool Spot. It was the world's first multinational exchange for trading electric power. Nord Pool Spot offers both day-ahead and intra-day markets, see [18].

2.2 The Main Features of Electricity Prices

The first characteristic is periodicity of different length. Electricity prices exhibit various seasonality over days, weeks and months. Weather conditions affect demand of electricity over months. In this paper we analyze electricity price on winter months. So, the periodic behaviors in daily and weekly are considered. And they explain periodicity components strongly, since the need for electricity is various during whole day and whole week. For example, the electricity demand is higher during daytime than at night. On the other hand, the electricity supply performs different ways between weekdays and weekends.

Secondly, presence of spikes is distinct in electricity price. This feature is related to instantaneous supply and demand, which is quite differ from stocks. And it should be treated by appropriate model.

The third one is stationary over short intervals. This means that electricity price is mean reversion over shorter periods. It can be observed that, during winter, the price fluctuate around a stable level but not follow a trend.

Last but not least, high volatility plays important role among features of electricity price. The volatility means the standard deviation of the hourly price. For electricity price, the volatility is not a constant but various from time to time.

3 Theoretical Background

3.1 ARMA Model

The general autoregressive and moving average (ARMA) statistical model is used to describe a time series that evolves over time. In this process there is a linear relationship between the values at a certain time point and past values, noise as well.

According to [1], time series $\{X_t\}$ is an ARMA(p, q) process if $\{X_t\}$ is stationary and if for every t,

$$X_t = \phi_1 X_{t-1} - ... + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q},$$

where $\{\epsilon_t\}$ is i.i.d. $N(0, \sigma^2)$ and the polynomials

$$(1 - \phi_1 \epsilon - ... - \phi_p \epsilon^p)$$

and

$$(1 + \theta_1 \epsilon + ... + \theta_q \epsilon^q)$$

have no common factors.

The process $\{X_t\}$ is said to be an ARMA(p, q) process with mean μ if $\{X_t - \mu\}$ is an ARMA(p, q) process. The time series $\{X_t\}$ is said to be an autoregressive process of order p, and a moving-average process of order q.

3.2 GARCH Model

According to [4], the generalized autoregressive conditional heteroskedasticit (GARCH) is a model that is used to estimate the volatility of an asset. It indicates that the present volatility depends on past observations and volatilities. The time series X_t can be modeled by

$$X_t = \sigma_t \epsilon_t,$$

where $\{\epsilon_t\}$ is i.i.d. $N(0, 1)$ random variables. GARCH model is used to estimate the variance σ^2

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

The GARCH (p, q) model is strictly stationary with finite variance when the conditions $\omega > 0$, and $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$ are required. We can see the GARCH model has similar pattern with ARMA model, which shows we can derive GARCH process using similar theory and method with ARMA.

Particularly, in most cases structure $p = q = 1$ is sufficient and it is sufficient for our purposes. GARCH(1, 1) model is the most widely used, which is given by

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2$$

To obtain strictly stationary solution, the conditions $\omega > 0, \alpha + \beta < 1$ are required. Looking at the formula we see that GARCH (1, 1) explains that the present volatility depends only on previous one. It is easy to calculate and simulate since there are only three parameters in GARCH(1, 1) model.

Though GARCH model successfully explain the volatility clustering, it does not capture the leverage effect. Next we will introduce another GARCH-type model, which could capture leverage effect.

3.3 Exponential GARCH Model

Accoding to [4], The exponential GARCH (EGARCH) is a model that is used to estimate the volatility of an asset. Time series X_t can be modeled by

$$X_t = \sigma_t \epsilon_t$$

where ϵ_t is i.i.d.$N(0, 1)$ random variables.

$$\log \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i |\epsilon_{t-i}| + \sum_{i=1}^{q} \gamma_i \epsilon_{t-i} + \sum_{j=1}^{p} \beta_j \log \sigma_{t-j}^2$$

EGARCH(1, 1) is frequently used to estimate the variance σ^2

$$\log {\sigma_t}^2 = \omega + \alpha |\epsilon_{t-1}| + \gamma \epsilon_{t-1} + \beta \log {\sigma_{t-1}}^2$$

The parameters ω, α, β and γ can be estimated by using the maximum likelihood method. Note that $|\beta| < 1$ is required, and γ is the parameter that indicates leverage effect.

3.4 Stationary Process

A discrete time stochastic process $\{X(n)\}_{n=-\infty}^{\infty}$ is simply an definite sequence of random variables defined on a common probability space. And electricity prices can be modeled as discrete time stochastic processes. The stationary process is required, which means that finite dimensional distributions of the process are invariant for time translations

$$P\{X(n_1+m) \in A_1, ..., X(n_k+m) \in A_k\} = P\{X(n_1) \in A_1, ..., X(n_k) \in A_k\}$$

for any $n_1, ..., n_k, m \in \mathbf{Z}, A_1, ..., A_k \subseteq \mathbf{R}$ and $k \in \mathbf{N}$. From this it follows in particular that the random variables $\{X(n)\}_{n=-\infty}^{\infty}$ are identically distributed. Also, it follows that $\{X(n)\}_{n=-\infty}^{\infty}$ is a so called weakly stationary process, which is to say that

$$m_X = \mathbf{E}\{X(n)\}$$

does not depend on time n, and the covariance function

$$r_X(k) = \mathbf{Cov}\{X(n), X(n+k)\}$$

only depends on the distance in time k between a pair of process values considered. It is generally to assume that the hourly electricity price data are observations of a stationary process within short period after removing 24-hours and 168-hours periodicities.

3.5 Extreme Value Theory

3.5.1 Generalized Pareto Distribution

From [11], Generalized Pareto distribution(GPD) is mostly used, when we focus on the behavior of large observations that exceed a high threshold. Given a high threshold u, the distribution of excess values of x over threshold u is defined by

$$F_u(y) = P\{X - u \leq y | X > u\} = \frac{F(y+u) - F(u)}{1 - F(u)}$$

which represents the probability that the value of x exceeds the threshold u by at most an amount y given that x exceeds the threshold u. A theorem by

Balkema and de Haan and Pickands shows that for sufficiently high threshold u, the distribution function of the excess may be approximated by the generalized Pareto distribution (GPD) such that, as the threshold gets large, the excess distribution $F_u(y)$ converges to the GPD, which is

$$G(x) = \begin{cases} 1 - (1 + \gamma \frac{x}{\beta})^{-1/\gamma} & \text{if } \gamma \neq 0 \text{ ;} \\ 1 - e^{-x/\beta} & \text{if } \gamma = 0. \end{cases}$$

where γ is the shape parameter. For ordinary Pareto distribution γ is positive.

A graphical test for assessing the tail behavior may be performed by studying the sample mean excess function based on the sample $X_1, ..., X_n$. With N_u being the number of exceedances of u by $X_1, ..., X_n$, the sample mean excess function is given by

$$e_n(u) = \frac{1}{N_u} \sum_{k=1}^{n} (X_k - u) 1_{u,\infty}(X_k).$$

The mean excess plot is the plot of the points

$$\{(X_{k,n}, e_n(X_{k,n})) : k = 2, ..., n\}.$$

If the mean excess plot is approximately linear with positive slope the X_n may be assumed to have a heavy-tailed Pareto-like tail.

Figure 1 shows the mean excess plot for GP distribution.

3.5.2 Peak Over Threshold

Assume an i.i.d. sample of random variables $X_1, ..., X_n$ from an unknown distribution function F with a right tail, the distribution of excesses $X_k - u$ over a high threshold u is approximated by a distribution called the generalized Pareto distribution (GPD). This situation can be applied to construct estimates of tail probabilities.

For $\gamma > 0$ and $\beta > 0$, the generalized Pareto distribution (GPD) function is given by

$$G_{\gamma,\beta}(x) = 1 - (1 + \gamma x/\beta)^{-1/\gamma}, x \geq 0$$

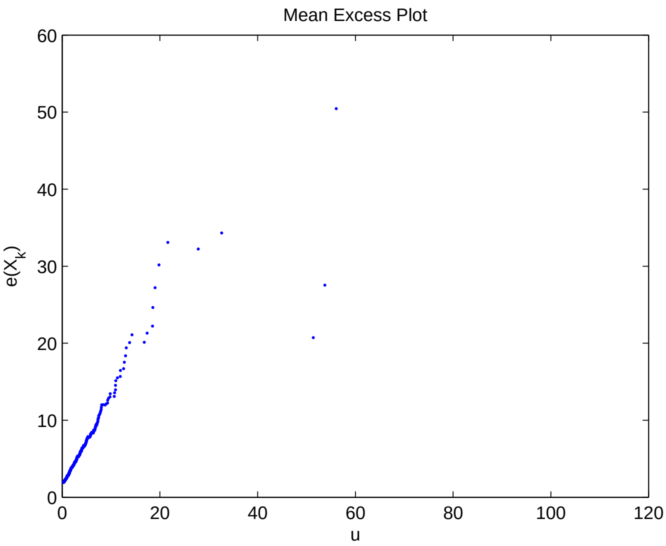

Figure 1: Mean Excess Plot for GP Distribution

Suppose that X is a random variable with distribution function F that has a regularly varying right tail so that

$$\lim_{u\to+\infty}\frac{1-F(\lambda u))}{1-F(u)}=\lambda^{-\alpha}$$

for all $\lambda>0$ and some $\alpha>0$. Then

$$\lim_{u\to\infty}P(\frac{X-u}{u/\alpha}>x|X>u)=\lim_{u\to\infty}\frac{P(X>u(1+x/\alpha))}{P(X>u)}$$

$$=(1+x/\alpha)^{-\alpha}=1-G_{1/\alpha,1}(x)$$

The excess distribution function of X over the threshold u is given by

$$F_u(x)=P(X-u\leq x|X>u),x\geq 0.$$

Notice that

$$1-F_u(x)=\frac{1-F(u+x)}{1-F(u)}=\frac{1-F(u(1+x/u))}{1-F(u)}.$$

Since F is regularly varying with index $-\alpha<0$ it holds that $(1-F(\lambda u))/(1-F(u))\to\lambda^{-\alpha}$ uniformly in $\lambda\geq 1$ as $u\to\infty$, i.e.

