NTNU – Trondheim
Norwegian University of
Science and Technology

The use of extreme value statistics in risk management of the electricity market

Ampem Darko Owusu

Master of Science in Mathematics (for international students)
Submission date: June 2013
Supervisor: Arvid Næss, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Abstract

In this thesis, we investigate the success of extreme value theory in managing electricity price risk. We specifically deals with the behaviour of the tails of financial time series.
The theory provides well established statistical models for which extreme risk measures like the Value at Risk, Expected Shortfall and Return level can be computed. We use daily electricity price data from Nord Pool and compare distributions that effectively estimates the tail quantile. We propose a new method which employs extreme value for estimating the tail risk measure. Our method provides the exact empirical distribution without independence assumption and applicable to non-stationary data. This method is briefly known as ACER. We show that the recently proposed approach gives better tail quantile estimates, have nice features and it is easy to implement.

Preface

This Master Thesis, carried out in autumn 2012 and spring 2013 is the last requirement of my MSc (Mathematics) degree under the Department of Mathematical Sciences at the Norwegian University of Science and Technology. The work is supervised by Professor Arvid Næss.

Ampem Darko Owusu NTNU, Trondheim 1st June, 2013.

Acknowledgements

I would like to thank all those people who have encouraged and supported me in writing this thesis. I own a considerable debt especially to my supervisor, Professor Arvid Næss for the numerous discussions and meetings with me, his help and advice. My sincere appreciation goes to Professor Sjur Westgaard for providing me with the electricity price data and to Oleh Karpa for for his help with the ACER implementation in Matlab. I would also like to thank my family and good friends: Paul Wilpert, Ron Drodzik and Petter Nygård.

Contents

List of Figures

List of Tables

1 Introduction

Worldwide financial markets have experienced significant fluctuations over several of the past few years. Many have come to question the viability of existing risk management methodology in dealing with outcomes in these volatile circumstances. There are those who ask, how bad can things really get before a solution is found? It appears that present systems must be restructured to include reliable methods and strategies to address quickly the potential problems, regardless of how unpredictable these financial instabilities might be. The main problem have to do with modeling events that fall outside the range of available observations. It is therefore necessary to depend on well founded methodology which Extreme Value Theory provides the key.
Extreme Value Statistics, the statistical analysis of extreme events, is the key to many problems in the applied sciences. Extreme Value Theory (EVT) provides a concrete theoretical background on which statistical models describing extreme events can be built. Its main purpose is to provide asymptotic models with which the tail of a distribution can be modeled. It is applied in modeling events that occur with a very small probability, hence a tool for modeling and managing risk, since risk, in practice, occurs with low probability. It has some applications in insurance, finance, hydrology, and other institutions dealing with extremes.

Risk is the possibility of suffering from harm or loss. Its probability distribution has never been observed exactly, although data from past losses are accessible. Since financial risk management is primarily concerned with tail quantiles, extreme quantiles are of much interest, because being able to assess them well translates into the ability of effectively managing extreme financial risk. EVT plays an outstanding role in determining whether events (extreme) that occurred are in fact extremely low probability events, or just recurring situations.

In this thesis, we aim to investigate the status of development of EVT in risk management of the electricity market, and comparing results obtained by traditional parametric methods to those obtained by applying a recently proposed method. The traditional estimation methods, based on density estimation, are not very favourable in assessing extreme quantiles, since they produce good fit in areas where most of the data fall, but in reality, only fewer observations fall. Typical method including the Normal distribution, the student t distribution models will be briefly discussed and use as benchmark, but our concern will be narrower to the recently proposed theorem. This work is divided into six sections. Chapter 2 discusses the electricity market structure and its main features, including the most common forecasting models for conventional spot price. The time series models considered in addition to Extreme Value Theory, Peak Over Threshold, Average Conditional Exceedance Rate is discussed in chapter 3. The data used in this thesis is presented in chapter 4 and relevant analysis of data carried out in chapter 5. We then look at the role of EVT in managing price risk of the electricity market in chapter 6.

2 The Electricity Market

The deregulation of the electricity market has led to perfect competition of the previously monopolized market, thereby reducing the risk of loss and creating a platform for spot electricity and their derivatives trading, just as stocks and securities traded on the financial market. Electricity cannot be stored on a large scale, as a result, electricity generated is balanced with electricity used each time. The market is, therefore, highly volatile when compared to any other security since its product is traded primarily for consumption.

In competitive market, prices are no longer regulated by the market operators but determined by the market through demand and supply. Generally, large electricity consumers are more concerned about unexpected price increases whereas producer's main interest is to monitor extreme price drops. As a result of this extreme behaviour, the increased interest in risk management of electricity positions as well as assessment of worst-case scenarios in the electricity market stresses the important of forecasting models which Extreme Value Theory has produced significant results.

2.1 Nord Pool

The Nordic countries: Norway, Sweden, Denmark and Finland operate a single market for trading electricity known as the Nord Pool. It was started in 1992 after the Norwegian government decision of deregulating the electricity market came into effect. Sweden later joined in 1996 followed by Finland and then Denmark. Each of Nord Pool member country retains the right to set the market rules and each country had different reasons of liberalization of their markets. Currently, Nord Pool is the world largest market for trading electricity, handling over 70% of the total power consumption traded in the Nordic regions.

Two separate markets operating on Nord Pool are the Physical and Financial market. The physical commodity market comprises the Elspot and Elbas. Elspot is traded for the following 24 hours while Elbas is an hour-ahead market. Elspot provides the electricity price for the system. The financial market covers derivatives like forwards, future, and options. Any credit worthy institution can trade on Nord Pool financial market irrespective of where they are operating.

2.2 Electricity Price Formation and Features

The spot market, a short-term future contract, is a market of physical delivery of electricity. Because electricity cannot be stored, the electricity spot market are more of auction-based type

whose prices are determined by the market equilibrium model where demand and supply of all participants are set one day ahead. The spot market receives bid and offers from alike participants to determine prices and balance opposing sides.
Careful analysis of demand and supply reveals price series tend to show large complexity than might be expected, since in some instances power plants have to be scheduled in order to meet fluctuations in demand. Hence spot prices time series tends to be more intricate in structure than just a reshaping demand exhibiting marginal cost.
The electricity price differs in periodicity. They exhibit seasonality over hours, days, weeks and years. Weather conditions affect electricity prices depending on the location. Whilst cold countries have greatest demand of electricity during the winter (due to heating), there might be reversible situation in other parts of the world.
In order to have a clear picture of seasonal variation effects over years, we will use the electricity prices of a 17 year period which captures the spring, summer, autumn and fall seasons including day and night variation.
Another important feature of spot price is stationarity: Contrary to stock prices, electricity prices tend to exhibit stationarity over short intervals. This implies that they are mean reversion over short period and from our time series plot, we have observed that prices tend to fluctuate about a stable level but follows no specific trend.

2.3 Spot Price Forecasting Models

In deregulated market, electricity spot prices are highly random events which are determined by the market and other factors including weather condition, fuel price, demand and supply and operating cost. This has call for proper forecasting models capable of capturing all price extremities. Numerous methods that achieve success in electricity price forecasting concentrate mainly on normal price forecasting but give less attention to spikes. As a result, accurate prediction of spot prices may be somehow difficult to achieve when there is extreme price behaviour and seasonal variation. Because spot price could be as high as 50 (which is more than ten times higher than that of many products obtained from energy), the choice of forecasting models geared towards extreme spikes are critical.

Analysis of the European electrical power spot prices by [4] revealed that electricity prices exhibit seasonal behaviour in daily, weekly and yearly cycles. It was also noticed that spot prices tend to show stationary behaviour and are mean reverting to a trend, moreover they show impressive spikes in upward jumps followed by swift return which is nearly around the same level and non-Gaussian which tends to have slightly positive skewed distribution and strong leptokurtic.

The mean-reverting process with trend proposed by [2] model the daily average or periodic spot prices. There was a slow speed of the mean reversion after jump which was a drawback.

Since spot prices in the electricity markets, in our case Nord pool, are extremely volatile which is time varying with evidence of heteroscedascity both in conditional and unconditional variance, modeling the residuals with GARCH is critical. This model have a limit due to extreme values and can be corrected when a regression model with implicit jump component for prices and leptokurtic distributions innovations are used.

A more accurate description of electricity prices where utilised the assumption an AR-GARCH price (return) phenomena that includes a seasonal volatility component and models the residuals with extreme value theorem was proposed by Byström in [3]. His theorem inculcates all the important price feature models for spot modeling. Upon reviewing several literatures, we will be considering the Byström innovation in this thesis.

3 THEORY

3.1 Value At Risk

Sophisticated financial trading systems has led to the increase of awareness of harm associated with very large losses. The most widely tool used in risk management purposes is the VaR. VaR tends to quantify the probability of returns (losses), and it became very renowned when its Risk Metrics system was published by J.P. Morgan in 1994. VaR quantifies probability of loss by providing a single number about loss at particular point but provides less information about the loss that exceeds it. For details about potential size of loss, we will be considering [10] innovation where the use of conditional tail expectation or Expected Shortfall (ES) instead of VaR was proposed.

The non-parametric VaR estimation methods includes Historical simulation, Monte Carlo simulation, the Variance-Covariance method and the Exponential Moving Average Method. It is important to highlight some few points of the non-estimation method but we are not getting into the details of that.

The Historical Simulation (HS) is a powerful VaR method because its simple to use. It works by assuming an unknown price distribution, further regarding future and past risks as similar. It relies less on the stochastic nature of the data. It enhances non-normal distribution thereby accounting for fat-tailed distribution. However, the assumption that future and past risk is the same constitutes major drawback, that is extreme observations that remained unnoticed in the past cannot be predicted in the future. In this era of technology advancing, there is rapid increase in the amount of data in all sectors. For very large sample data, the HS approach functions poorly.

Instead of relying on past values of market factors for prediction, The Monte Carlo Simulation technique builds on an assumed knowledge about the stochastic mechanism used in generating the returns, by employing historical values for unknown parameters estimation. A set of possible future values of the asset are then generated, using stochastic simulation. The main advantage of this method is that, there is the possibility of enlarging the data set. On the other hand, its major drawback is based on the distributional assumptions of returns.

In the variance-covariance approach, emphasis is made on summary statistics based on the magnitude of price volatilities and later utilizes correlation between price dynamics to estimate for potential losses. This method underestimates risk at high quantiles and for fat tailed data we are considering, normality assumption of which this approach works on is rejected. The normal distribution case is discussed later in this thesis.

Broadly speaking, the most serious drawback of using VaR as a risk management tool in the electricity market lies in the fact that, normality assumption is invalid when prices or observations have heavy tails, which are characterised by extreme events left outside the bounds of a normal distribution in the VaR modeling. We therefore tackle the issue of normality assumption of the prices or return series through the introduction of distribution based on the modeling of tail quantiles, or by applying the newly statistical extreme distribution statistics approach such as the Extreme Value Theory.

3.2 Extreme Value Theory

We will consider a series $(X_1, ...X_n)$ of random independent and identically distributed (iid) variables with cumulative distribution function (cdf) $F(x)$ whose extreme value is given as

$$M_n = \max(X_1, ..., X_n), \tag{3.1}$$

The exact distribution of M_n by is given

$$P(M_\eta \leq \eta) = P(X_1 \leq \eta)...P(X_N \leq \eta) = F_X^N(\eta), \tag{3.2}$$

In practise, the distribution F is not usually known making equation (3.2) somehow wasteful [5]. Obviously, small errors in estimating F could lead to large errors in estimating $F_X^{(N)}$.

3.3 The Generalized Extreme Value Distribution (GEV)

The basic idea is to explore the behaviour of F^η as $\eta \to \infty$,and hence $M_\eta \xrightarrow{P} \infty$, but this not sufficient to explore the the behaviour as the distribution of M_η degenerates to the upper end point. Therefore M_η has to be standardized to achieve a non-degenerate limit.

If for a non-degenerate distribution function G, There exist constants $a_n > 0$ and b_N, such that

$$M_n^* = \frac{M_n - b_n}{a_n} \tag{3.3}$$

and

$$P\{M_n^* \leq \eta\} \to G(\eta) \tag{3.4}$$

Then G belongs to one of the following families:

$$\mathrm{I} : G(\eta) = \exp\left\{-\exp\left[-\left(\frac{\eta - b}{a}\right)\right]\right\}, \quad -\infty < \eta < \infty; \tag{3.5}$$

$$\mathrm{II} : G(\eta) = \begin{cases} 0, & \eta \leq b, \\ \exp\left\{-\left(\frac{\eta-b}{a}\right)^{-\alpha}\right\}, & \eta > b \end{cases} \tag{3.6}$$

$$\mathrm{III} : G(\eta) = \begin{cases} 0, & \eta < b, \\ \exp\left\{-\left[-\left(\frac{\eta-b}{a}\right)^{-\alpha}\right]\right\}, & \eta \geq b \end{cases} \tag{3.7}$$

for parameters $a, b > 0$, and $\alpha > 0$ for families II and III. The distributions I, II and III are the Gumbel, the Fréchet and the Weibull distribution respectively. These three distributional classes are called the Extreme Value Distribution and can be expressed in Generalized Extreme Value (GEV) as:

$$G(\eta) = \exp\left\{-\left[1+\xi\left(\frac{\eta-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \tag{3.8}$$

defined on $\eta : 1+\xi(x-u)/\sigma > 0$ where $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \xi < \infty$ and with location parameter μ, scale parameter δ and shape parameter ξ. The shape parameter is also the tail index and it determines the tail behaviour of $G(\eta)$. Assuming $w = \frac{n-u}{\sigma}$ in equation (3.8), We can say that w belongs to the domain of attraction of $G(\eta)$. In cases where the tail of the distribution F declines by power function, $G(\eta)$ is of Fréchet type and $\xi > 0$. Fat-tailed or heavy-tailed is used to describe distributions in this type of domain, (examples include Cauchy, Pareto). If F has finite tail, $G(\eta)$ is Weibull with $\xi < 0$. Distributions in this attraction domain have some kind of upper bound support, (examples are beta and uniform distributions). For exponentially declining tail of the distribution F, $G(\eta)$ is Gumbel and $\xi = 0$. For such scenario, the domain of attraction are thin-tailed, (examples are gamma, exponential, log-normal and normal). Both the Fréchet and Gumbel distributions have infinite upper tails with the upper tail of the Gumbel shorter than of the Fréchet distribution. $\xi = 0$ is the limit of equation (3.8) as $\xi \to 0$ which gives the distribution function of the Gumbel family:

$$G(\eta) = \exp\left\{-\left[1+\xi\left(\frac{\eta-\mu}{\sigma}\right)\right]\right\}, \quad -\infty < z < \infty \tag{3.9}$$

One of the extreme value limit laws, equations (3.5) - (3.7) is fitted to the annual maxima of a series in the classical EVT whiles the GEV can be directly fitted to the sample maxima. The classical EVT is very narrow to be applied to modern extreme value problems which is the main weakness. With the necessary parameters in equation (3.8) estimated, extreme events can be forecasted.

3.4 Parametric and EVT Estimation of VAR and Expected Shortfall

Parametric Estimation

3.4.1 Normal Distribution

When the mean u and standard deviation σ of data is estimated, we calculate VaR_p as:

$$\text{VaR}_p(R) = u + \sigma z_p \tag{3.10}$$

where R is the return (loss) of our portfolio. Here we assumed normal innovations and the standard normal table gives the quantiles needed to calculate VaR. Since electricity prices are more heavy tailed, far from being normal, we expect estimated VaR under the normal distributional innovations to perform worse than the other models.