$$\limsup_{\lambda\geq 1}|(1-F(\lambda u))/(1-F(u))-\lambda^{-\alpha}|=0.$$

Hence,

$$\lim_{u\to\infty}\sup_{x>0}|1-F_u(x)-(1-G_{\gamma,\beta(u)}(x))|=0,$$

where $\gamma=1/\alpha$ and $\beta(u)\sim u/\alpha$ as $u\to\infty$

Next, choose a high threshold u and let

$$N_u=\#\{i\in\{1,...,n\}:X_i>u\}$$

be the number of exceedances of u by $X_1,...,X_n$. On the other hand,

$$1-F(u+x)=(1-F(u))(1-F_u(x)).$$

If u is not too far from the tail, then the empirical approximation $1-F(u)\approx F_n(u)=N_u/n$ is accurate. Moreover,

$$1-F_u(x)\approx 1-G_{\hat{\gamma},\hat{\beta}}(x)=(1+\hat{\gamma}\frac{x}{\hat{\beta}})^{-1/\hat{\gamma}}$$

where $\hat{\gamma}$ and $\hat{\beta}$ are the estimated parameters. So,

$$1-\hat{F}(u+x)\approx\frac{N_u}{n}(1+\hat{\gamma}\frac{x}{\beta})^{-1/\hat{\gamma}}$$

4 Methods

4.1 Maximum Likelihood Estimation Method

The maximum likelihood estimation is a most widely used method to estimate parameters in statistical model base on a known data set. For time series $y_1,y_2,...,y_n$, assume the density function is known, the parameters can be estimated through maximizing the probability of getting the observed data from the known density function [13]. For an i.i.d. sample, their joint density function is

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

The likelihood function for the given time series is defined by

$$L(\theta|Y)=\prod_{i=1}^{n}f(y_i|\theta),$$

Our purpose is to estimate parameters through maximizing above likelihood function. In practice it is easy to calculate the logarithm of likelihood function, called log-likelihood function:

$$l(\theta|Y) := \log L(\theta|y_1, y_2, ..., y_n) = \sum_{i=1}^{n} \log f(y_i|\theta).$$

To maximize the log-likelihood function we let the partial derivatives equal to zero.

4.2 Estimation for ARMA Model

Assume time series $\{X_t\}$ is mean zero, we aim to fit this process to ARMA(p, q) model, which satisfied following equation:

$$x_t = \phi_1 x_{t-1} + ... + \phi_p x_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}, \tag{1}$$

where ϵ_t is an i.i.d. $N(0, \sigma^2)$. Let $r = \max(p, q+1)$, and rewrite the model as

$$x_t = \phi_1 x_{t-1} + ... + \phi_r x_{t-r} + \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_{r-1} \epsilon_{t-r+1}. \tag{2}$$

Denote $\phi_i = 0$ for $i > p$ and $\theta_j = 0$ for $j > q$.

Kalman filter, a recursive estimator, is applied in the estimation procedure. Only the estimated state from the previous time step and the current measurement are needed to compute the estimate for the current state. The Kalman filter can be written as a single equation for prediction:

$$\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1} + B_k u_k,$$

where $\hat{x}_{k|k}$ is a posteriori state estimate at time k given observations up to and including at time k, F_k and B_k are parameter matrices and u_k is residual in our model. Based on the Kalman filter, we transform the model to following state-space:

$$Y_{t+1} = AY_t + B\epsilon_{t+1}, \tag{3}$$

$$x_t = C'Y_t, \tag{4}$$

where Y_t is an $r \times 1$ state vector, A is an $r \times r$ matrix, and B and C are $r \times 1$

vectors, which is given by:

$$A = \begin{pmatrix} \phi_1 & 1 & 0 & 0 & ... & 0 \\ \phi_2 & 0 & 1 & 0 & ... & 0 \\ \phi_3 & 0 & 1 & 0 & ... & 0 \\ \vdots & & & & \ddots & \\ \phi_{r-1} & 0 & 1 & 0 & ... & 0 \\ \phi_r & 0 & 0 & 0 & ... & 0 \end{pmatrix}; B = \begin{pmatrix} 1 \\ \theta_1 \\ \theta_2 \\ \\ \vdots \\ \theta r - 1 \end{pmatrix}; C = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \\ \vdots \\ 0 \end{pmatrix}.$$

We can see that system (3) and (4) is equivalent to (2). Now denote by $\hat{Y}_{t+1|t} = E[Y_{t+1}|x_0, ..., x_t; Y_0]$ the expected value of Y_{t+1} conditional on the history of observations $(x_0, ..., x_t)$. Associated with each of these forecasts is a mean squared error matrix, defined as

$$Z_{t+1|t} = E[(Y_{t+1} - \hat{Y}_{t+1|t})(Y_{t+1} - \hat{Y}_{t+1|t})'].$$

Given the estimate $\hat{Y}_{t|t-1}$, we use (4) to calculate the residuals

$$\begin{aligned} e_t &= x_t - E[Y_t|x_0, ..., x_{t-1}; Y_0] \\ &= x_t - C'\hat{Y}_{t|t-1}. \end{aligned}$$

The innovation variance, denoted by ω_t, satisfies

$$\omega_t = E[(x_t - C'\hat{Y}_{t|t-1})(x_t - C'\hat{Y}_{t|t-1})'] \tag{5}$$

$$= E[(C'Y_t - C'\hat{Y}_{t|t-1})(C'Y_t - C'\hat{Y}_{t|t-1})'] = C'Z_{t|t-1}C.$$

Moreover, to estimate $\hat{Y}_{t+1|t}$, the Kalman filter equations imply the following evolution of the matrices $Z_{t+1|t}$

$$Z_{t+1|t} = A[Z_{t|t-1} - Z'_{t|t-1}CC'Z_{t|t-1}/\omega_t]A' + BB'\sigma^2. \tag{6}$$

Given the initial value $\hat{Y}_{1|0} = \overline{0}$, which is the unconditional mean of Y_t, the likelihood function of the observation vector $x_0, x_1, ..., x_T$ is given by

$$L = \prod_{t=1}^{T} (2\pi\omega_t)^{-1/2} \exp(-\frac{e_t^2}{2\omega_t}).$$

Taking logarithms and dropping the constant, we obtain

$$l = -\sum_{t=1}^{T} [\log(\omega_t) + e_t^2/\omega_t]. \tag{7}$$

To find estimations, we should maximize log-likelihood function (7) with the parameters θ_i, ϕ_j, and σ^2. However, above function does not include the term σ^2, and only involve the parameters θ_i, ϕ_j. Suppose we initialize the filter with the matrix $\tilde{Z}_{1|0} = \sigma^2 Z_{1|0}$. Then from (6) it follows that each $Z_{t+1|t}$ is proportional to σ^2, and from (5) it follows that the residual variance is also proportional to σ^2. So, we can optimize with respect to σ^2 first, replace the result into the log-likelihood function, then maximize the function with the parameters θ_i, ϕ_j. Note that (7) becomes

$$l = -\sum_{t=1}^{T}[\log(\sigma^2\omega_t) + \frac{e_t^2}{\omega_t\sigma^2}] \tag{8}$$

Optimize (8) with respect to σ^2, we obtain

$$\sigma^2 = \frac{1}{T}\sum_{t=1}^{T} e_t^2/\omega_t.$$

Replacing above result into (8) and ignoring constants, we obtain following log-likelihood function

$$l = -[T\log\sum_{t=1}^{T} e_t^2/\omega_t + \sum_{t=1}^{T}\log\omega_t] \tag{9}$$

At last, we can estimate ARMA parameters through maximizing equation (9), see [4].

4.3 Estimation for GARCH Model

For GARCH model, the quasi-maximum likelihood method is particularly used to estimate parameters. Since we present an iterative procedure when we calculate the log-likelihood based on an initial values, see [4].

We fit the time series X_t in GARCH (p, q) process, assume the residuals $\{\epsilon_t\}$ are i.i.d $N(0, 1)$, the Gaussian quasi-likelihood function is given by

$$L_n(\theta) = L_n(\theta; X_1, ..., X_n) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp(-\frac{X_t^2}{2\sigma_t^2}),$$

the Gaussian log-likelihood function is given by:

$$l_n(\theta) = -\frac{1}{2}\sum_{t=1}^{n}(\log(2\pi) + \log(\sigma_t^2) + \frac{X_t^2}{\sigma_t^2}).$$

So, maximizing the likelihood is equivalent to minimize $\sum_{t=1}^{n}(\frac{X_t^2}{\sigma_t^2} + \log\sigma_t^2)$. Notice $\sigma_t = \omega + \alpha X_{t-1}^2 + \beta\sigma_{t-1}^2$, in GARCH model.

Iterative procedure is generally used in GARCH(1, 1) model. And it is common to choose $\sigma_0 = Var(X)$ and $X_0 = \sqrt{Var(X)}$ as initial values for GARCH(1, 1). In this project we choose initial values by this way.

4.4 AIC

Order selection is quite important when we use ARMA process. As we know, the higher order in model may result in smaller estimated errors. But the higher-order model is complex. When we apply the model into forecasting, the mean squared error of the forecasts will be large, which depends on errors from estimation of the parameters of the fitted model. So, we should choose order considering both sufficient and simple factors.

Many criteria are used to select order. In this paper we introduce one of them, AIC criterion. AIC, a generally applicable criterion for model selection, is an approximately unbiased estimate of the Kullback-Leibler index of the fitted model versus to the true model. According to [1], AIC criterion is defined as

$$AIC_k = 2k - 2\log(L)$$

where k is the number of parameters in the fitted model and L is the value of the maximized likelihood-function for the model. Assume the model errors have the same variance, then it holds that

$$AIC_k = 2k + N[\log(2\pi\sum_{i=1}^{N}{\epsilon_i}^2/(N-1)) + 1],$$

which means that it is enough to minimize

$$AIC_k = 2k + N\log(\sum_{i=1}^{N}{\epsilon_i}^2/(N-1)).$$

4.5 Analysis of Residuals

Stationarity are required for residuals in our models. To test for stationarity, Ljung-Box test is widely used, which has the null hypothesis that the time series are independently distributed. This test considers the sample autocorrelation functions simultaneously by test statistics Q_{LB} , defined as

$$Q_{LB} = n(n+2)\sum_{i=1}^{h} \hat{\rho}^2(i)/(n-i)$$

where n is the sample size, $\hat{\rho(i)}$ is the sample autocorrelation at lag i, and h is the number of lags being tested. For large n , Q_{LB} can be approximately the chi-squared distribution with degrees of freedom h. The assumption of an i.i.d. sequence is rejected at significance level α if $Q_{LB} > \chi^2_{1-\alpha}(h)$, see [1].

The residuals ϵ_t are also assumed to be normal distribution. To investigate this assumption, several methods and tests can be performed on the residuals. Using visual methods, we can check for normal distribution with a QQ-plot and independence by plotting the autocorrelation function. We can also compare the empirical density plot with normal density plot in the figure. Statistical hypothesis tests, such Pearson Chi-Square test, can be used to test the residuals for normal distribution. Kolmogorov-Smirnov test can be used to check the distance between residuals empirical density function and normal density function. The distance between the fitted distribution function F and the empirical distribution function F_n of the sample $(X_1, ..., X_n)$, given by

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}},$$

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

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

We will check the Kolmogorov distances between residuals and normal distribution for our models.

5 Modeling

In this section the time series model will be applied for the hourly system price from Nord Pool.

5.1 Data Set

The price period from January to February 2012 is used to estimate the model. This data set is obtained from Nord Pool's market data service. The original price is shown in Figure 2. The presence of spikes and volatility clustering is quite obvious. We explore the basic statistics from our data set in Table 1. We also get histogram plot and QQ-plot to compare with normal distribution.

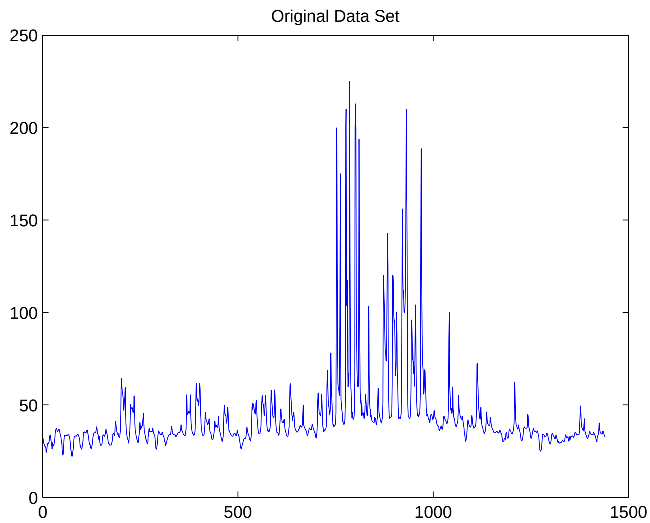

Figure 2: the Original Data Set

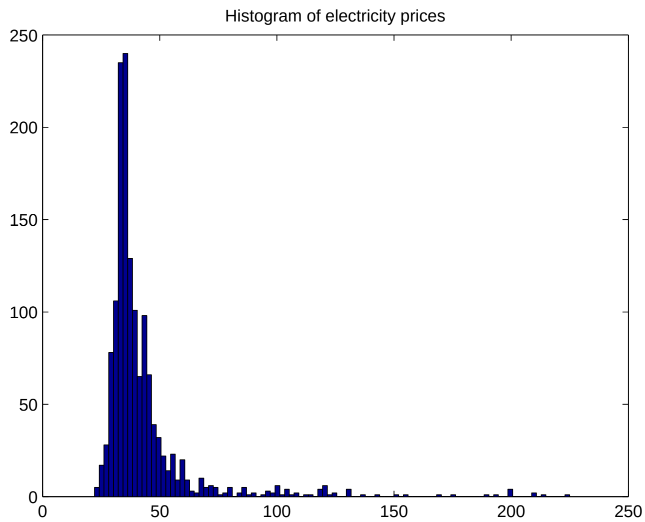

Figure 3: Empirical Histogram of Electricity Prices

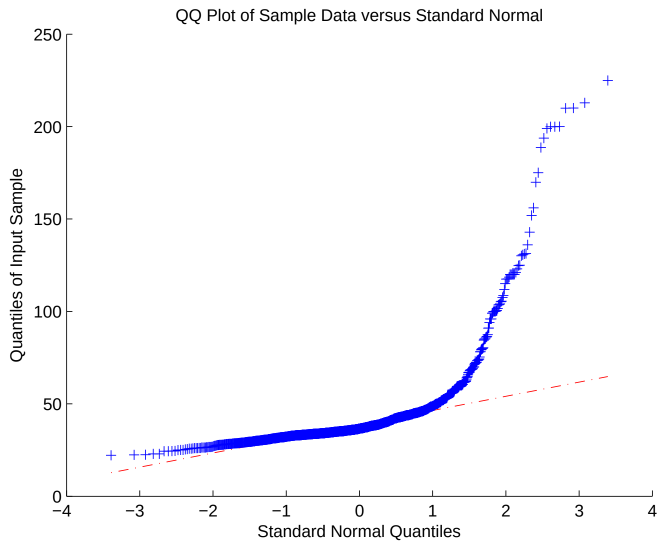

Figure 4: QQ-plot of Original Data Set

Table 1: Descriptive Statistics of Electricity Price

Statistic	Value
Mean	42.9219
Maximum	224.97
Minimum	22.16
Standard Deviation	21.7299
Skewness	4.4526
Kurtosis	28.0635

As we can see, from histogram and QQ-plot, our data set is long tailed comparing with normal distribution. Table 1 shows some descriptive statistics of the electricity price. We can find our data set is heavily tailed according to positive skewness and higher kurtosis than the normal distribution. The statistic values are largely affected by extreme values. So, it is necessary to deal with these extreme observations when we fit the model.

5.2 Fit the Model

From above plots and statistics, it is necessary to identify and remove extreme values. We apply extreme value theory to deal with them. Secondly, considering seasonal components, we remove seasonal periods of 24-hours and 168-hours.

5.2.1 Remove Extreme Values

We should remove some outliers and analyze the extreme values using statistical model. We just remove 1 % left outliers, and deal with right heavy tail based on extreme value theory. There are 84 extreme observations above threshold being removed. For both left outliers and right extreme values, we use their previous values instead of the old observations to protect the periodic feature. However, how to choose the appropriate threshold is not an easy issue, which we will explain in Section 4.6.

5.2.2 Remove the Seasonal Components

As we predict, the seasonal components of 24-hours and 168-hours are removed. We can see that 24-hours period of electricity prices is very clear from ACF plot. After comparing two figures, we can see that the ACF becomes much smoother when we remove periodic components. After these preprocessing, we can regard our data set as shortly stationary and model them applying ARMA process.