3.4.2 The Student t-Distribution

The Value at Risk under the student's t-distribution at quantile p is given as:

$$\text{VaR}_p(R) = u + \sqrt{\frac{v-2}{v}}\sigma t_{p,v} \tag{3.11}$$

The students $p-th$ quantile, like the Normal is estimated with the t-table. The degree of freedom v is estimated when AR-GARCH with t-distributed innovations is fitted to the data. The student's t-distribution is reliable when extreme values at very high confidence intervals are analysed.

3.4.3 Normal and Student t-Distribution with Time Varying Volatility

In the two parametric distributions described so far, focus have been on time constant models. If our interest is to estimate VaR for both the normal and student's t-distributions with time-varying volatility, we replaced σ in equations (3.10) and (3.11) with

$$\sigma_{t+1}^2 = (1-0.94)\varepsilon_t^2 + 0.94\sigma_t^2$$

where σ_{t+1}^2 expresses the volatility of first loss in the sample period.

3.5 EVT Estimation

3.5.1 VaR and Expected Shortfall

In this section of Value at Risk and Expected Shortfall estimation, help have been sought from the article [1].

For an extreme value distribution with right endpoint x_F and for some threshold u, If $x \geq u$, we have

$$\begin{aligned}
\bar{F}(x) &= P(X > u)P(X > x \mid X > u) \\
&= \bar{F}(u)P(X - u > x - u \mid X > u) \\
&= \bar{F}(u)\bar{F}_u(x - u) \\
&= \bar{F}(u)\left(1 + \xi\frac{x - u}{\sigma}\right)^{-1/\xi} \qquad (3.12)
\end{aligned}$$

This gives the tail of the distribution for known $F(u)$. Taking the inverse of equation (3.12) gives the tail quantile of the distribution or the VaR:

$$\mathrm{VaR}_\alpha = q_\alpha(F) = u + \frac{\sigma}{\xi}\left(\left(\frac{1-\alpha}{\bar{F}(u)}\right)^{-\xi} - 1\right) \qquad (3.13)$$

for $\xi < 1$ the ES is given by

$$ES_\alpha = \frac{\mathrm{VaR}_\alpha}{1-\xi} + \frac{\sigma - \xi u}{1-\xi} \qquad (3.14)$$

Analytical expression of VaR and ES can be defined as a function based on estimated GPD parameters. We therefore have:

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

Let n represent the total number of observations and N_u the number of observations above a threshold u. If we replace F_u by the GPD and $F(u)$ by $(n - N_u)$, we arrive at the estimator for tail probabilities;

$$\widehat{\mathrm{VaR}} = u + \frac{\sigma}{\xi}\left(\left(\frac{n}{N_u}p\right)^{-\xi} - 1\right) \qquad (3.15)$$

$$\widehat{\mathrm{ES}_p} = \frac{\mathrm{VaR}_p}{1-\xi} + \frac{\sigma - \xi u}{1-\xi} \qquad (3.16)$$

3.6 Peak Over Threshold

Classical approach to modeling of extreme events deal with block maxima. The events or data are classified in blocks. This method is wasteful if we have other data on extremes or in cases where large variation of data within blocks is present. A more recent than the classical EVT approach is the Peak Over Threshold (POT) which model observations that exceed a high certain threshold. The limiting distribution of the normalized excess of the threshold gives the expression:

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

As u approaches the endpoint of interest is the Generalized Pareto Distribution. The POT method basically involves the selection of a suitable threshold, the exceedance above which are fitted to the GPD function, then point and interval estimates of VaR and ES calculated [1].

3.7 Generalized Pareto Distribution

The Generalized Pareto Distribution (GPD), as an asymptotic distribution in EVT is based on idea that exceedances of high threshold are uncommon events to which we can easily apply the Poisson distribution.

Suppose we have sequence of independent random variables with common distribution function F, and let

$$M_n = max\{X_1,,,X_n\}$$

Denote an arbitrary term in the X_i sequence by X, and suppose F satisfies equation (3.4), so that for large n,

$$Pr\{M_n \leq \eta\} \approx G(\eta)$$

where $G(\eta)$ is given by equation (3.8). When the threshold u is large enough, the distribution function of $(X-u)$, conditional on $X > u$, is approximately

$$Pr(X-u > y \mid X > u) \approx H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \tag{3.17}$$

defined on $\{y : y > 0$ and $(1 + \frac{\xi y}{\tilde{\sigma}} > 0\}$ where $\tilde{\sigma} = \sigma + \xi(u-\mu)$. Equation (3.17) is the Generalized Pareto family, which in order wordsis a block maxima with an approximating distribution G, whose threshold excesses have corresponding approximate distribution within the Generalized Pareto family.

3.8 Return Period

When the Generalized Extreme Value is fitted to a dataset, It can be used to capture extremes and describe potential losses. The first way to do this is by the return level estimation where we define the return period of the occurrence of extreme event, predicting its magnitude and finally the return period estimation.

The return period R of a given value η of random variable X is given by

$$R = \frac{1}{1 - F_X(\eta)} \tag{3.18}$$

It follows that

$$P(Y \leq y) = 1 - \frac{1}{\lambda y} \tag{3.19}$$

where λ denotes the mean exceedance rate of the threshold in any interval.

Since Y is known distribution,we have from equation (3.17) that

$$\eta_R = \mu - \frac{\sigma}{\xi}(1 - (\lambda R)^{\xi}) \tag{3.20}$$

for $\xi \neq 0$, and

$$\eta_R = \mu + \tilde{\sigma}\ln(\lambda R) \tag{3.21}$$

for $\xi = 0$.
Assuming multi normally distributed maximum likelihood estimator, The variance of the return level is calculated using

$$Var(\eta_r) \approx \nabla\eta_R^T V \nabla\eta_R \tag{3.22}$$

where V is the variance-covariance matrix for $(\hat{\lambda}, \hat{\sigma}, \hat{\xi})$ and given as:

$$\nabla\eta_R = \left[\frac{\partial\eta_R}{\partial\lambda}, \frac{\partial\eta_R}{\partial\sigma}, \frac{\partial\eta_R}{\partial\xi}\right]^2 \tag{3.23}$$

Equations (3.21) gives the return period for the POT method for given values of ξ. The return periods of the ACER method can be found when we invert the empirical ACER function with the values of the parameters a, b, c and q for the exceedance rate of interest.

3.9 Threshold Selection

The selection of an appropriate threshold is one of the most crucial factors when considering the Peak Over Threshold method. From theory, a high threshold is desirable for asymptotic assumption to be satisfied, but the higher the threshold the less the observations that are left to estimate for the tails of the distribution leading to poor parameter estimation of the GPD. The choice of the threshold is chosen by the user and a good threshold value might be the lowest value for which excess threshold fits the Generalized Pareto Distribution.

We consider the methods for selecting the lowest threshold values. The first one is the explanatory technique which is carried out prior to model estimation and the second involves assessing the stability of the estimated parameters based on fitting models across a range of different thresholds.

Method 1 is based on the mean of the GPD with parameters σ and ξ.

Assuming Y has a GPD , then

$$E(Y) = \frac{\sigma}{1-\xi} = \frac{\sigma + \xi(u-\mu}{1-\xi} \tag{3.24}$$

defined for $\xi < 1$ and linear in u.

The mean residual life plot is obtained by plotting the empirical mean exceedance of the threshold against the threshold itself, The locus of those points is

$$\left\{\left(u, \frac{1}{n_u}\sum_{i=1}^{n_u}(x_i - u) : u < x_{max}\right)\right\}, \tag{3.25}$$

$x_1, \ldots, x_{n_u}$ are observed values exceeding the threshold u, $x_{\max}$ is the largest of the x_i and is termed as the mean residual life plot. The GPD provides a valid approximation to the excess distribution above a threshold u_0, and the mean residual life plot must be approximately linear in u. Confidence intervals can always be added to the plot based on the approximate normality of the sample means. The interpretation of these life residual plot can be difficult in practise, because it is approximated linear in areas where the GDP appears to have a good fit, which could be a misleading in cases where several areas are providing good fit.

Method 2 involves observing the estimated scale and shape parameters. The shape parameter is the same from GEV distribution. The scale parameter $\tilde{\sigma} = \sigma + \xi(u - \mu)$ is modified to $\tilde{\sigma} = \sigma - \xi u$ which is independent of u and must be constant over the chosen threshold. We therefore get better indication of what to choose by plotting estimates against set of threshold and observing when estimates become constant.

3.10 Average Conditional Exceedance Rates

The POT model works on the assumption of independent observations. The setting in the ACER method is that the data can be dependent and non-stationary, and it provides estimates of the exact extreme value distribution of the data when properly implemented. For thorough understanding of the ACER method, the article "Estimation of extreme values from sampled time series", [9] is used.

Let M_n be defined just as in equation (3.1), The exact distribution of M_n is given as

$$\begin{aligned} Pr(M_n \leq \eta) &= \text{Prob}(X_1 \leq \eta, ..., X_n \leq \eta) \\ &= \text{Prob}(X_n \leq \eta \mid X_{n-1} \leq \eta, ..., X_1 \leq \eta)\text{Prob}(X_1 \leq \eta, ..., X_{n-1} \leq \eta) \\ &= \prod_{j=2}^{n} \text{Prob}(X_j \leq \eta \mid X_1 \leq \eta, ..., X_{j-1} \leq \eta)\text{Prob}(X_1 \leq \eta) \end{aligned} \tag{3.26}$$

The variables X_j are statistically dependent. Assuming that all the X_j are statistically independent, as in the POT method, leads to the classical approximation:

$$\text{Prob}(M_n \leq \eta) \approx \prod_{j=1}^{n} \text{Prob}(X_j \leq \eta)$$

We will therefore use the following one-step memory approximation which will to some extent account for the dependence between them.

For $k = 1$; and $2 \leq j \leq n$, we have

$$\text{Prob}\{X_j \leq \eta \mid X_{j-1} \leq \eta, ..., X_1 \leq \eta\} \approx \text{Prob}\{X_j \leq \eta \mid X_{j-1} \leq \eta\} \tag{3.27}$$

and $3 \leq j \leq n$ we have

$$\text{Prob}\{X_j \leq \eta \mid X_1 \leq \eta, ..., X_{j-1} \leq \eta\} \approx \text{Prob}\{X_j \leq \eta \mid X_{j-2} \leq \eta, X_{j-1} \leq \eta\} \tag{3.28}$$

For general k, we obtain the expressions:

$$\begin{aligned} \text{Prob}(M_n \leq \eta) = P(\eta) &\approx \prod_{j=k}^{n} \text{Prob}(X_j \leq \eta \mid X_{j-k+1} \leq \eta, ..., X_{j-1}) \leq \eta) \\ &\quad .\text{Prob}(X_{k-1} \leq \eta \mid X_1 \leq \eta, ..., X_{k-2} \leq \eta) \\ &\quad .\text{Prob}(X_2 \leq \eta \mid X_1 \leq \eta)\text{Prob}(X_1 \leq \eta) \\ &= \prod_{j=k}^{n} (1 - \alpha_{kj}(\eta))(1 - \alpha_{k-1,k-1}(\eta)), ..., (1 - \alpha_{11}(\eta)) \end{aligned} \tag{3.29}$$

and

$$\alpha_{kj}(\eta) = \text{Prob}(X_j > \eta \mid X_{j-k} \leq \eta, ..., X_{j-k+1} \leq \eta), \text{for} \quad 2 \leq k \leq j.$$

$\alpha_{kj}(\eta)$ denotes the exceedance probability conditional on $k-1$ previous non-exceedances. Assuming $X's$ are independent and for $k=1$, we have

$$\alpha_{1j}(\eta) = \text{Prob}(X_1 > \eta)$$

Equation (3.22) can be further expressed as

$$P(\eta) \approx \prod_{j=k}^{n}(1-\alpha_{kj}(\eta))\prod_{j=k}^{n}(1-\alpha_{kj}(\eta)) \tag{3.30}$$

Using the relation $(1+x) \approx e^x$, $P(\eta)$ is deduced to

$$P(\eta) \approx P_k(\eta) \approx exp(-\sum_{j=k}^{n}\alpha_{kj}(\eta) - \sum_{j=1}^{k-1}\alpha_{jj}(\eta), \eta \to \infty \tag{3.31}$$

It follows that

$$P(\eta) \approx P_2(\eta) = \exp(-\sum_{j=2}^{n}\alpha_{2j}(\eta) - \alpha_{11}(\eta)) \tag{3.32}$$

$$P(\eta) \approx P_3(\eta) = \exp(-\sum_{j=3}^{n}\alpha_{3j}(\eta) - \alpha_{22}(\eta)) - \alpha_{11}(\eta)) \tag{3.33}$$

$$P(\eta) \approx P_4(\eta) = \exp(-\sum_{j=4}^{n}\alpha_{4j}(\eta) - \alpha_{33}(\eta)) - \alpha_{22}(\eta)) - \alpha_{11}(\eta)) \tag{3.34}$$

where α_{2j}, α_{3j}, and α_{4j} defined respectively as:

$$\alpha_{2j} = P(X_j > \eta \mid X_{j-1} \leq \eta) \tag{3.35}$$

$$\alpha_{3j} = P(X_j > \eta \mid X_{j-1} \leq \eta, X_{j-2} \leq \eta) \tag{3.36}$$

$$\alpha_{4j} = P(X_j > \eta \mid X_{j-1} \leq \eta, X_{j-2} \leq \eta, X_{j-3} \leq \eta) \tag{3.37}$$

The modified ACER function can be applied when we have non-stationary time series data. It is seen that:

$$P_k = \exp(-\sum_{j=k}^{N}\alpha_{kj}(\eta)), \qquad k \geq 1$$

$$P_k(\eta) \approx (-(n-k+1)\varepsilon_k(\eta) \tag{3.38}$$

(3.39)

The ACER function is then deduced from the approximation as

$$\varepsilon_k(\eta) = \frac{1}{n-k-1}\sum_{j=k}^{n}\alpha_{kj}(\eta), \qquad k \geq 1 \tag{3.40}$$

where n is the number of data points. $\sum_{j=k}^{n}\alpha_{kj}(\eta)$ is the expected number of clumps of exceedances above η conditioned on $k-1$ previous non-exceedances.

3.10.1 Empirical Estimation of ACER

In the empirical estimation of the ACER function, we count the total number of exceedances, together with the required non-exceedance number for the total time series data and finally dividing by $n-k+1$.

Consider the random functions below.

$$A_{kj}(\eta) = I\{X_j > \eta, X_{j-1} \leq \eta, ..., X_{j-k+1} \leq \eta\}, j = k, ..., n, k \geq 2 \tag{3.41}$$

$$B_{kj}(\eta) = I\{X_{j-1} \leq \eta, ..., X_{j-k+1} \leq \eta\}, j = k, ..., n, k \geq 2 \tag{3.42}$$

where $I\{A\}$ denotes the indicator function of some A. Then

$$\alpha_{kj}(\eta) = \frac{E[A_{kj}(\eta)]}{E[B_{kj}(\eta)]} \tag{3.43}$$

where $E[.]$ denotes the expectation operator. Assuming an ergodic process, then obviously $\varepsilon_K(\eta) = \alpha_{kk}(\eta) = ... = \alpha_{kn}(\eta)$. It may be assumed that

$$\varepsilon_k(\eta) = \lim_{n\to\infty} \frac{\sum_{j=k}^{n} A_{kj}(\eta)}{\sum_{j=k}^{n} B_{kj}(\eta)} \tag{3.44}$$

obviously, $\lim_{\eta\to\infty} E[B_{kj}(\eta)] = 1$.