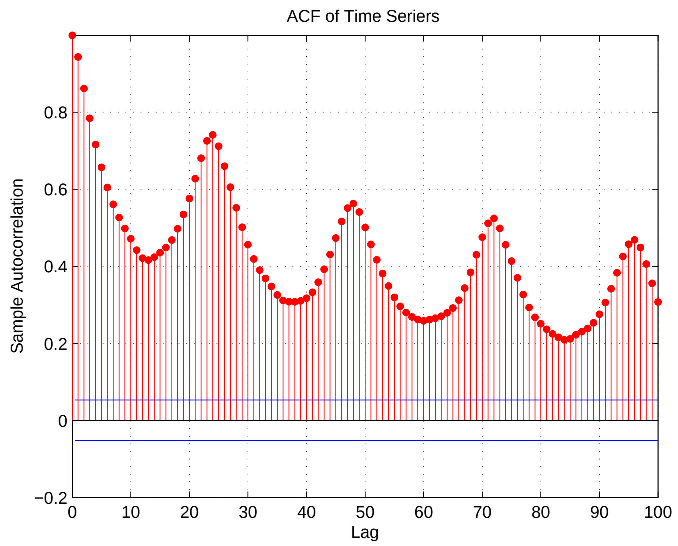

Figure 5: the ACF of Time Seriers after Removing Extreme Values

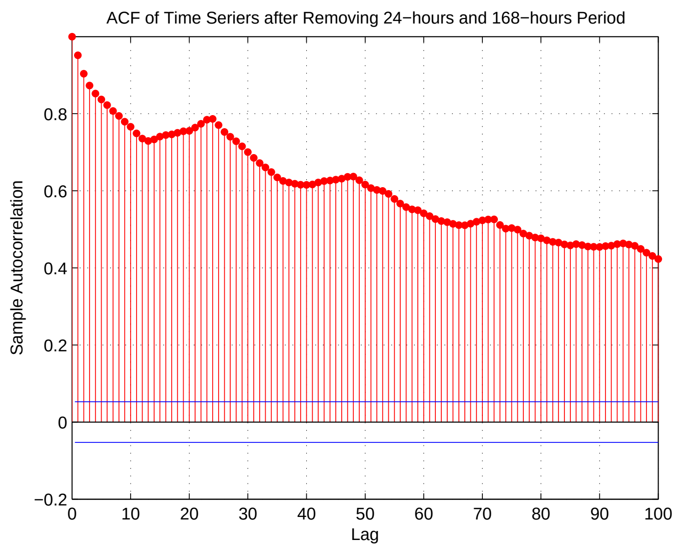

Figure 6: the ACF of Time Series after Removing Extreme Values and 24-hours and 168-hours Periodic Components

5.3 ARMA Model

We calculate AIC values using different orders of ARMA model.

Table 2: AIC Values using Different Parameters of ARMA Model

(p, q)	1	2	3	4	5	6	7	8
1	1.7530	1.7131	1.6990	1.6921	1.6888	1.6882	1.6829	1.6841
2	1.6891	1.6762	1.6762	1.6769	1.6780	1.6905	1.6787	1.6844
3	1.6786	1.6773	1.6790	1.6782	1.6796	1.6796	1.6485	1.6800
4	1.6771	1.6779	1.6797	1.6713	1.6816	1.6773	1.6480	1.6860
5	1.6788	1.6800	1.6769	1.6712	1.6736	1.6788	1.6483	1.6752
6	1.6803	1.6817	1.6652	1.6799	1.6788	1.6413	1.6820	1.6698
7	1.6822	1.6823	1.6784	1.6753	1.6506	1.6645	1.6768	1.6752
8	1.6837	1.6831	1.6742	1.6780	1.6729	1.6743	1.6639	1.6645

The smallest value is 1.641 from ARMA(6, 6) model according to above table. And when we increase the order of parameters, the AIC value becomes smaller. We calculate AIC values until p=25, q=25. The smallest value appears at ARMA(21, 24). But the model will be complex when we use the optimal order of model. And if we apply the optimal model, the mean squared error of the simulation will be large, which depends on error from estimation of the parameters of this model. To solve this problem we try to find sufficient and simple model through checking and comparing residuals from different models. We can see that AIC value of ARMA(2, 2) is not much bigger than the smallest value and smaller than other values around it. We explore the residuals of ARMA(2, 2), ARMA(6, 6) and ARMA(21, 24) as follows.

Table 3: Residuals Analysis for ARMA Model with Different Parameters

residuals	AIC value	Kolmogorov distance
ARMA(2, 2)	1.6762	0.1255
ARMA(6, 6)	1.6413	0.1304
ARMA(21, 24)	1.6007	0.1207

Above table shows that we can get smallest AIC value from ARMA(21, 24) model. For all ARMA models, however, the kolmogorov distances between residuals and normal distribution are quite close. Hence, we can conclude that higher order might get smaller AIC value, but it contributes limitedly to reduce distance between residuals and normal distribution. To decrease distance between residuals and normal distribution it is not enough to increase the order of ARMA model. So, we may need another type model to analyze residuals for ARMA model. And it is not a good choice to use high order ARMA model. In this case, we decide to try simpler model ARMA(2, 2) to deal with our data set.

Table 4: Estimated Parameter and Standard Error of ARMA(2, 2) Process

Parameter	Value	Standard Error	T-Statistic
ϕ_1	1.6587	0.04232	39.1943
ϕ_2	-0.66046	0.041914	-15.7575
θ_1	-0.73096	0.04404	-16.5974
θ_2	-0.1625	0.025755	-6.3092

It is obvious that both parameters of AR and MA are significant from T-statistic. Since the absolute value of T-statistic is larger than 2, we can get significant conclusion.

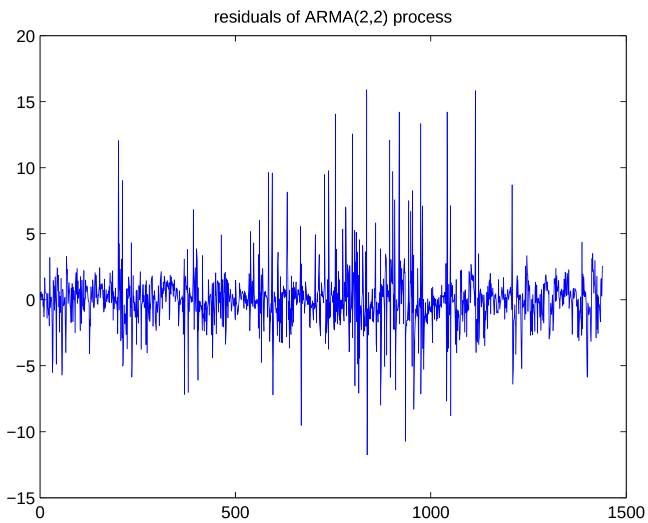

Figure 7: Residuals of ARMA(2, 2) Process

It seems that the residuals are stationary and volatility clustering. Then we check the ACF and PACF of residuals. We also check the ACF of squared residuals later.