We have that

$$\lim_{\eta\to\infty} \tilde{\varepsilon}_k(\eta)/\varepsilon_k(\eta) = 1 \tag{3.45}$$

where

$$\tilde{\varepsilon}(\eta) = \frac{\sum_{j=k}^{n} E[A_{kj}(\eta)]}{n-k+1} \tag{3.46}$$

$\tilde{\varepsilon}(\eta)$ is the modified ACER function for $k \geq 2$. The modified function can be used when the underlying time series is non-stationary or for long term statistics. The sample estimate of $\tilde{\varepsilon}(\eta)$ is

$$\hat{\varepsilon}_k(\eta) = \frac{1}{R}\sum_{r=1}^{R} \hat{\varepsilon}_k^{(r)}(\eta), \tag{3.47}$$

where R is the number of realizations and

$$\hat{\varepsilon}_k^{(r)}(\eta) = \lim_{n\to\infty} \frac{\sum_{j=k}^{n} A_{kj}^{(r)}(\eta)}{n-k+1} \tag{3.48}$$

3.10.2 Confidence Interval Estimation

The sample standard deviation is estimated as defined

$$\hat{s}_k(\eta)^2 = \frac{1}{R-1}\sum_{r=1}^{R}\left(\hat{\varepsilon}_k^{(r)}(\eta) - \hat{\varepsilon}_k(\eta)\right) \tag{3.49}$$

For independent realizations, the estimated standard deviation gives a good approximation to the 95% confidence interval CI with

$$\mathrm{CI}^{\pm}(\eta) = \hat{\varepsilon}_k(\eta) \pm 1.96\hat{s}_k(\eta)/\sqrt{R} \tag{3.50}$$

3.10.3 Fitting Asymptotic Distribution for the General Case

The Weibull, Gumbel, and Fréchet asymptotic distributions are special cases of the Generalized Extreme Value distribution as outlined in section (3.3). The Weibull distribution is for maxima, more of an upper bond limit and is irrelevant to the data we are considering.

Originally, the ACER programme is targeted at asymptotic Gumbel extremes. As a result, it gives bad fit when the underlying extreme value distribution differs from Gumbel. Elspot prices seemed to have a heavy-tailed distribution, and for some of these models we need more of a Fréchet type distribution. Fitting the empirical ACER function to the asymptotic Fréchet distribution, we can use the parametric form in making predictions.

For independent data, the first ACER function can be expressed asymptotically as:

$$\varepsilon_1(\eta) \approx [1+\xi(a(\eta-b))]^{-\frac{1}{\xi}} \tag{3.51}$$

where $a > 0$ and b, ξ are constants. The parameters a and b corresponds to $\frac{1}{\sigma}$ and u in the GEV expression respectively. If we assume that

$$\varepsilon_k(\eta) = q_k(\eta)[1+\xi_k(a_k(\eta-b_k)^{c_k})]^{-\frac{1}{\xi_k}}, \eta \geq \eta_1 \tag{3.52}$$

where $q_k(\eta)$ is weakly varying when compared with $[1+\xi_k(a_k(\eta-b_k)^{c_k})]^{-\frac{1}{\xi_k}}$, and a_k, b_k, c_k and $\xi_k > 0$ are k dependent constants. When $c_k = 1$ and $q_k(\eta) = 1$, we obtain the asymptotic limit of the expression. Some adjustment will be made about the above definition of $q_k(\eta)$. Since $q_k(\eta)$ is a weakly varying, we can assume it has a slowly varying tail sufficient enough to be replaced by constant for $\eta > \eta_1$, $q_k(\eta)$
For simplicity we write

$$\varepsilon(\eta) \approx [1+\xi(a(\eta-b))]^{-\frac{1}{\xi}} \tag{3.53}$$

To estimate the parameters of the ACER function by optimization, we minimized the mean square error function which is defined as

$$F(\tilde{a}, b, , c, q, \gamma) = \sum_{j=1}^{n} w_j \mid \log \hat{\varepsilon}(\eta_j) - \log q + \gamma \log[1 + \tilde{a}(\eta_j - b)^c] \mid^2 \tag{3.54}$$

$$= \sum_{i=1}^{n} w_i (y_i - \log q - \xi x_i)^2 \tag{3.55}$$

where w_j is the weight factor that emphasis less extreme data points and defined as:

$$w_i = (\log \mathrm{CI}^+(\eta_i) - \log \mathrm{CI}^-(\eta_i))^{-2}$$

$\mathrm{CI}^+(\eta_i)$ is the upper bond and $\mathrm{CI}^-(\eta_i)$ is the lower bond of the 95% confidence intervals at η_i x_i and y_i defined respectively as $\log(1 - \tilde{a}(n_i - b)^c$ and $\log \hat{\varepsilon}(\eta_i)$.

The Levenberg-Marquardt least square optimization method is employed to estimate the five parameters: $\tilde{a}, b, c, q, \gamma$. To make the optimization easier, the parameters $\tilde{a}, b, c$ are fixed reducing the optimization problem to a weighted linear regression of the standard form. The optimal values of γ and $\log q$ then estimated respectively with expressions:

$$\gamma^*(\tilde{a}, b, c) = -\frac{\sum_{j=1}^{n} w_j (x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^{n} (x_j - \bar{x})^2} \tag{3.56}$$

and

$$\log q^*(\tilde{a}, b, c) = \tilde{y} + \gamma^*(\tilde{a}, b, c)\bar{x} \tag{3.57}$$

where $y_j = \log \hat{\varepsilon}(\eta_j)$ and $x = 1 + \tilde{a}(n_j - b)^c$ The optimal parameters of $\tilde{a}^*$, b^* and c^* is found by applying Levenberg-Marquardt on the function $\tilde{F}(\tilde{a}, b, c) = F(\tilde{a}, b, c, q^*(\tilde{a}, b, c), \gamma^*(\tilde{a}, b, c))$.

3.11 AR-GARCH Model

As previously mentioned, The Byström price modeling innovation will be implemented where combined time series AR-GARCH with Extreme Value Theory was used. AR(k)-GARCH(p, q) is an autoregressive model of order k with GARCH noise of order p, q and can be generally expressed as:

$$r_t = a_0 + \sum_{i=1}^{k} a_i r_{t-i} + \varepsilon_t \tag{3.58}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_j \sigma_{t-i}^2 \tag{3.59}$$

where $\varepsilon_t = \sigma_t \eta_t$ with $\eta_t \sim iid(0,1)$. η_t could either be normally distributed scaled to unit variance or student t-distributed with ν degrees of freedom. The AR is stationary when$\mid \sum_{i=1}^{k} a_i < 1 \mid$ and GARCH(p, q) is strictly stationary with finite variance if $a_0 > 0$ and $\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_j < 1$ The extreme observations will be filtered by the AR-GARCH model.

The electricity spot market exhibits strong seasonality pattern and significant clustering of volatility, hence the use of AR-GARCH model is critical. An AR(5) have been used to model the weekly seasonality since we have 5 observations each week. In order to capture the monthly as well as quarterly seasonality, AR(20) and AR(60) terms are included in our model. For simplicity $p,q = 1$ is sufficient in our case. GARCH(1,1) models conditional volatilities as function of past conditional volatilities. (The present volatility been dependent only on the first).

Our method is therefore a basic AR[1,5,20,60]-GARCH(1,1) model and would be represented as shown:

$$r_t = a_0 + a_1 r_{t-5} + a_2 r_{t-20} + a_3 r_{t-60} + \varepsilon_t. \tag{3.60}$$

$$\sigma_t^2 = \phi_0 + \phi_1 \varepsilon_{t-1}^2 + \phi_2 \sigma_{t-1}^2 \tag{3.61}$$

where σ_t^2 is the conditional variance of ε_t.

4 Data

Data used in this thesis is the elspot daily system price data from January 1, 1996 to March 19, 2013. This data set is obtained from Prof Sjur Westgaard of the Economics department, NTNU. It consist of 4352 electricity price observations, [quoted Norwegian kroner (NOK/MW h)] which represent data collected over 17 years. 5 observations were recorded every week and there were days no price data was recorded.

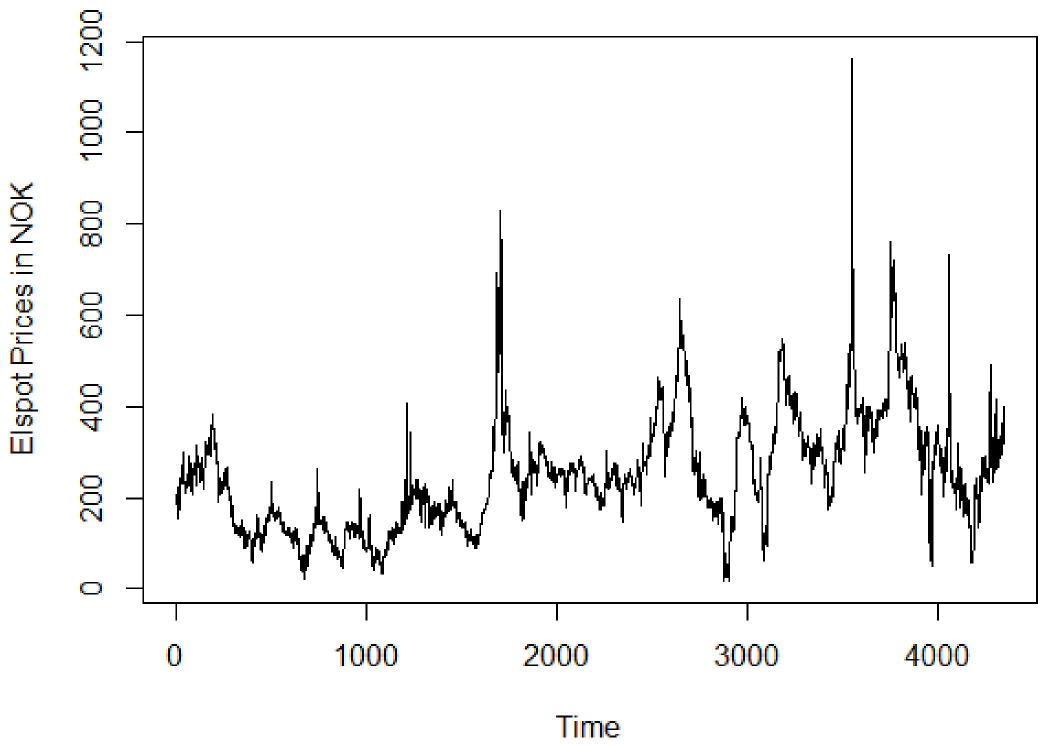


Figure 4.1: *Daily electricity prices from January 1, 1996 to March 19, 2013*

We observe from Figure 4.1 that, prices could be very extreme attaining as high as over 1164 NOK/MW h and as 1ow as 17 NOK/MW h. Also our time series tend to show no specific increasing or decreasing pattern. We therefore might suspect a stationary one but conclusion will only be made after a valid stationarity test have been performed.

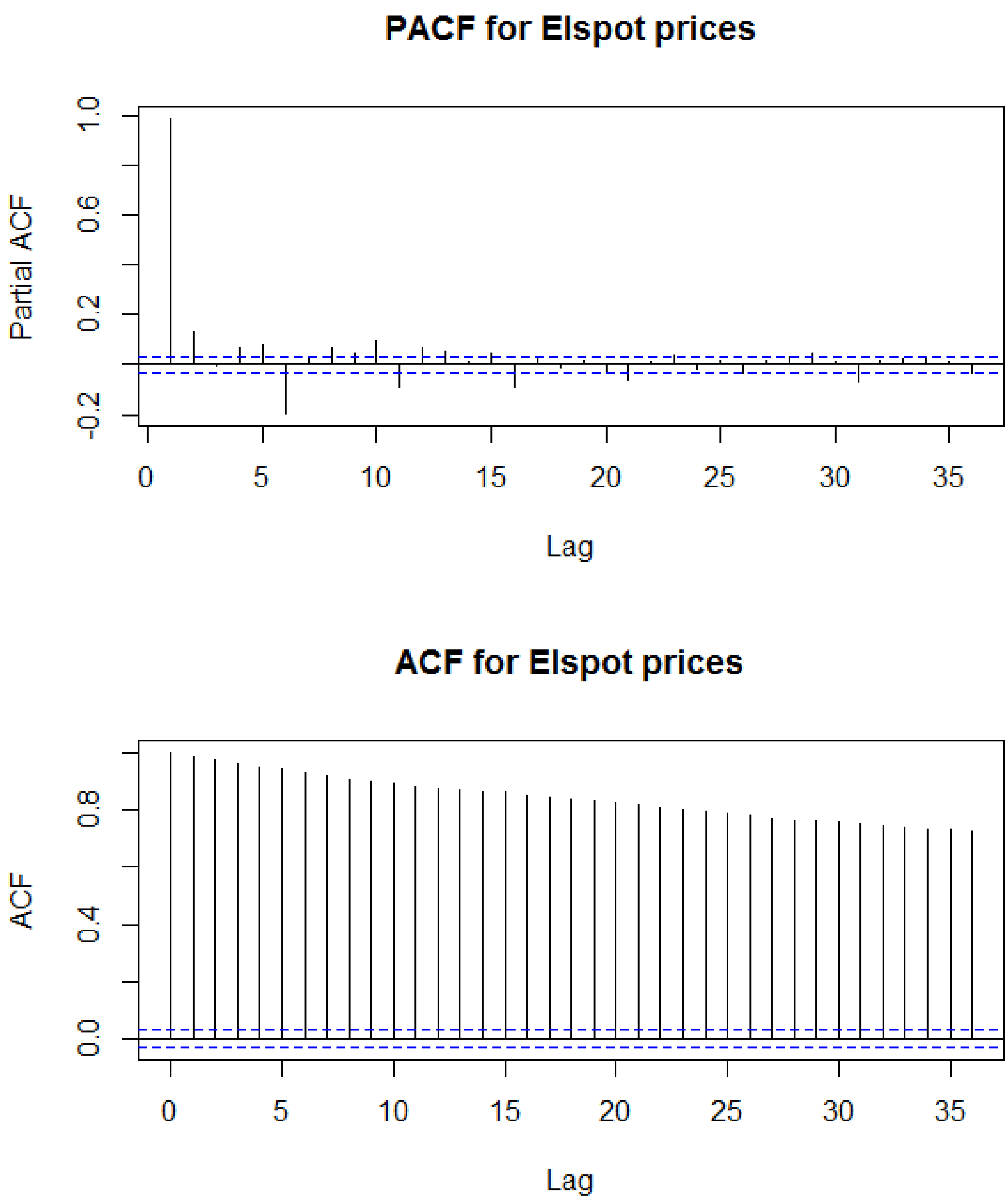


Figure 4.2: *PACF and ACF of the electricity prices*

The appropriate choice of time series is determined with the ACF. It shows the lag from one observation to the other. A geometrically decaying ACF implies that an AR model is appropriate or combination of AR and MA. We also need to remember that in most cases the best model is the one that uses either only AR terms or MA terms. One that shows no regular pattern could just be random noise.

The PACF is used to determine the order of our AR model by examining the line that falls outside the confidence bounds and counting how many lags it takes for the data to fall outside the dotted lines. In our case it appears it takes only one significant lag. Moreover there is a slowly decaying, but strongly significant ACF. Prices tend to show strong dependence, hence the need of differencing is critical and have been carried out through returns.

4.1 Returns

Returns ensure that prices are measured in comparable metric, and mainly employed to ensure normalization in prices. Thus bringing out the analytical relationship between prices though returns are originally from prices series of different values. Our data have be analysed with log returns and simple net returns.