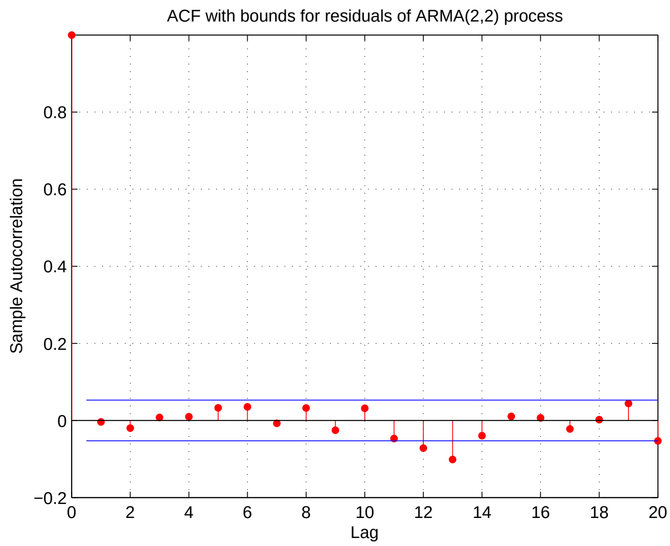

Figure 8: ACF with Bounds for Residuals of ARMA(2, 2) Process

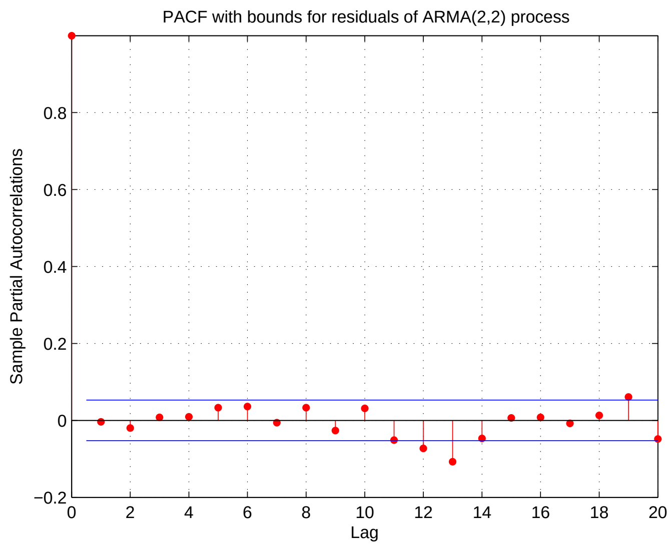

Figure 9: PACF with Bounds for Residuals of ARMA(2, 2) Process

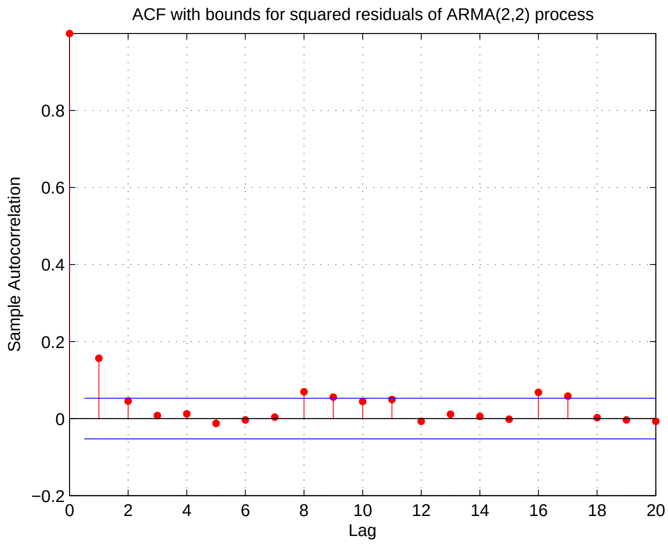

Figure 10: ACF with Bounds for Squared Residuals of ARMA(2, 2) Process

From ACF and PACF plots of residuals we can see that most values are within bounds, which are 95% confidence interval for Gaussion write noise. ACF for squared residuals shows that although the row data themselves are largely uncorrelated, the variance process exhibits some correlation. This indicates the possibility of a variance process close to being non-stationary and GARCH-type model may suit to this situation.

We also use formal test to check stationary for residuals and squared residuals.

Table 5: Results of Ljung-Box Test for Residuals

lag	H	P-value	State value	Critical Value
5	0	0.7967	2.3649	11.0705
10	0	0.6174	8.1168	18.3070
15	1	0.0017	36.1535	24.9958

Table 6: Results of Ljung-Box Test for Squared Residuals

lag	H	P-value(10^{-6})	State value	Critical Value
5	1	0.2016	39.3537	11.0705
10	1	0.0529	53.7993	18.3070
15	1	0.6307	57.6728	24.9958

Ljung-Box test may confirm the conclusion from plots. It indicates that the residuals of ARMA(2, 2) model are stationary, at least shortly stationary (within 12 lags). However the variances do not hold his property. Next, we will compare the residuals with normal distribution.

From following two plots we can see the empirical cdf of residuals is a bit far from normal distribution. The kolmogorov distance between residuals and normal distribution is 0.1255. According to Kolmogorov-smirnov test, we have to reject the null hypothesis that the residuals follow normal distribution at 5% significance level. Hence, the residuals of ARMA(2, 2) is stationary but a bit far from normal distribution and variance shows non-stationary. Next step we will use both GARCH and EGARCH model in order to describe volatilities.

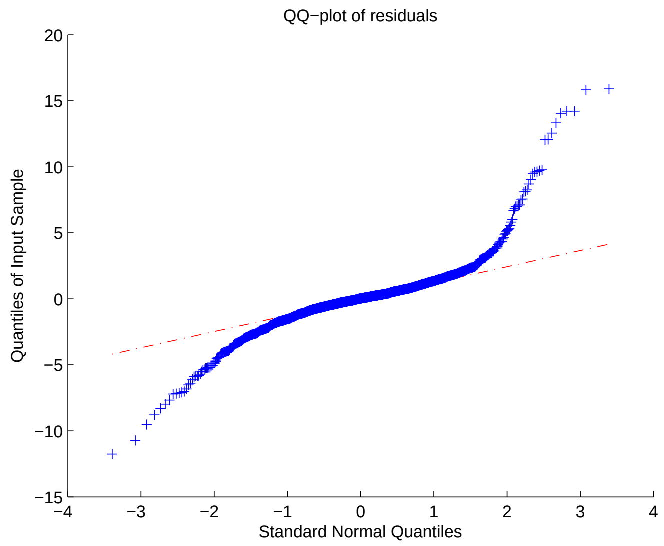

Figure 11: QQ-plot for Residuals of ARMA(2, 2) Process

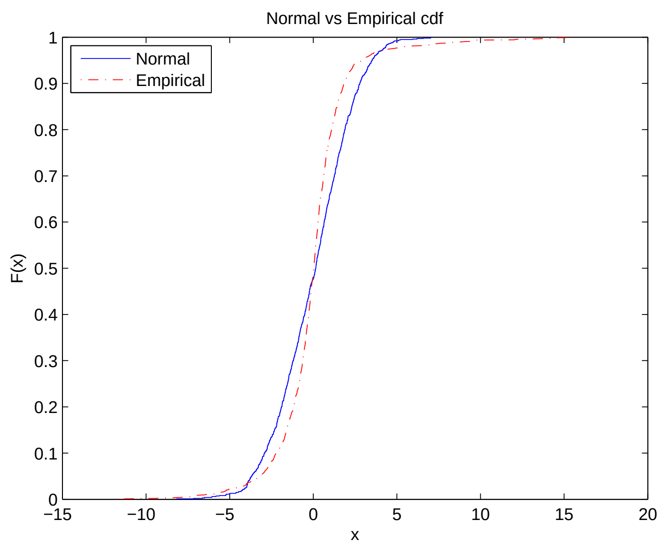

Figure 12: Empirical cdf for Residuals of ARMA(2, 2) Process vs Normal Distribution

5.4 ARMA-GARCH Model

Table 7: Estimated Parameter and Standard Error of GARCH(1,1) Process

Parameter	Value	Standard Error	T Statistic
ω	0.14599	0.010132	14.4090
α	0.90769	0.0049621	182.9240
β	0.072466	0.0056852	12.7463

It is clear that ω, α and β are significant from T-statistic. And we get the plot of residuals of ARMA-GARCH process.

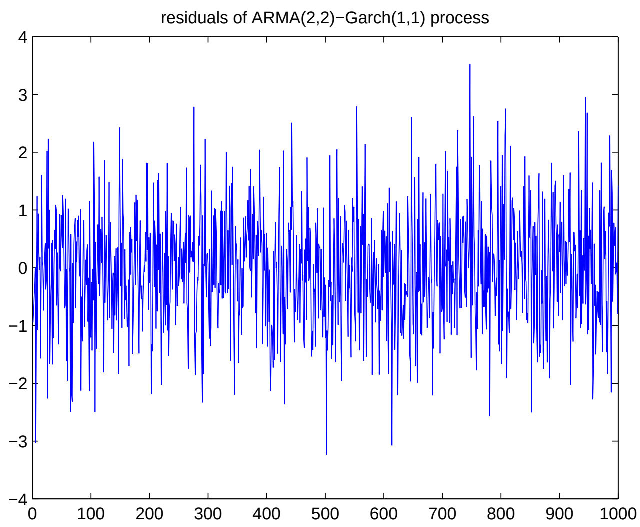

Figure 13: Residuals of ARMA(2,2)-GARCH(1, 1) Process

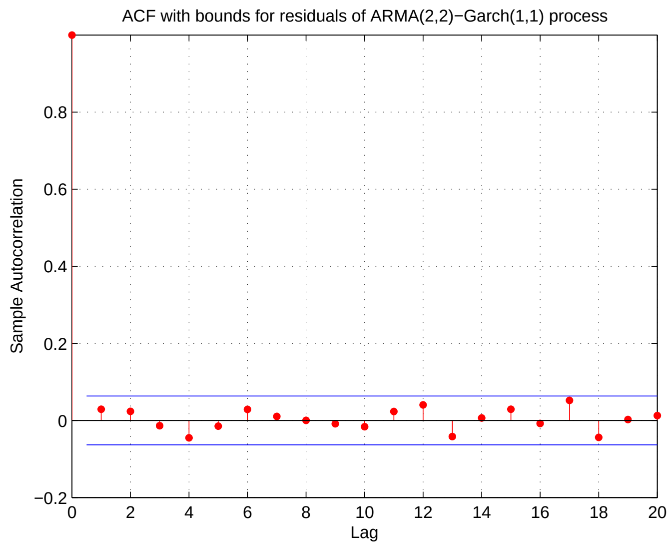

Figure 14: ACF for Residuals of ARMA(2, 2)-GARCH(1, 1) Process

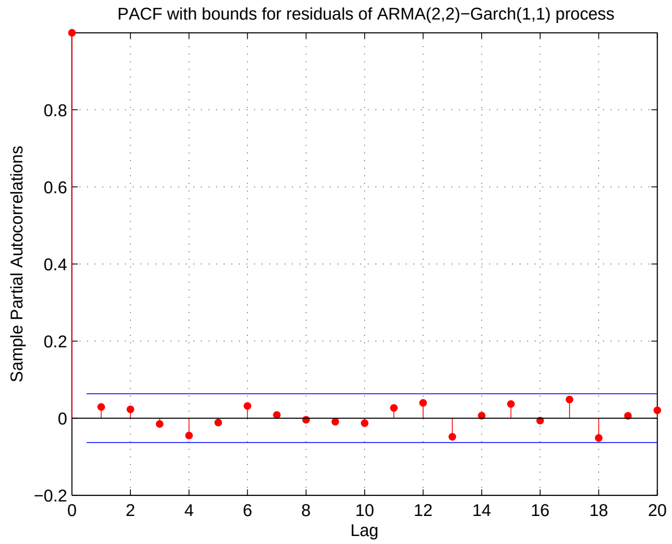

Figure 15: PACF for Residuals of ARMA(2, 2)-GARCH(1, 1) Process

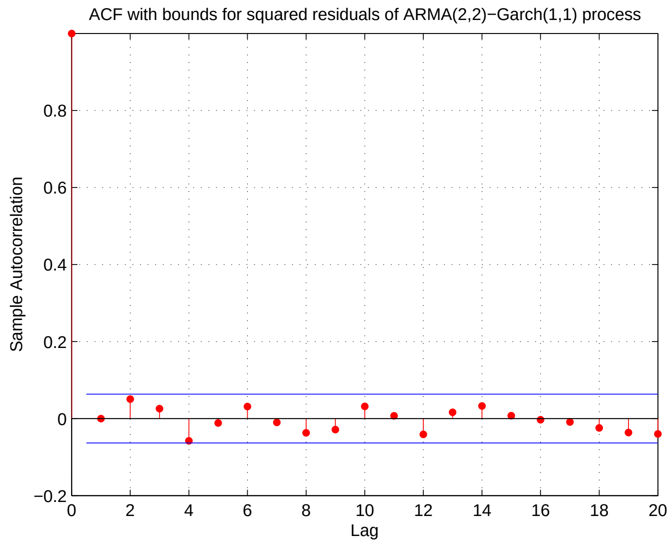

Figure 16: ACF for Squared Residuals of ARMA(2, 2)-GARCH(1, 1) Process

We can see all values are within bounds. We may draw a conclusion that both residuals themselves and variance are uncorrelated. It also shows that the GARCH(1, 1) model sufficiently explains the heteroscedasticity. We also use formal test to check stationary for residuals and squared residuals.

Table 8: Results of Ljung-Box Test for Residuals

lag	H	P-value	State value	Critical Value
5	0	0.5721	3.8443	11.0705
10	0	0.8835	5.1128	18.3070
15	0	0.8200	9.9953	24.9958

Table 9: Results of Ljung-Box Test for Squared Residuals

lag	H	P-value	State value	Critical Value
5	0	0.2480	6.6505	11.0705
10	0	0.3755	10.7721	18.3070
15	0	0.5218	14.0496	24.9958

Ljung-Box Test can confirm the conclusion of plots.

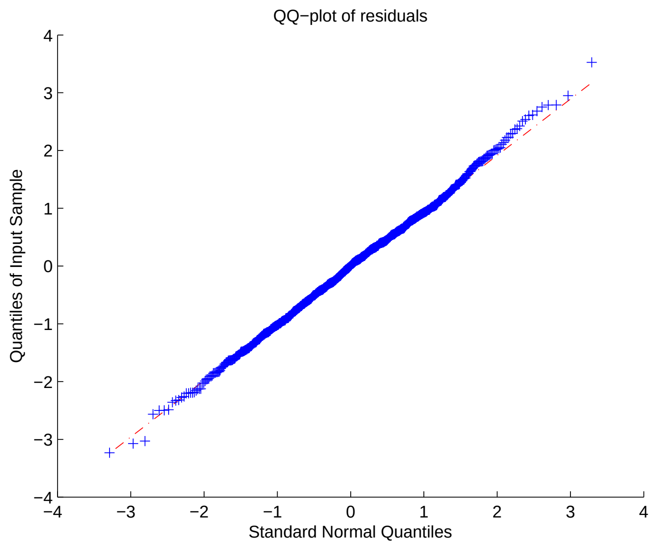

Figure 17: QQ-plot for Residuals of ARMA(2, 2)-GARCH(1, 1) Process

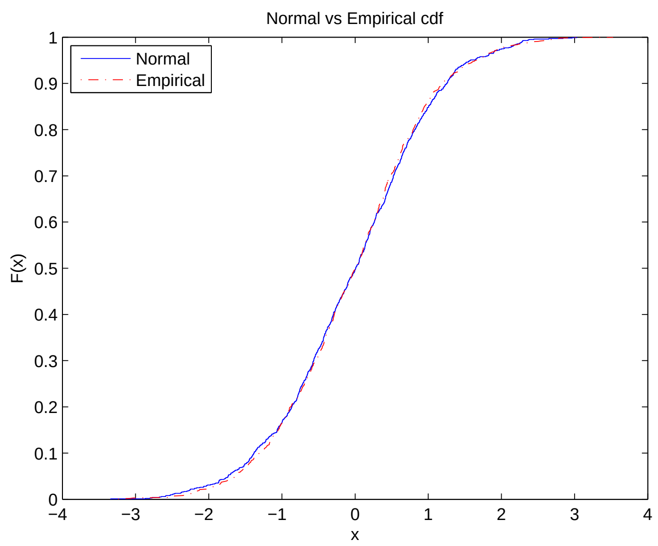

Figure 18: Empirical cdf of ARMA(2, 2)-GARCH(1, 1) Process vs Normal Distribution

The residuals are quite close to normal distribution.

We also check the Kolmogorov distance between residuals of our model and normal distribution.

Table 10: Residuals Analysis for ARMA-GARCH Model

residuals	Kolmogorov distance	P-value	Test (h)
ARMA-GARCH model	0.0255	0.5254	0

where $h = 0$ indicates that we can not reject the null hypothesis at the 5% significance level. From above plots and tables, it can be observed that the distance between residuals of ARMA-GARCH model and normal distribution become much smaller than residuals of ARMA model.

5.5 ARMA-EGARCH Model

In this section I employ EGARCH model to deal with residuals of ARMA(2, 2) process.

Table 11: Estimated Parameter and Standard Error of EGARCH(1, 1) Process

Parameter	Value	Standard Error	T Statistic
ω	0.098413	0.0063656	15.4600
α	0.95767	0.003133	305.6779
β	0.18685	0.0097585	19.1472
γ	0.058309	0.0080679	7.2272

It is obvious that ω, α, β and γ (leverage) are significant from T-statistic. The plot of residuals for ARMA-EGARCH process is following. To analyze them we obtain ACF and PACF of residuals and squared residuals. Later formal test of stationary is used. We also compare empirical distribution and normal distribution.