4.1.1 Log Returns

Assuming prices(returns) are distributed log normally , then $\log(1+r_t)$ is normally distributed, since: $(1+r_t) = \frac{P_t}{P_{t-1}}$. For very small returns, $\log(1+r) \ll 1$, we have the logarithmic return

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

.

From Figures 4.5 and 4.6, It appears there is no significant difference among ACF and PACF plot. The PACFs show an alternating pattern with spikes at lags 1,4,5,10 and 15. The lag axis of the ACFs starts from 0. The 0 lag means that the returns does not appear to have greater correlation between the observations, with quite significant spikes at few lags. Thus there are five significant lags outside the confidence limit of the PACF and a model with an AR(5) might be reasonable. In fact according to our ACF and PACF of the return series, the data looks random and shows no discernible pattern. But this should not rule the fact that our data fits an AR(1) with weak autocorrelation. We therefore add the seasonal pattern and arrive at the model used in this thesis, atleast for the conditional mean as expressed in equation (3.55).

Our preferred choice of residual series are those close to being an iid than the actual prices. With log and simple net residual series, we can observe from their time series plot that, it's essential to use log returns when we are concerned with price drops. Most importantly, the log-returns is log-normally distributed and seems quite illogical when combined with very fat-tailed distributional assumptions though might be tempting since fat-tails have additional kurtosis (table 4.1) and provides better risk estimates. Since our interest is to examine extreme return when there are high price increase (extremely high risk) with heavy tailed distribution, the simple net return is a more preferred choice but its worth saying that there is no significant difference between the two returns.

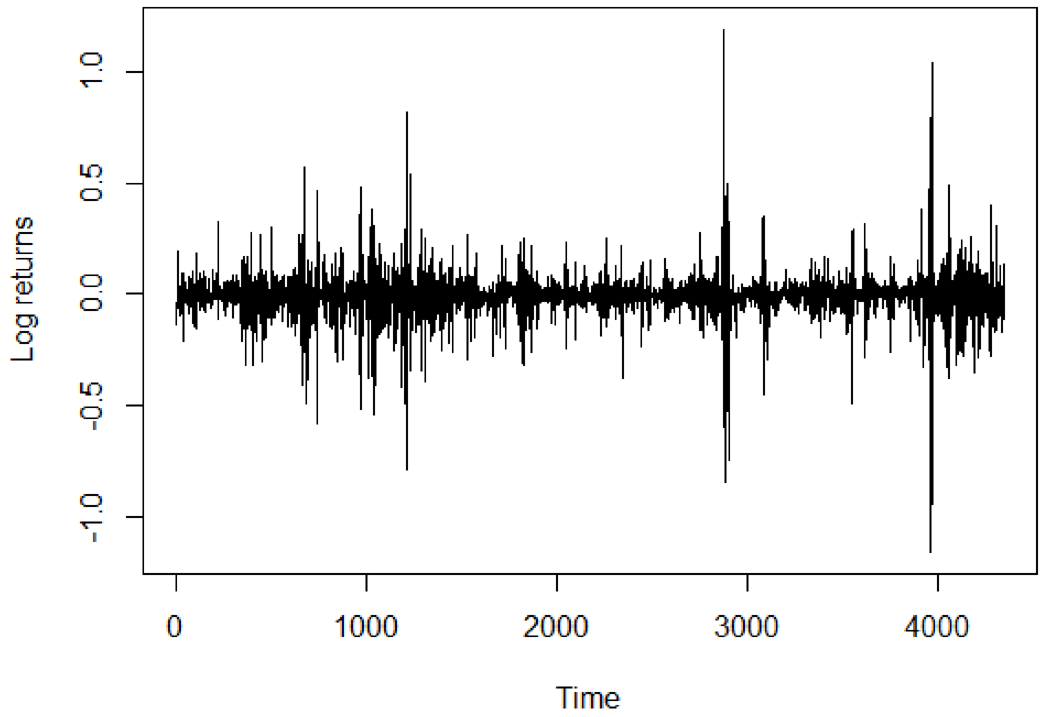


Figure 4.3: *Daily observation of log return from Nord Pool January 1, 1996 to March 19, 2013*

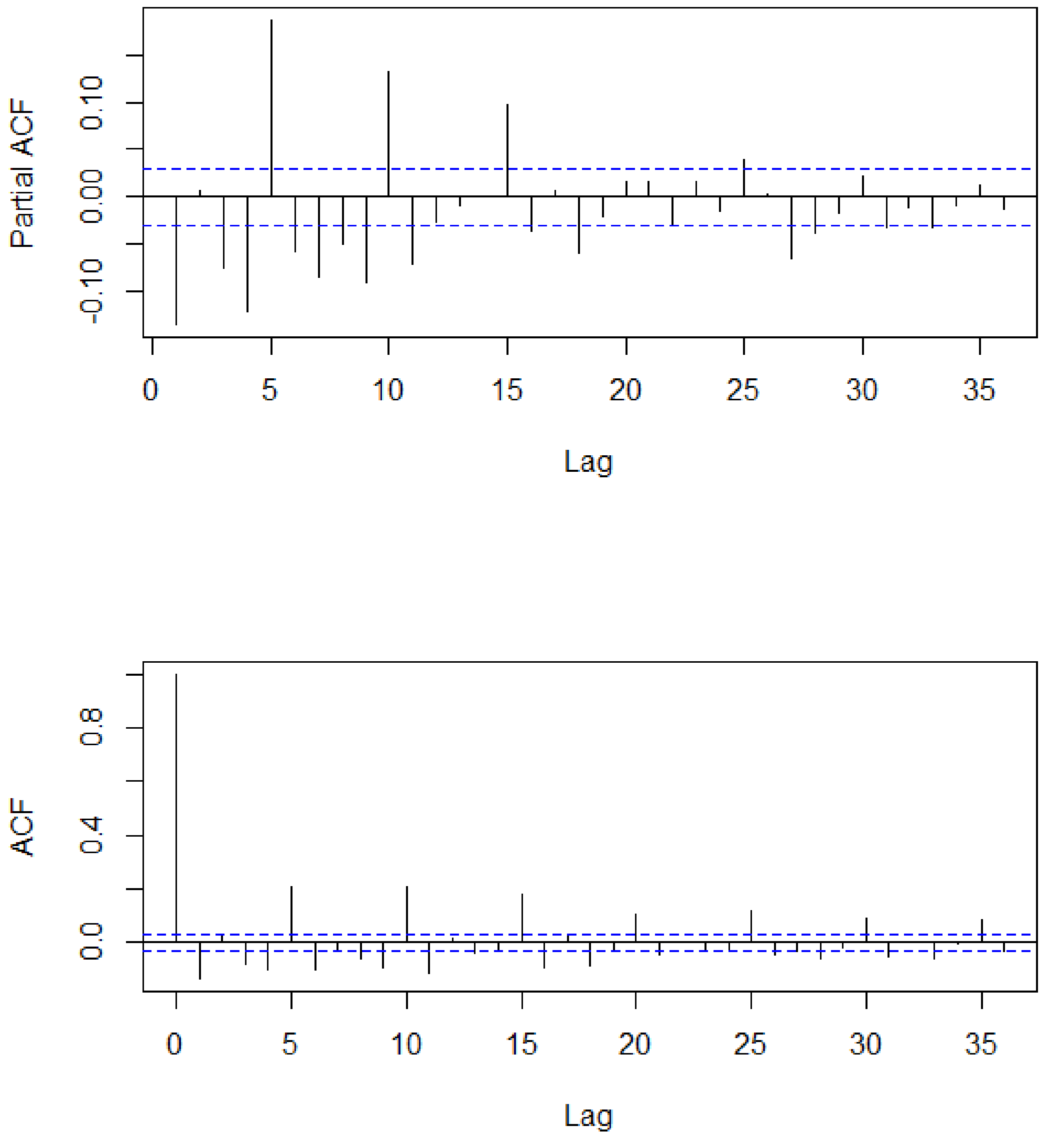


Figure 4.4: *PACF and ACF of the daily log returns*

.

4.1.2 Simple Net Returns

The simple net return is given as:

$$\frac{P_t - P_{t-1}}{P_{t-1}} \tag{4.1}$$

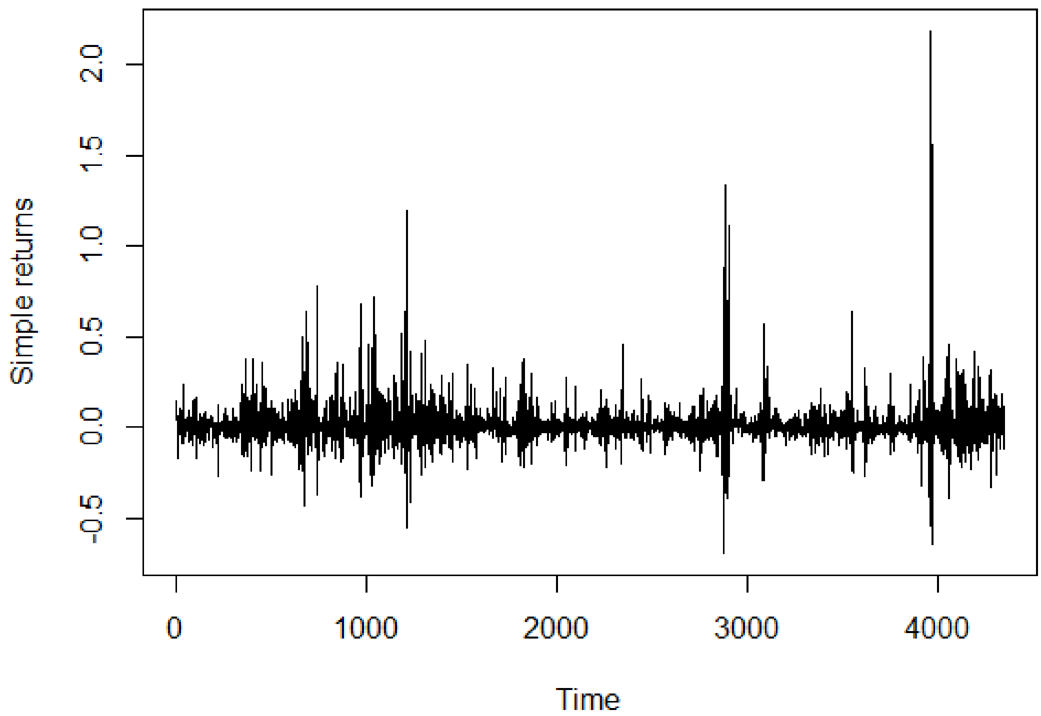
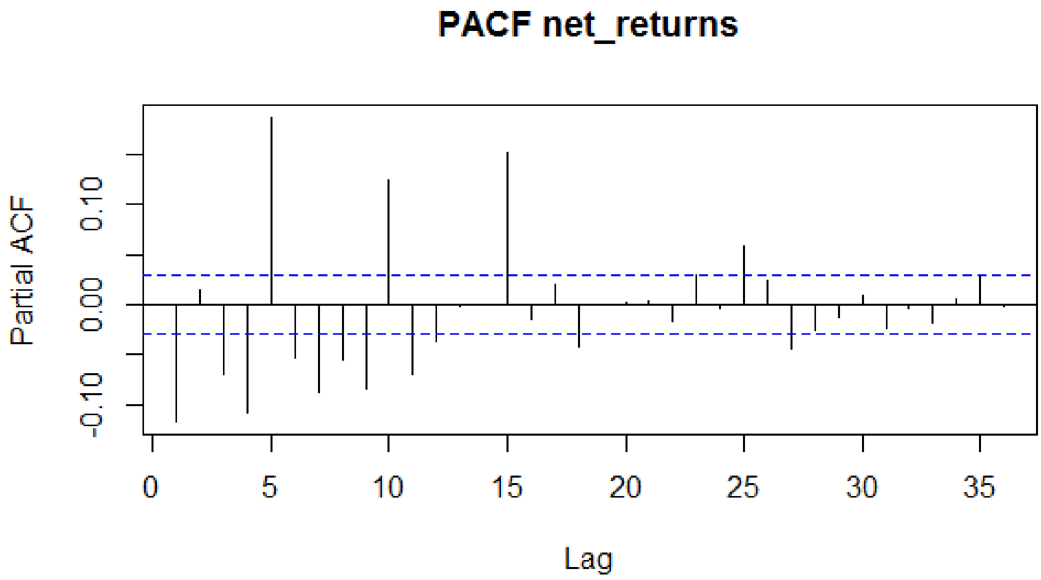


Figure 4.5: *Simple net returns plot of elspot January 1, 1996 to March 19, 2013*

.

This time series plot of the simple net returns tends to show higher extreme spikes when compared to the one with logarithmic returns .

.