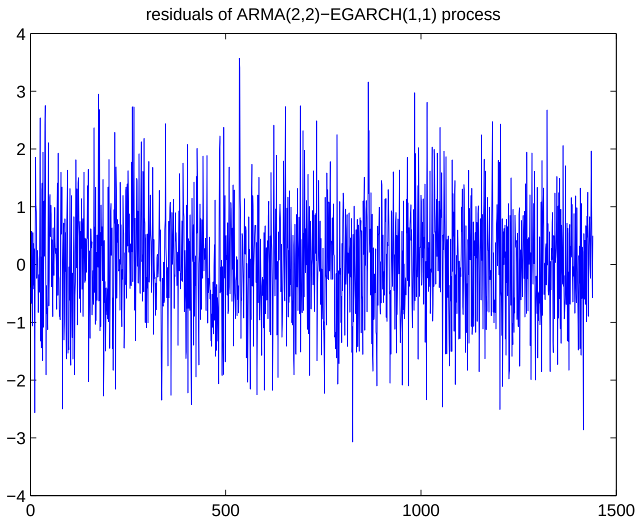

Figure 19: Residuals of ARMA(2, 2)-EGARCH(1, 1) Process

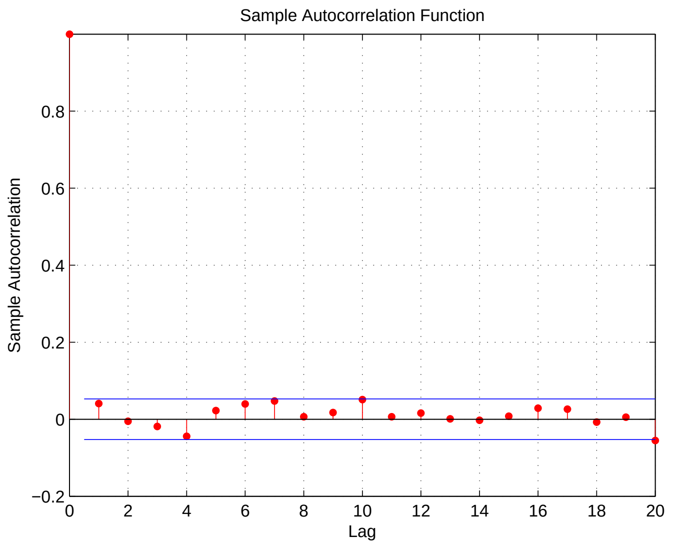

Figure 20: ACF for Residuals of ARMA(2, 2)-EGARCH(1, 1) Process

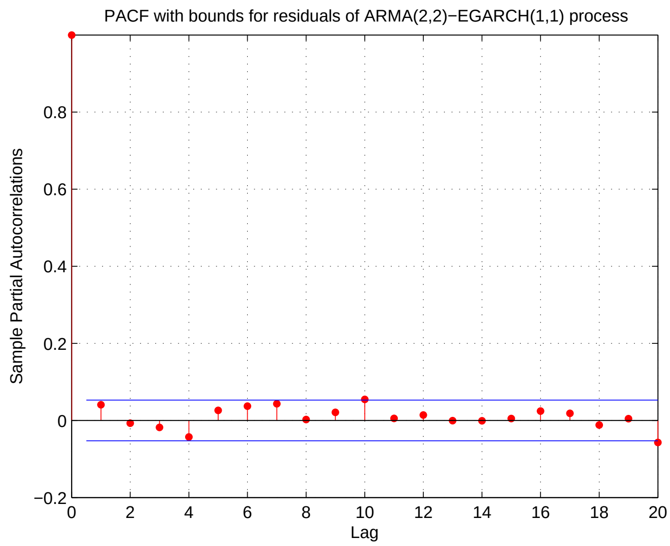

Figure 21: PACF for Residuals of ARMA(2, 2)-EGARCH(1, 1) Process

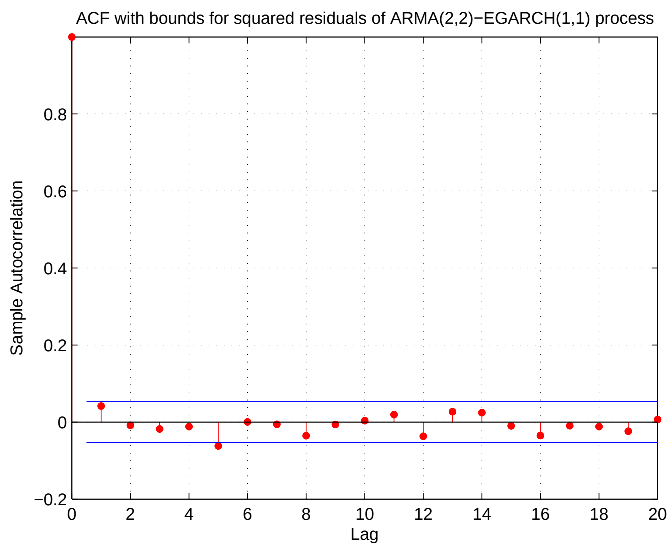

Figure 22: ACF for Squared Residuals of ARMA(2, 2)-EGARCH(1, 1) Process

From ACF and PACF plot we can see all values are within bounds. We may draw a conclusion that both residuals themselves and variance are uncorrelated. It also shows that the EGARCH(1,1) model sufficiently explains the heteroscedasticity. We also use formal test to check stationary for residuals and squared residuals. They confirm our conclusion from plots.

Table 12: Results of Ljung-Box Test for Residuals

lag	H	P-value	State value	Critical Value
5	0	0.8316	2.1247	11.0705
10	0	0.8647	5.3758	18.3070
15	0	0.8646	9.2410	24.9958

Table 13: Results of Ljung-Box Test for Squared Residuals

lag	H	P-value	State value	Critical Value
5	0	0.3780	5.3209	11.0705
10	0	0.1369	14.8695	18.3070
15	0	0.4081	15.6149	24.9958

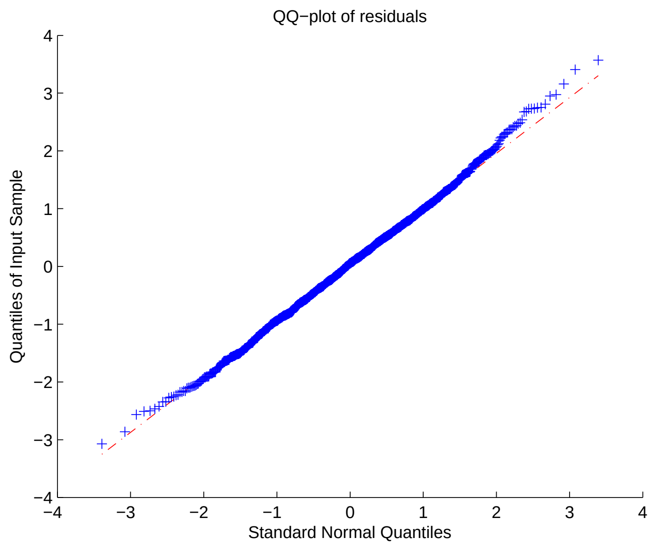

Figure 23: QQ-plot for Residuals of ARMA(2, 2)-EGARCH(1, 1) Process

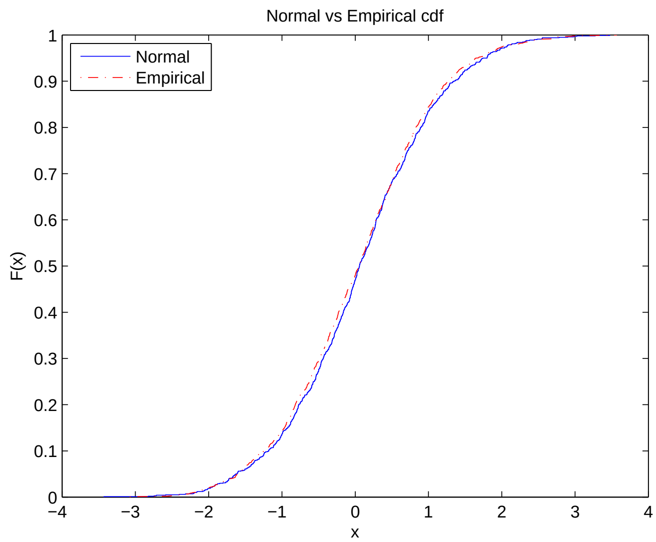

Figure 24: Empirical cdf of ARMA(2, 2)-EGARCH(1, 1) Process vs Normal Distribution

From above plots we can get the conclusion that the residuals are independent and very close to normal distribution. Next we check the Kolmogorov distance between residuals of our model and normal distribution.

Table 14: Residuals Analysis for ARMA-GARCH Model

residuals	Kolmogorov distance	P-value	Test (h)
ARMA-EGARCH model	0.0182	0.8878	0

where $h = 0$ indicates that we can not reject the null hypothesis at the 5% significance level. From above plots and tables, it can be observed that the distance between residuals of ARMA-EGARCH model and normal distribution become much smaller than residuals of ARMA model. In addition, the Kolmogorov distance in EGARCH model is slightly smaller than one in GARCH model.

5.6 Extreme Value Distribution Model

The time series indicates a behavior that makes the price spikes tend to accumulate. This characteristic should be taken into consideration in model of extreme values. The observations that are larger than threshold u will be treated into two steps. Firstly, we fit spikes into GP-distribution and obtain the parameters. Secondly, we simulated each extreme value using parameters and add them to the simulated processes of above models. We have assigned the positions of extreme values when we remove them. We add the simulated extreme values on the signed positions earlier.

Firstly, we get the mean excess plot of our data set. Our mean excess plot follows linear, which shows that we can fit extreme values with GP distribution. In GP distribution, it is important to find a threshold. We will choose threshold based on mean excess function and mean residual life plot as follows.

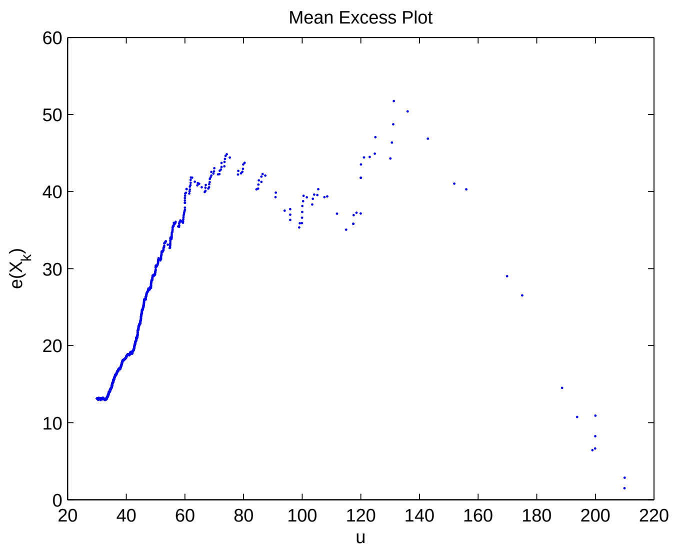

Figure 25: Mean Excess Plot

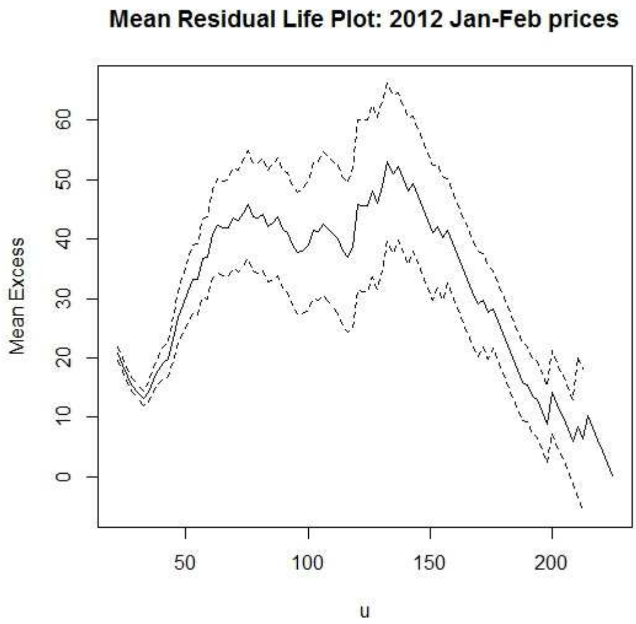

Figure 26: Mean Residual Life Plot

It is not simple to find a threshold in practice. Above figure shows the mean residual life plot with approximate 95% confidence intervals for the electricity price. At first glance, we may conclude that there is no stability until $u = 130$, after which there is approximate linearity. However, there are only 20 exceedances above the threshold $u = 130$, too few to get meaningful conclusions. In addition, from the plot, large values of u is unreliable because of the limited amount of data on which the estimate and confidence interval are based. We intend to choose the threshold about 70. From mean excess plot, we also find it is better to choose threshold u=70. We remove 84 observations above this threshold u.

For every observation above our threshold, we use its previous value instead of current one. The size of our data set after moving extreme values is the same with original time series. So, this data set still possess the same periodic property with the original electricity price. We also assign the position in each extreme value for later simulating.

5.7 Simulation of GARCH-Type model and Extreme Value Distribution Model

We simulate our GARCH-type model and GP distribution model separately. At first, we simulate time series using estimated parameters for ARMA-GARCH and ARMA-EGARCH model. On the other hand, we use the GP distribution to simulate extreme values. Then we obtain the final simulation by following way. On Section 4.6 we assign the position of each extreme value when we remove them. We add our estimated extreme values to the positions that assigned earlier.

From following figures we find that two simulated processes with extreme values are similar with the original data. It can be observed that the simulated time series is quite similar with the original data set for both ARMA-GARCH model and ARMA-EGARCH model with GP model. But we know that there are also some distances between original data and simulated processes. The main reason is that the residuals are not exactly normal that is our assumption when we fit our models. They look very slight since we add the simulated extreme values at the same position with our data, and the non-extreme values are in a small range.

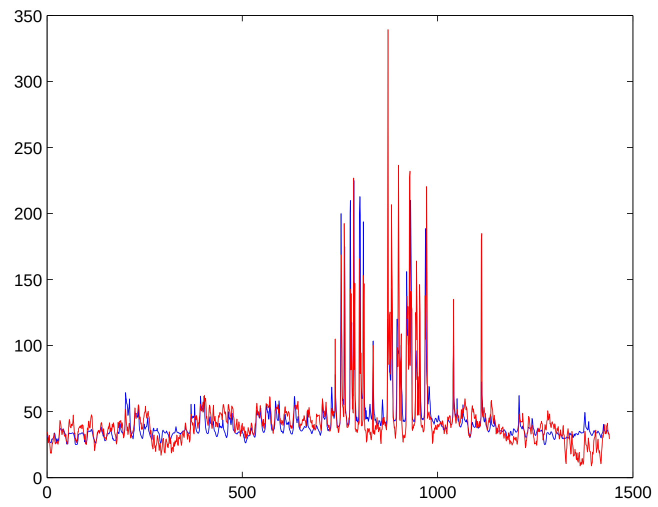

Figure 27: Simulated Processes for Data From January 2012 - February Using ARMA-GARCH Process and EVT

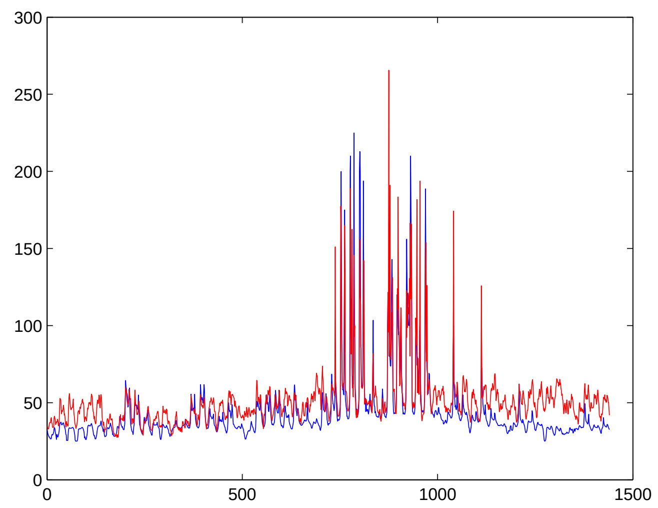

Figure 28: Simulated Processes for Data From January - February 2012 Using ARMA-EGARCH Process and EVT

6 Conclusions

We can draw the conclusions as follow:

Predictable periods are appropriate to the data set. After removing them the data set tend to stationary. We can see that the ACF become much smoother then before.

For this kind of shortly stationarity, ARMA can successfully decompose autoregressive process and moving average components, then obtain stationary residuals.

On the other hand, GARCH-type models are quite necessary since the volatility is still not independent.

It is not easy to say that GARCH or EGARCH is better to fit the time series. The residuals from both models are quite close to normal distribution.

AIC may be over estimated for order selection.From this criterion we can get the best choice of parameter order, however, more parameters will increase errors in prediction in practice. In stead of it, fewer parameters are chosen based on AIC values.

GP distribution is fit for extreme values of electricity prices.

We also need develop following issues in the future:

The residuals are assumed to be normal distribution. Actually, there is more or less distance between residuals and normal distribution, which may affect our result.

Order selection for ARMA model is a complex problem. We may need further research and find a general way for electricity price.

Finding threshold in GP distribution is quite important. We will compare different thresholds and results, then study some criteria to judge different thresholds.

When we add the simulated extreme values, it is better to find the position randomly. How to estimate the occasion that may appear extreme value is a future task.

External factors that affect electricity price are not ignored completely. How to indicate these factors and remove them becomes a task in the future.

References

[1] P. J. Brockwell and R. A. David. . *Time Series: Theory and Methods.* Springer, New York, NY, USA, 2 edition, 1990

[2] Beth Andrews *Rank-Based Estimation for GARCH Processes* Northwestern University

[3] Chang Su *Application of EGARCH Model to Estimate Financial Volatility of Daily Returns: The empirical case of China* University of Gothenburg, Sweden, 2010

[4] Christian Francq and Jean-Michel Zakoian. *GARCH Models: Structure, Statistical Inference and Financial Applications* University Lille3, France, 2010

[5] Christopher J.Neely and Paul A. Weller *Predicting Exchange Rate Volatility: Genetic Programming vs. GARCH and RiskMetrics* Working Paper 2001-009B

[6] Chunyan Zhou, Chongfeng Wu, Hailong Liu, Fubing Liu *A New Method to Choose the Threshold in the POT Model* Financial Engineering Research Center for Shanghai Jiaotong University 2000

[7] Henrik Hult and Filip lindskog *Mathematical Modeling and Statistical Methods for Risk Management Lecture Notes* 2007

[8] Kallberg, D. *Modeling electricity prices with seasonal long memory time series.* Master thesis, Mathematical Sciences, Gothenburg.2006

[9] Niklas Axelsson, David Eriksson, Osman Malik and Jens Rï¿½nnqvist *Stokastisk modellering av elpriser med autoregressiva processer, Ornstein-Uhlenbeck-processer och extremvï¿½rdesteori* candidate thesis, Mathematical Sciences, Gothenburg.2010

[10] Patrick Burns *Robustness of the Ljung-Box Test and its Rank Equivalent* 6 October 2002

[11] Ramazan Gencay, Faruk Selcuk and Abdurrahman Ulugulyagci *EIVM: A Software Package for Extreme Value Analysis in MATLAB* Studies in Nonlinear Dynamics and Econometrics, October 2001, 5(3):213-239

[12] Reinaldo C.Garcia Javier Contreras, Marko van Akkeren *A GARCH Forecasting Model to Predict Day-Ahead Electricity Prices* September 27, 2003

[13] Sango, H. *Modeling electricity prices in the German market. Master thesis* Mathematical Sciences, Gothenburg.2008

[14] Stuart Coles and Anthony Davison *Statistical Modelling of Extreme Values* 2008

[15] Stuart Coles *An Introduction to Statistical Modeling of Extreme Values*Springer, 2001

[16] Stuart Thomas and Heather Mitchell *GARCH Modelling of High-Frequency Volatility in Australia's National Electricity Market*

[17] Svetlozar T.Rachev, Stoyan V.Stoyanov, Almira Biglova, Frank J.Fabozzi *An Empirical Examination of Daily Stock Return Distributions for U.S. Stocks*

[18] Nord pool spot. http://www.nordpoolspot.com/, Apr 2012.