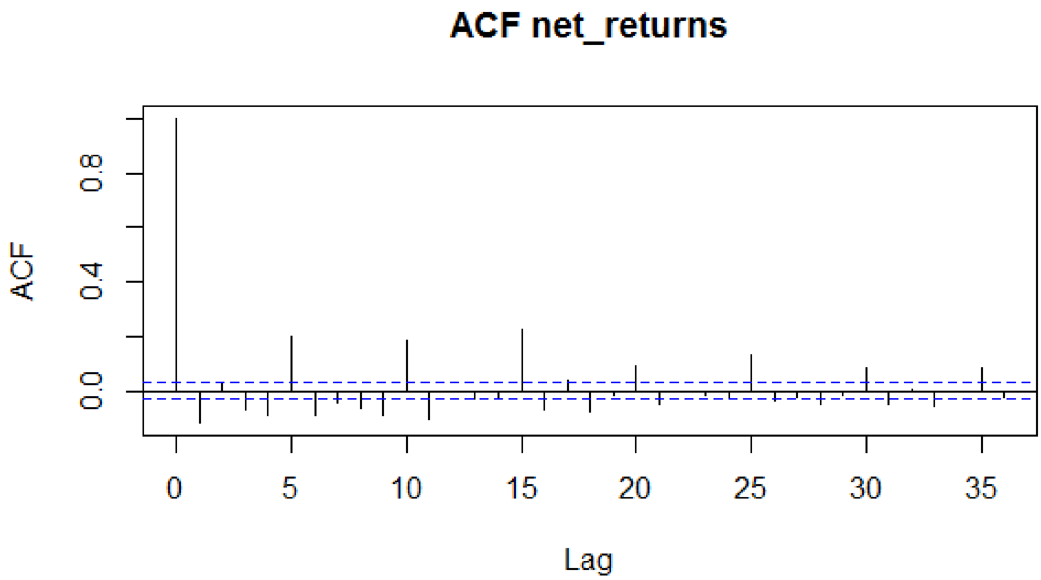


Figure 4.6: *PACF and ACF of the simple net returns*

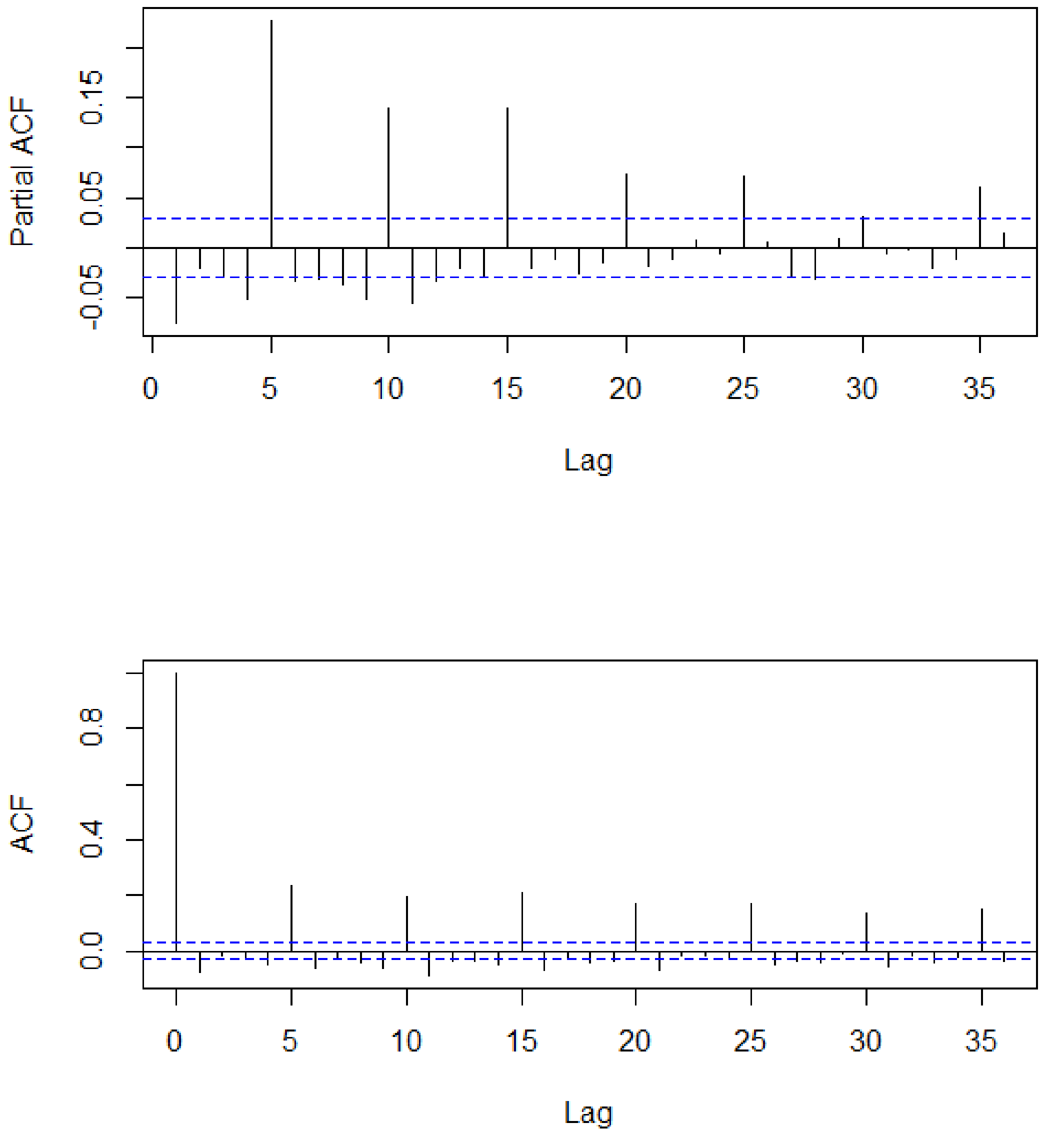


Figure 4.7: *PACF and ACF of the standardized residuals of AR-GARCH model of simple net returns*

Again, there happens to be some level of time dependence in the residuals series. Much of the serial autocorrelation have been removed by the AR-GARCH filter as compared to figure 4.6. In the POT analysis, we will assume that the residuals series are close as possible to being an iid and therefore our data would not be declustered as in [3].

4.2 Quantile-Quantile (Q-Q) Plot

A quantile-quantile is a graphical technique to ascertain if two data sets are derived from a population with a common distribution. Our definition of quantile simply means the point below which a given percentage of data lies.

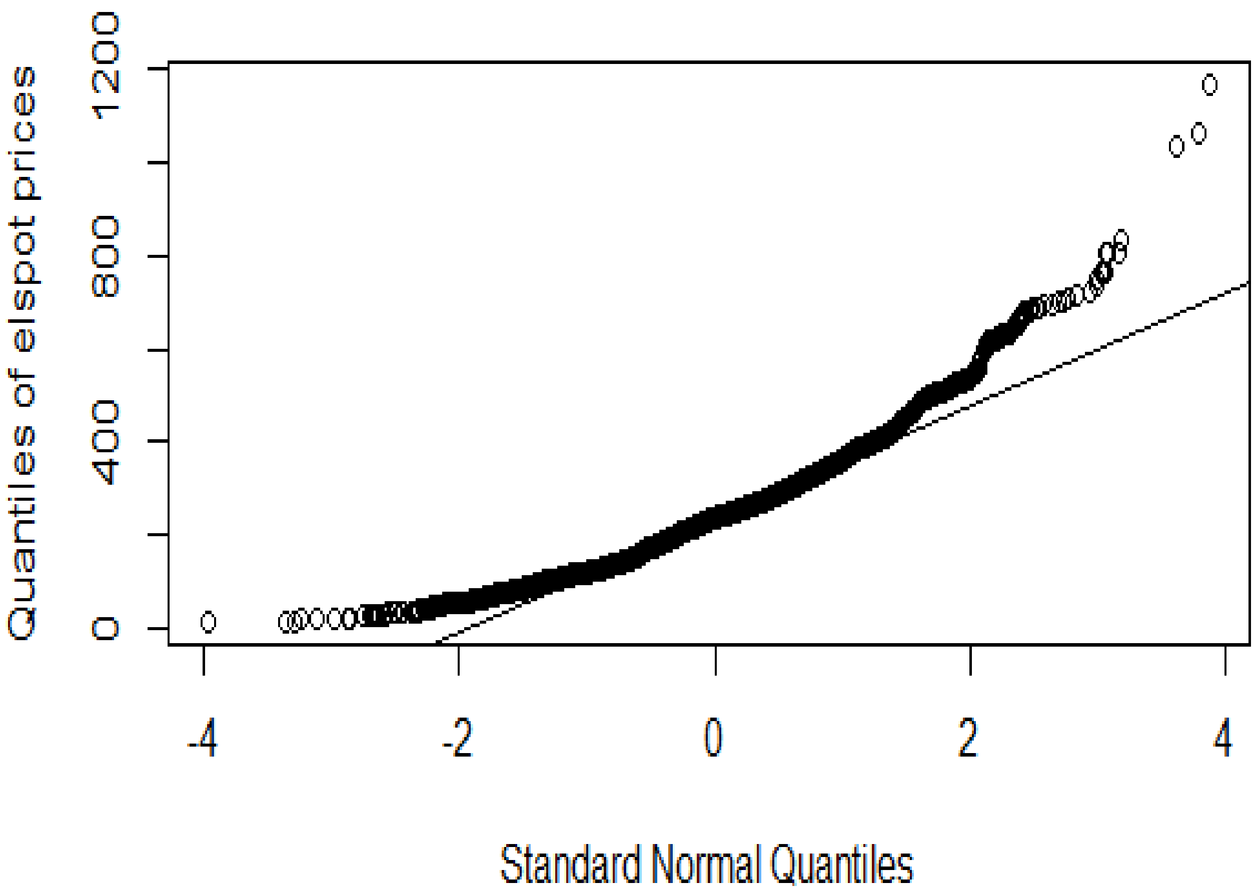


Figure 4.8: *Q-Q Plot of electricity prices*

To compare our empirical distribution at hand to the normal distribution, we use the Q-Q plot. The Q-Q plot of daily electricity prices over the 17 years period is shown in Figure 4.8, the electricity prices shows a concave curve with the normal line. From R output, the skewness and kurtosis of the price data is found to be 1.0286 and 5.227 respectively.

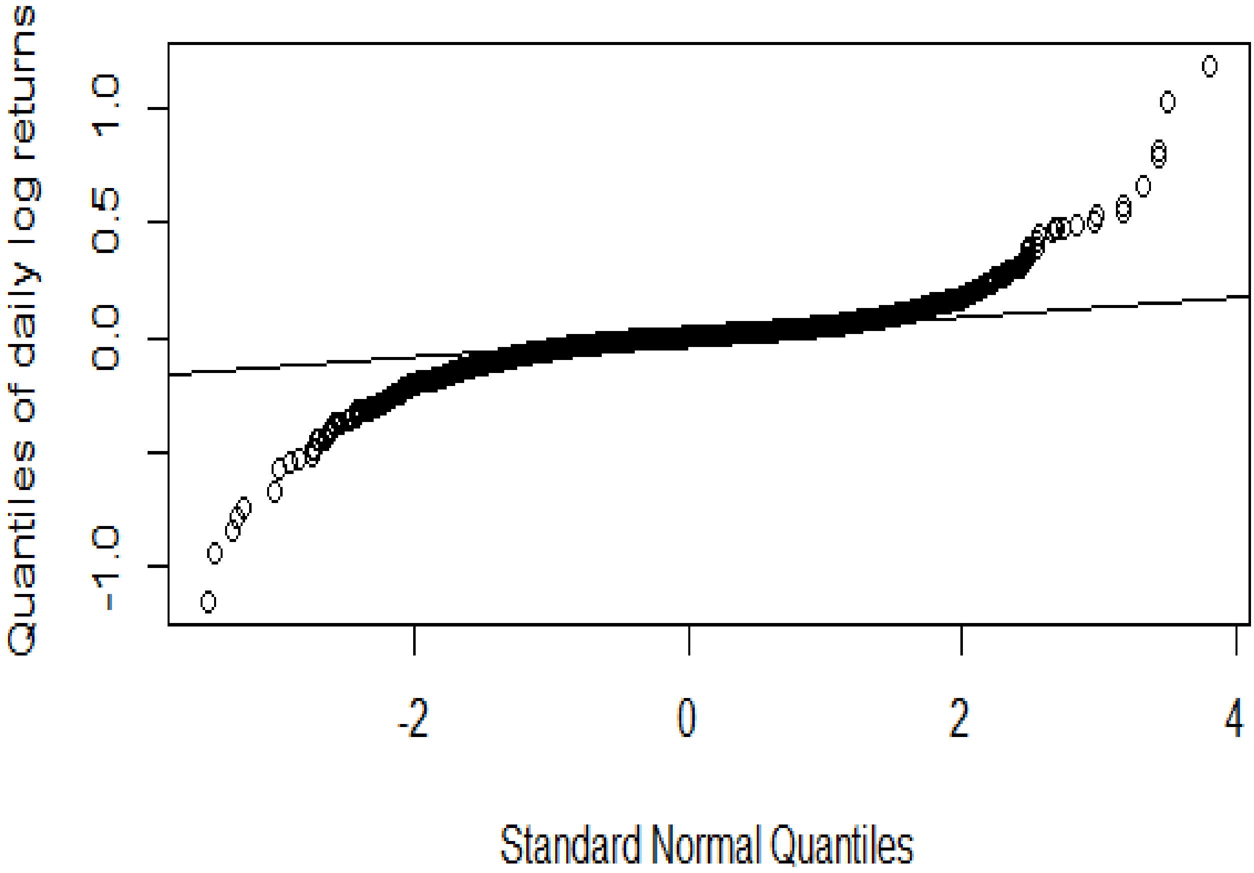


Figure 4.9: *Q-Q Plot of daily logarithmic returns*

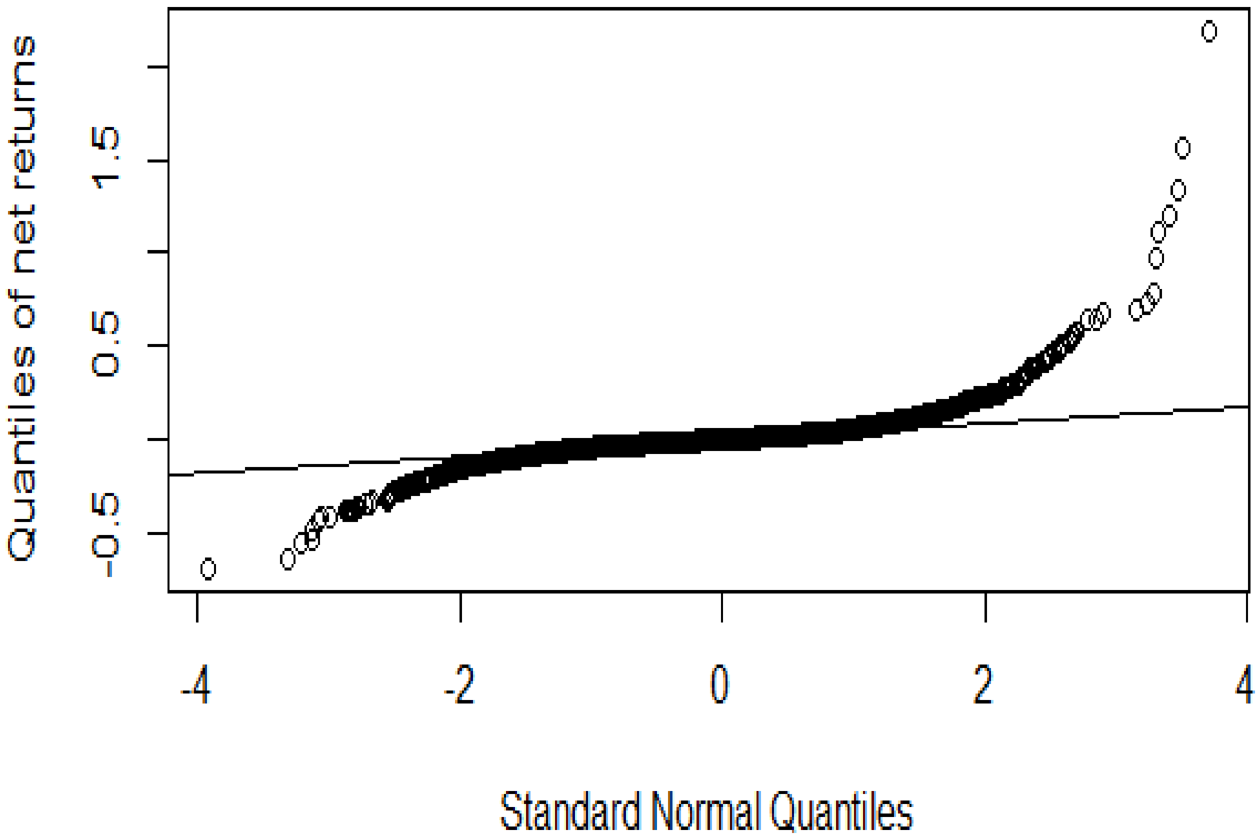


Figure 4.10: *Q-Q Plot of net returns with normal distribution as the line*

From the above, the daily net returns tends to behave significantly different from the standard normal distribution.

4.3 Normality Test

If an empirical distribution coincides with a normal distribution, the Q-Q plot gives a straight line. The S-shape in Figure 4.10 and Figure 4.9 is clearly far from been straight line and we might suspect a fatter tail distribution skewed to the right. Kurtosis of 75.22 and positive skewness of 4.76 support our thought. Sometimes it could be misleading to make conclusions from plots of data set without a valid statistical test. We therefore employ the Kolmogorov-Smirnov and Shapiro-Wik tests to compare the empirical data to the normal distribution. The test result shown in Table 4.1

	Statistic	P-value
Kolmogorov-Smirnov	0.4084	2.2e-16
Shapiro-wik	0.6863	2.2e-16

Table 4.1: *Normality test of daily electricity prices*

The null hypothesis: The two distributions shows no significant difference, is tested against the alternative hypothesis that the two distributions were drawn from the same distribution. If the p-value is smaller than the significance level, 5% in this case, we reject the null hypothesis. From the test output shown above, we can confidently confirm that our price (returns) is not normally distributed and that could be observed from the probability plots as well.

4.4 Test of Stationarity

The table below presents the stationarity test of daily price observations. Under the null hypothesis of data been hardly stationary, is tested against the alternative hypothesis that our data is from a stationary process, at 5% level of significance. The null hypothesis is rejected by both the Dickey-Fuller and Phillips-Perron tests and we can confirm that prices are from stationary process. We must however notice that result of these tests cannot always be trusted. An alternative to stationarity testing is by observing the plotted time series for upward and downward trends. Stationarity of prices is an important tool when considering the listed extreme value approaches. Whilst it's the means of obtaining consistent estimators in POT, ACER requires the stationarity specification of data. For both normality and stationariy test that have been carried out, the daily electricity price observation have been used. However, repeating the tests with simple net returns yield the same results.

	Statistic	P-value
Dickey-Fuller	0.4084	0.01
Phillips-Perron	-6.1196	0.01

Table 4.2: *Stationarity test of daily electricity prices*

5 Analysis of Data

In this chapter, the data presented in chapter 4 have been analysed. With exception of the ACER program which have been implemented in MATLAB, we have used R programming for rest of the analysis. Analysis have been done mostly with the simple net returns. We have used the Rugarch package in R for the AR-GARCH modeling, POT package for modeling GPD and ACER implementation done with help from [7].

5.1 Generalized Pareto Distribution Modeling

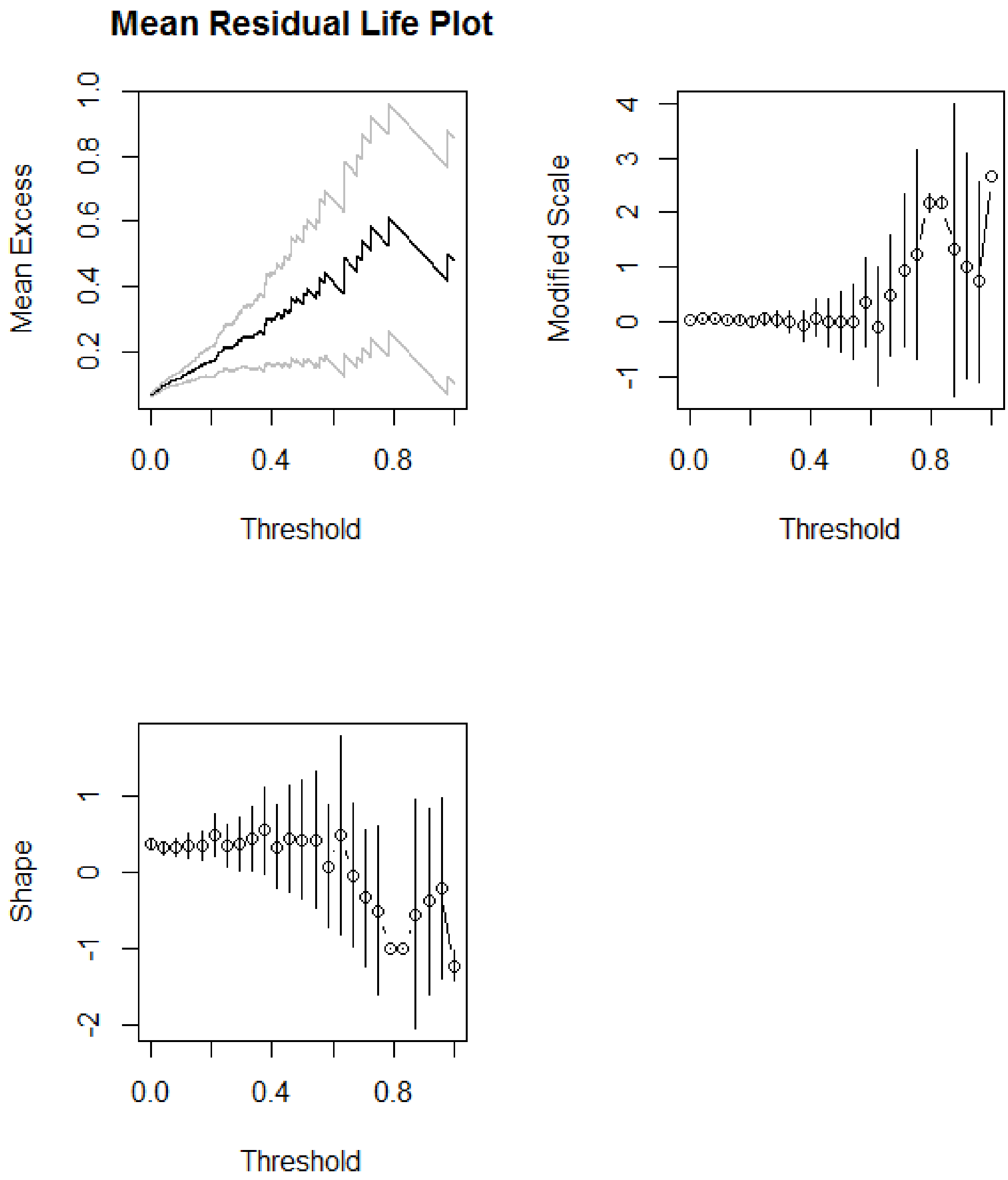


Figure 5.1: *Tools for threshold selection*

Figure 5.1 shows the mean residual life plot, together with the modified shape and scale parameters against the thresholds. The mean residual life plot tends to be linear around areas with reasonable choice of threshold and the threshold choice plots comprising the modified shape and scale estimates tend to be constant around (0.5,0.6). According to these figures, a threshold of 0.5 might be a good choice. For threshold greater than 0.5, both the modified scale and shape estimates appears to be constant only for a little while. The modified scale and shape parameters tends to be linear and after threshold of 0.5, they exhibit no linearity. A total of 20 observations were above this threshold value representing 0.46% of the total observations.
From equation (3.60) and (3.61), assuming Quasi Maximum Likelihood (QMLE) for the daily net returns, we get the estimated parameters for our model both with the AR and GARCH parts. The QMLE uses the normal distribution assumption and robust standard errors for inference.

	Normal	Student's t
AR-GARCH parameters		
a_0x10^{-3}	4.947 (1.841)	-4.973 (0.967)
a_1	0.100 (0.01485)	-0.0134 (0.01217)
a_2	-0.1794 (0.04263)	0.1703 (0.011663)
a_3	0.04647 (0.01529)	0.07978 (0.0096)
a_4	0.03358 (0.0149)	0.02348 (0.0783)
ϕ_0x10^{-5}	8.80 (1.495)	0.4698 (0.1052)
ϕ_1	0.1949 (0.01184)	0.50639 (0.09379)
ϕ_2	0.8394 (0.007594)	0.70439 (0.02488)
v		2.53 (0.1234)
Standardized residual statistics		
Mean	-0.0339	0.0561
Standard deviation	0.9993	0.85
Skewness	1.94	2.73
Excess kurtosis	19.13	32.40
Q(10)	488.3151	491.582
Q(20)	900.652	940.9171
$Q^2(10)$	6.1398	3.9211
$Q^2(20)$	23.5682	34.4983
GPD parameters		
ξ	0.164 (0.03876)	
α	0.718 (0.03689)	
u	0.5	

Table 5.1: *AR-GARCH parameters, statistics based on standardized residuals, as well as GPD parameters of our net returns*

Table 5.1 reports the estimated AR-GARCH parameters by the Normal and student t-distribution, the standardized residual statistics as well as GPD parameters. The Ljung-Box Q(10) and Q(20) statistics test autocorrelation for r_t and r_t^2 reported. The test was applied to the residual series and the result indicates there are still some level of autocorrelation left in the data though most of the heteroscedasticity has been removed. This is not surprise considering the high-frequency data we are working with. Standard errors are recorded in parenthesis. The GPD parameters are estimated, using maximum-likelihood estimator, by fitting GPD to the standardized residuals based on the normal distribution innovation. It should be noted that, the sum of the GARCH parameters slightly exceeds 1 following a large positive tail-index value. Also from theory, $(\xi > 0)$ reflects fat-tailed distribution, particularly $(\xi = 0.164)$ was significantly different from 0 at 1% level which suggests that the right tail of the standardized residuals follows a Fréchet distribution. The high excess kurtosis clearly depicts the non-normality of our prices as verified also with the normality test.

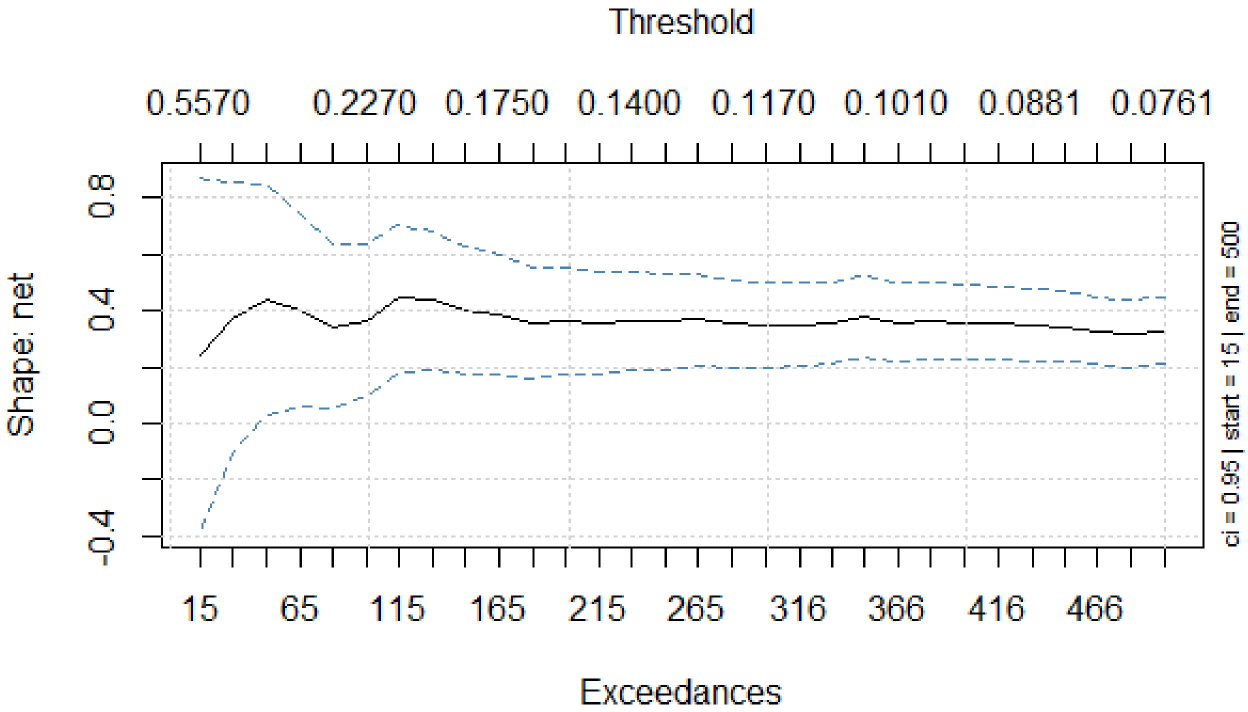


Figure 5.2: *GPD shape plot of the daily net returns*

Fig. 5.2 offers a visual display of how the estimate of the shape varies with the number of extremes or threshold. It could be noticed that, the threshold decreases with increasing number of exceedances.

Probability	Expected	AR-GARCH	AR-GARCH-t	Conditional GPD	ACER
.95	218	210	84	211	236
.99	44	111	9	48	42
.995	22	85	4	13	24
.999	4	58	1	7	5
.9995	2	48	9	3	2
.9999	1	30	0	1	1

Table 5.2: *Estimated tail quantiles at different probabilities(number of exceedances)*

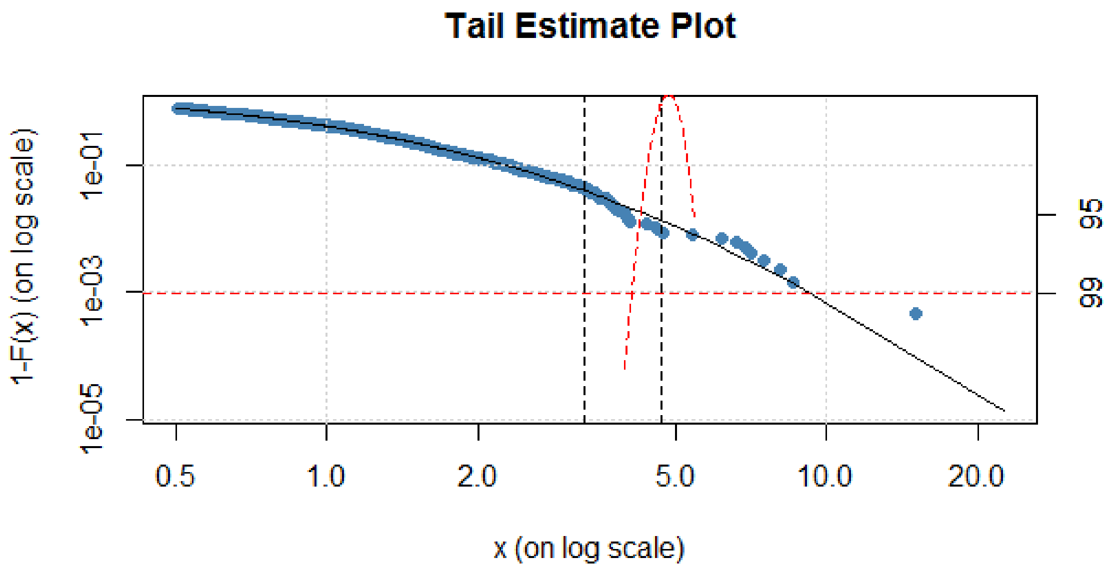


Figure 5.3: *Tail estimate plot of the GPD at threshold of 0.5*

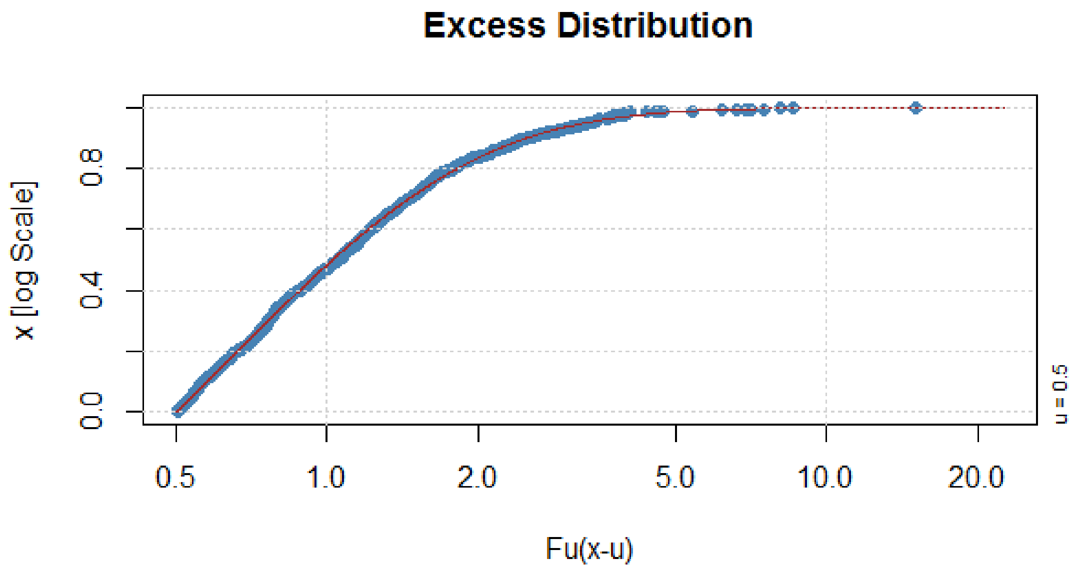
Probability	VaR estimate	C.I(VaR estimate)	ES estimate	C.I(ES estimate)
.95	1.6043	[1.526 1.698]	2.6796	[2.560 2.933]
.99	3.2601	[3.023 3.556]	4.6598	[4.161 5.409]
.995	4.1192	[3.777 4.615]	5.6871	[4.790 6.809]
.999	6.5336	[5.640 7.905]	8.5744	[7.0423 11.189]
.9995	7.7853	[6.546 9.760]	10.0724	[8.057 13.653]
.9999	11.3067	[8.893 15.545]	14.2823	[10.651 21.361]

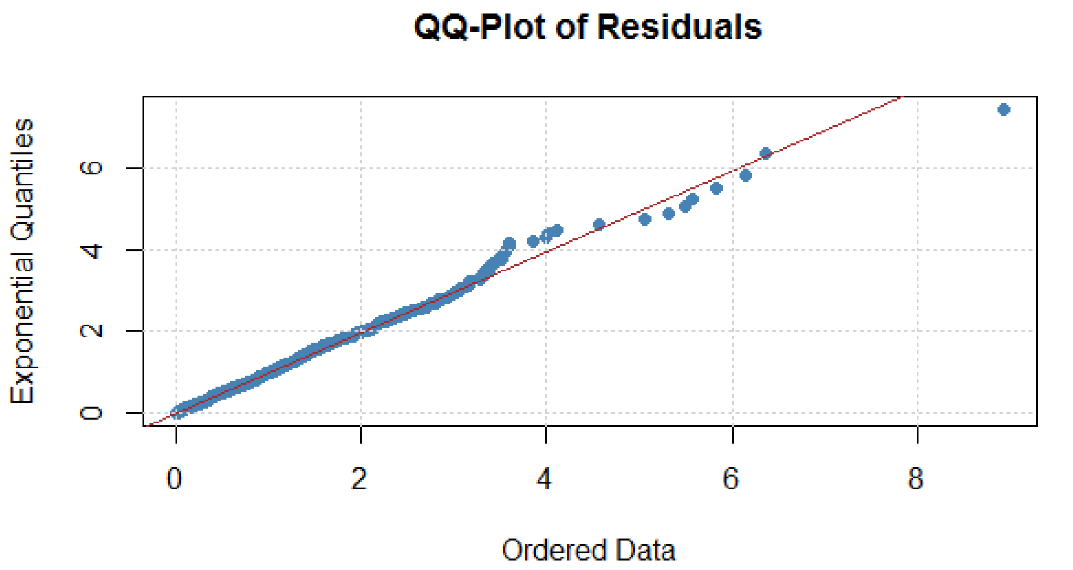
Table 5.3: *Tail Risk estimate at different probabilities based on the conditional GPD*

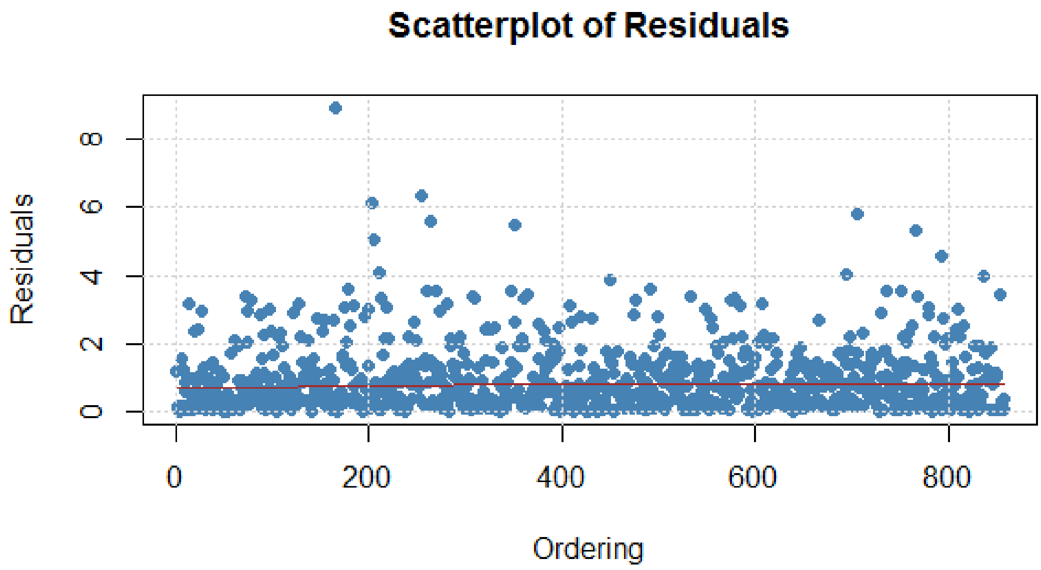
From Table 5.2, The EVT based conditional GPD model is compared to traditional method.

That is, we show how the traditional time series models and the POT could be used in describing the tail quantiles. The numbers represent empirically observed exceedance numbers and they should be close to the expected. If any of the distributions succeed in describing the marginal tail quantiles, then its empirically observed number of exceedances should be closed to the expected. At the 95% level of the normal AR-GARCH, the empirical number of exceedance was close to the expected but was far way off at extreme tails. However, the AR-GARCH-*t* seriously perform worse at more extreme tails and gets even worst at the upper tail quantiles (the first three quantiles). The results with the AR-GARCH-*t* was quite surprise, because considering it being a fat-tailed distribution, we expect that its exceedance becomes quite reliable at least at the most extreme quantiles. The reason of its poor performance could be that the data was very extreme beyond the limit it can handle making the distribution lack the capacity to effectively describe the tails. Hence, from our table neither of the two traditional time series was successful in describing the most extreme positive tails of our return series. Our results is in hand with that of Byström.

Figure 5.3 displays the 0.99 quantile plot together with the profile likelihood. The profile likelihood of the plot is used in finding the confidence intervals. This was the plot used to determine the confidence interval of the Value at Risk estimates at 0.99 level. The confidence interval of VaR and ES from Table5.1 tends to be asymmetric.


Excess Distribution
x [log Scale]
0.0
0.4
0.8
0.5
1.0
2.0
5.0
10.0
20.0
Fu(x-u)
u = 0.5


QQ-Plot of Residuals
Exponential Quantiles
0
2
4
6
0
2
4
6
8
Ordered Data



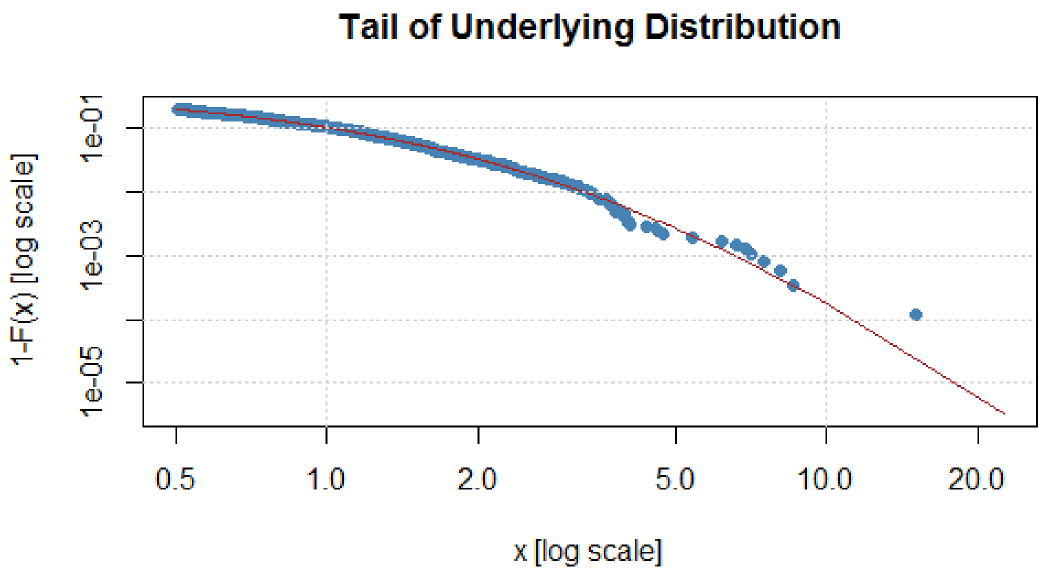


Figure 5.4: *Diagnostic plots of GPD model fitness*

In Figure 5.4, we have three plots of the GPD model fittings. The plots show good fit of the GPD models for the net returns and residual series. In the excess distribution plot, where mean of all excess values over the threshold u against the threshold itself is plotted, there's an indication of goodness of fit of the GPD. As a result, setting a high threshold at p=0.995 level is ideal.

The VaR and ES estimates above this level are then 0.999, 0.9995 and 0.9999 and their respective estimates are recorded in Table 5.3 together with corresponding confidence intervals. The confidence limits are obviously relatively wider at these three most extreme quantiles. Narrow confidence interval estimates both for VaR and ES at the first three upper quantiles provides high estimate accuracy. The Conditional GPD estimates in most extremes Table 5.2 are more closer to the expected than at the upper quantiles and outperformed much better than the traditional time series.

5.2 ACER Analysis

We continue our analysis with ACER. The ACER programs enables us to calculate and plot ACER functions, estimating the parameters in equation (3.55), and confidence interval for the predicted extreme value provided by the optimal curve. The ACER program does not assume the data process but provides the flexibility of choosing if we are modeling the tails of stationary or non-stationarity data. We will limit ourself to the result of the stationarity testing and treat our data as stationary in this analysis.

To take care of the seasonal components in our data with ACER, we have model the time dependence at $k = 1, k = 2, k = 6, k = 21$ and $k = 61$ which corresponds to the lags in AR filters as shown in equation (3.60) with estimates $\hat{\varepsilon_1}, \hat{\varepsilon_2}, \hat{\varepsilon_6}, \hat{\varepsilon_{21}}$and $\hat{\varepsilon_{61}}$. Modeling with these different k-values also captures the full dependence structure in the data.

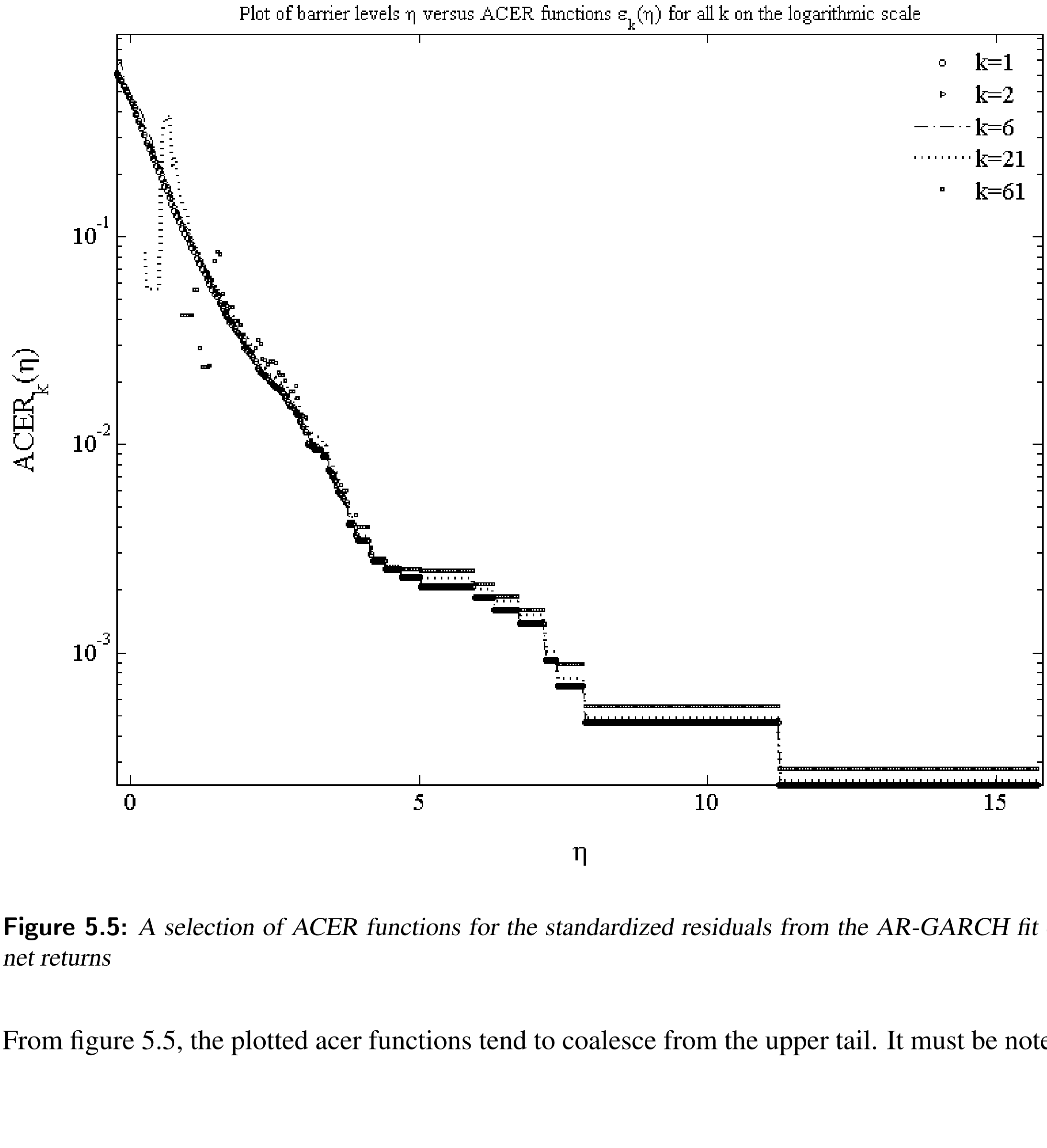


Figure 5.5: *A selection of ACER functions for the standardized residuals from the AR-GARCH fit of net returns*

From figure 5.5, the plotted acer functions tend to coalesce from the upper tail. It must be noted that there are lot of extreme observations at the far tail end as shown in figure. However, we can employ any of the ACER functions for further analysis because they converge. The first ACER function allows us to use all the data we have at hand. Accuracy is therefore improved and hence it is a preferred choice. As *k* gets larger, we loose some data during estimation but we still obtain good parametric estimates.

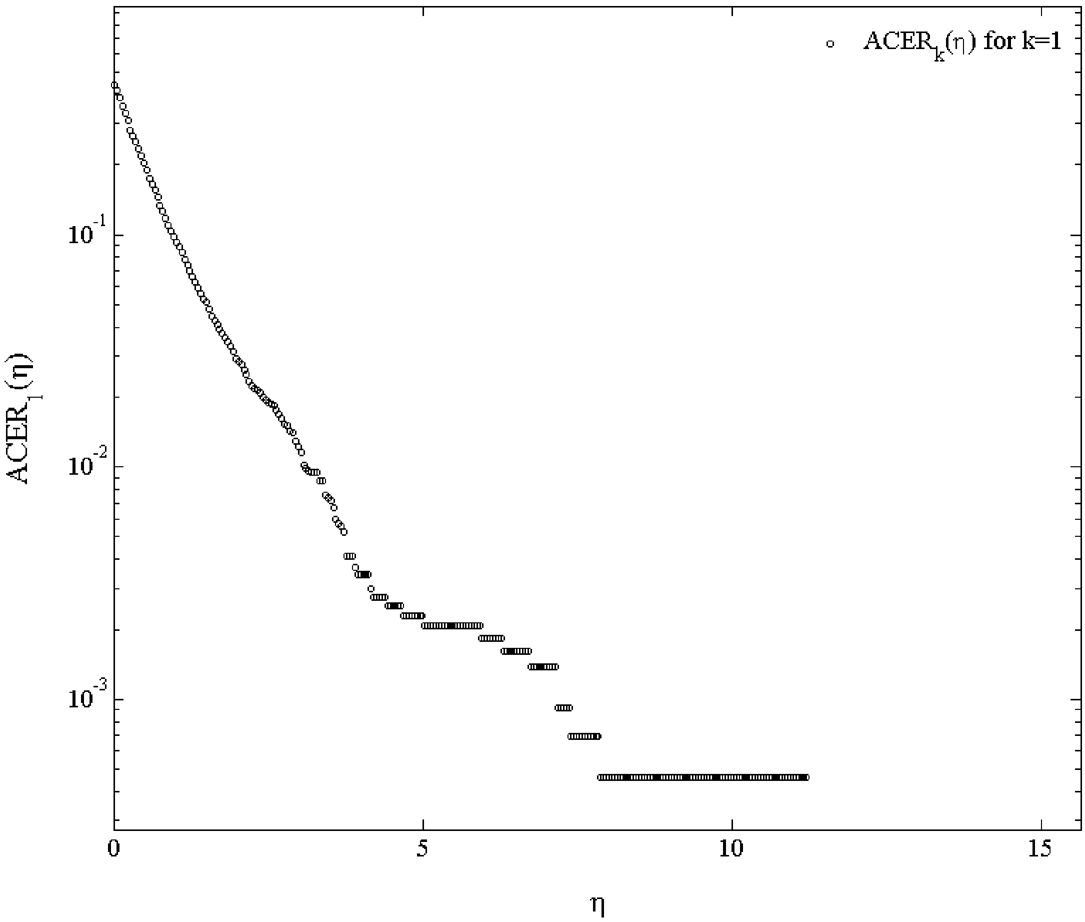


Figure 5.6: *Plot of chosen ACER function*

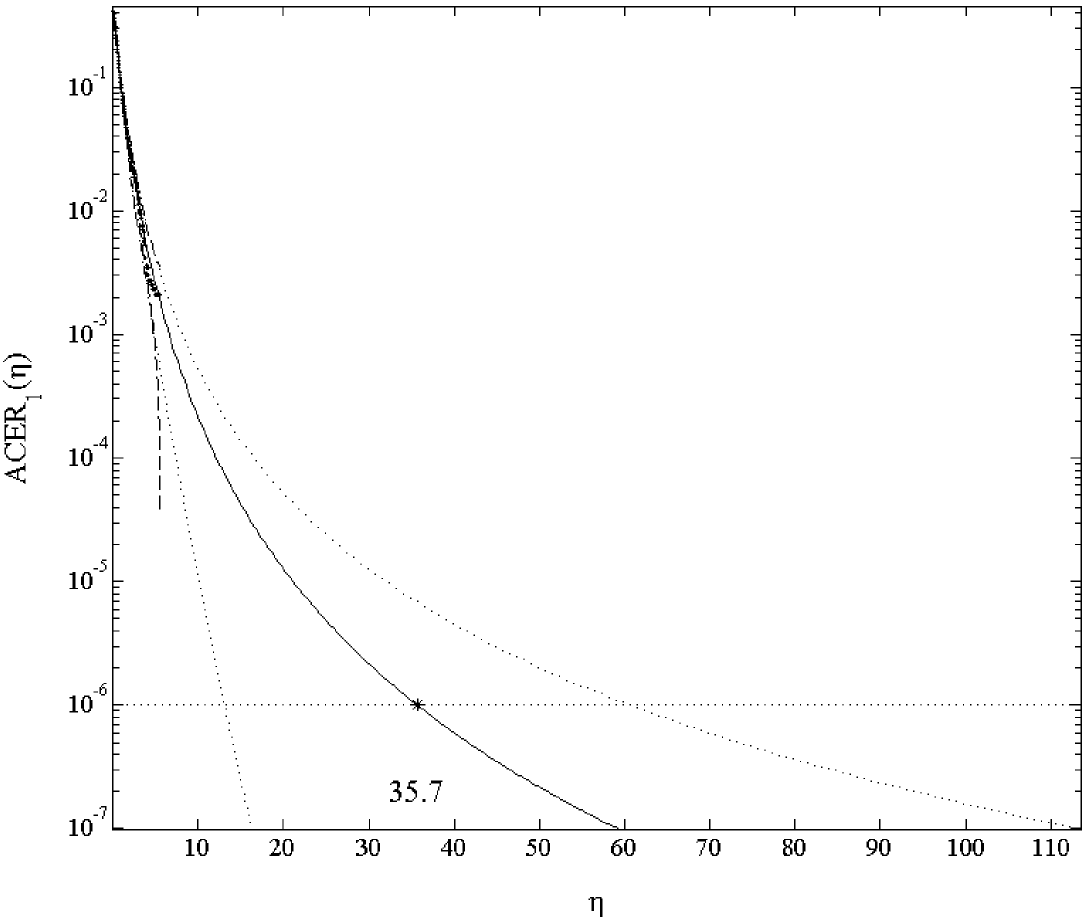


Figure 5.7: *Plot of extrapolated optimal curve and confidence bands for* ACER$_1$

The above figure displays the 100 year return level of the first ACER function. The plot also returns the estimated parameters of $\hat{\varepsilon}_1$. The tail marker η is the value of which ACER function behaves regularly. It is chosen either by visual inspection or automatically generated by the ACER program for $k \geq 2$. Just above the tail marker axis displays the estimated return level. From [9], the estimated return value is not very sensitive to the choice of η_1 provided it is selected with care. The value of parameter δ included in Table 5.5 shows the level of cutting of the most extreme observations, and our data set tends to exhibit very extreme tail behaviour as shown in Figure 5.5. In [7], δ lies within [0.5, 1]. A lower value of δ cuts more data. We have chosen $\delta = 1$ in above diagram which implies we are hardly cutting enough extreme tails in data. We will also see the effect of the return level and estimated parameters by varying δ. The value of fixed parameter c has been set to 1 during the optimization and we will therefore monitor how c changes. Detailed results shown in Table 5.7. Values in square bracket represent confidence intervals. In general, the value of the fixed

ACER k	η	q	b	a	c	γ	100yr Return Level
δ=2							
1	0.012	0.999	-0.068	0.899	1.667	2.057	59.13 [14.57, 142.3]
6	1.104	0.168	0.214	0.665	1.82	1.615	75.23 [11.67, 286.1]
21	2.5467	0.0198	2.309	8.818	4.226	0.342	566.4 [7.057, 6683]
61	5.957	-	-	-	-	-	-
$\delta = 1$							
1	0.012	0.4521	0.012	0.3931	1.014	4.723	35.67 [13.08, 60.61]
6	1.104	5.013	-1.627	0.5747	1.484	3.22	34.81 [11.21, 100.4]
21	2.5467	0.01904	2.547	22.72	3.383	0.3349	2378 [7.538, 21400]
61	-	-	-	-	-	-	-
δ=0.8							
1	0.012	7.36	-0.733	0.01161	0.5349	2805	19.51 [18.64, 20.5]
6	1.104	46.65	-2.619	0.0116	0.661	232.1	15.56 [11.89, 15.85]
21	2.5467	0.01975	2.404	3.718	2.903	0.5735	244.4 [5.869, 898]
61	-	-	-	-	-	-	-
δ=0.5							
1	0.012	0.4243	0.0426	0.01161	0.505	1.056	47.11 [46.11, 28.25]
6	1.104	3.394	-0.0728	-0.00382	0.4299	939	29.38 [505.4, 45.85]
21	2.5467	0.02	2.445	51.43	3.563	0.1979	4.16e5 [4.878, 7.45e8]
61	-	-	-	-	-	-	-

Table 5.4: *Estimated parameters of the sub asymptotic form together with the return level using ACER*

parameter c was not very far from 1, although we could observe the highest c value at 5, and the lowest at 0.4299. The deviation of the c value could be that we either overfit or underfit the sub-asymptotic distribution respectively. When cutting the extreme tails of the distribution, we wanted a value very close to 1, therefore these choices (0.99, 0.95, 0.90, 0.85) were selected but there was no significant difference between the results obtained with these values and 1. As a result we went as far down to 0.8. Being curious of the result when most of the extreme values is cut, $\delta = 0.5$ was analysed. The result seems quite interesting. For $k = 1$ and $k = 6$, The return level estimate lies outside the confidence limit and for $k = 21$ a very high return estimate was characterised by wide confidence interval. The reason simply could be that, much extreme observations were cut than we were supposed. An extremely high value of the shape parameter ξ associated with a very narrow confidence limits was observed at $k = 1$ and cutting at 0.8. From the table, we also investigated what will happen if we go beyond the range of δ. We therefore include the analysis of $\delta = 2$ with the view that much more extreme observations are added for estimation.

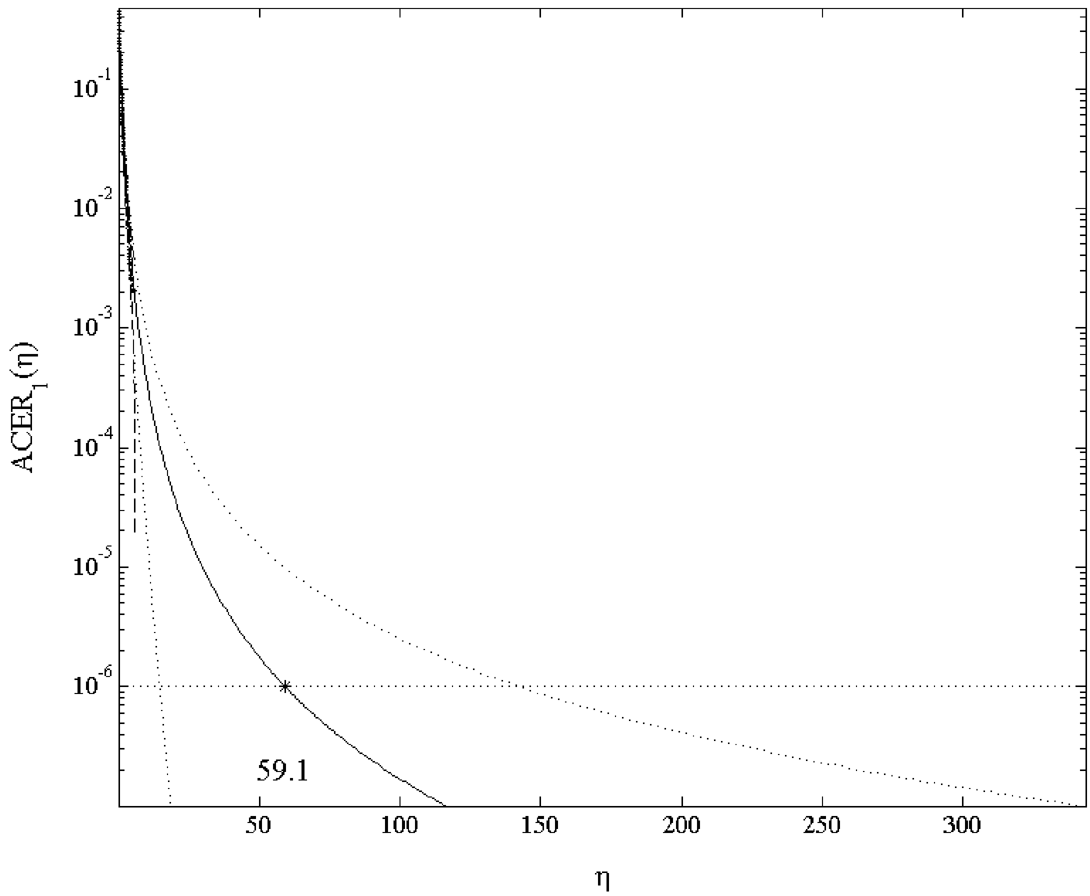


Figure 5.8: *Plot of extrapolated optimal curve* ACER_1 *using the standardized residual of net returns,* $\delta = 2$

The optimal values obtained for $\delta = 2$ were fairly good. From our ACER analysis so far, the parameter values may deviate from the optimal values in some cases, but we still get good estimates except of course when we are cutting most extreme values.

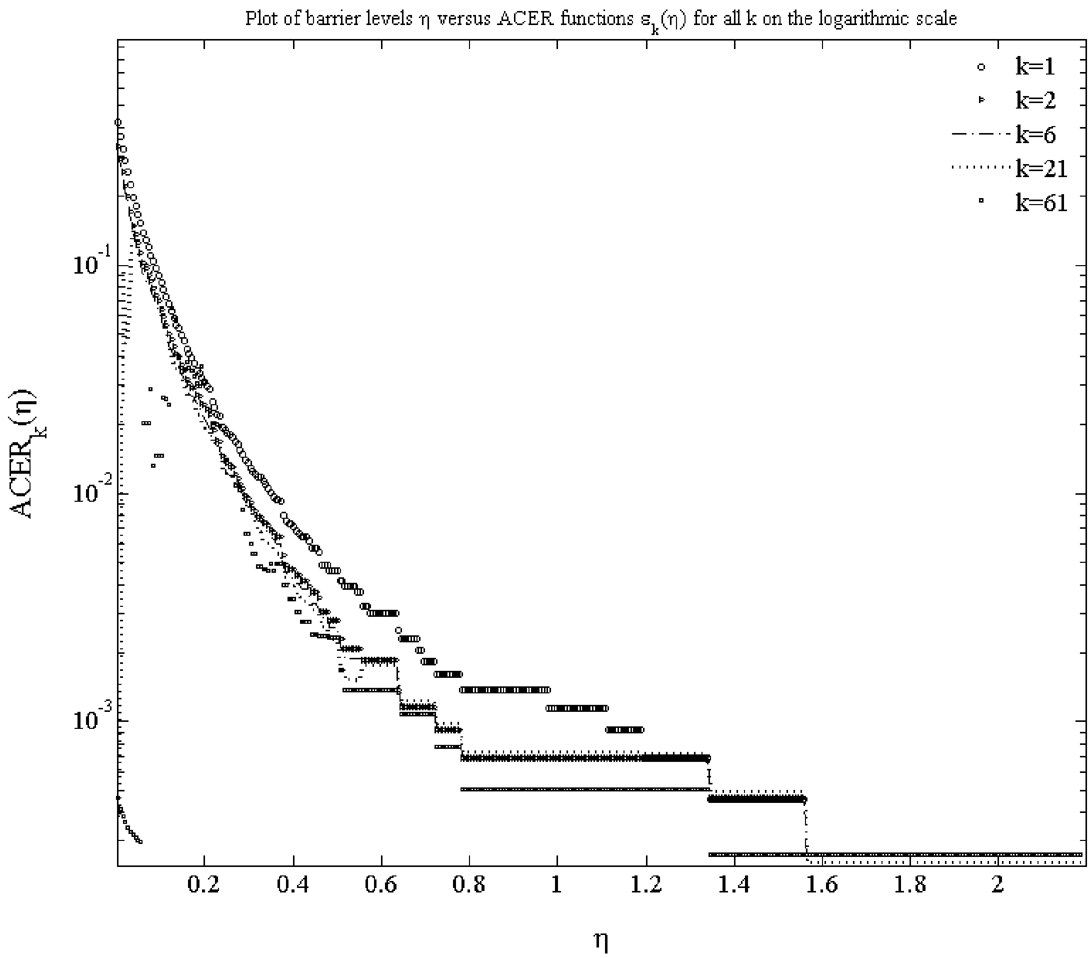


Figure 5.9: *A selection of ACER functions for the AR-GARCH fit of net returns*

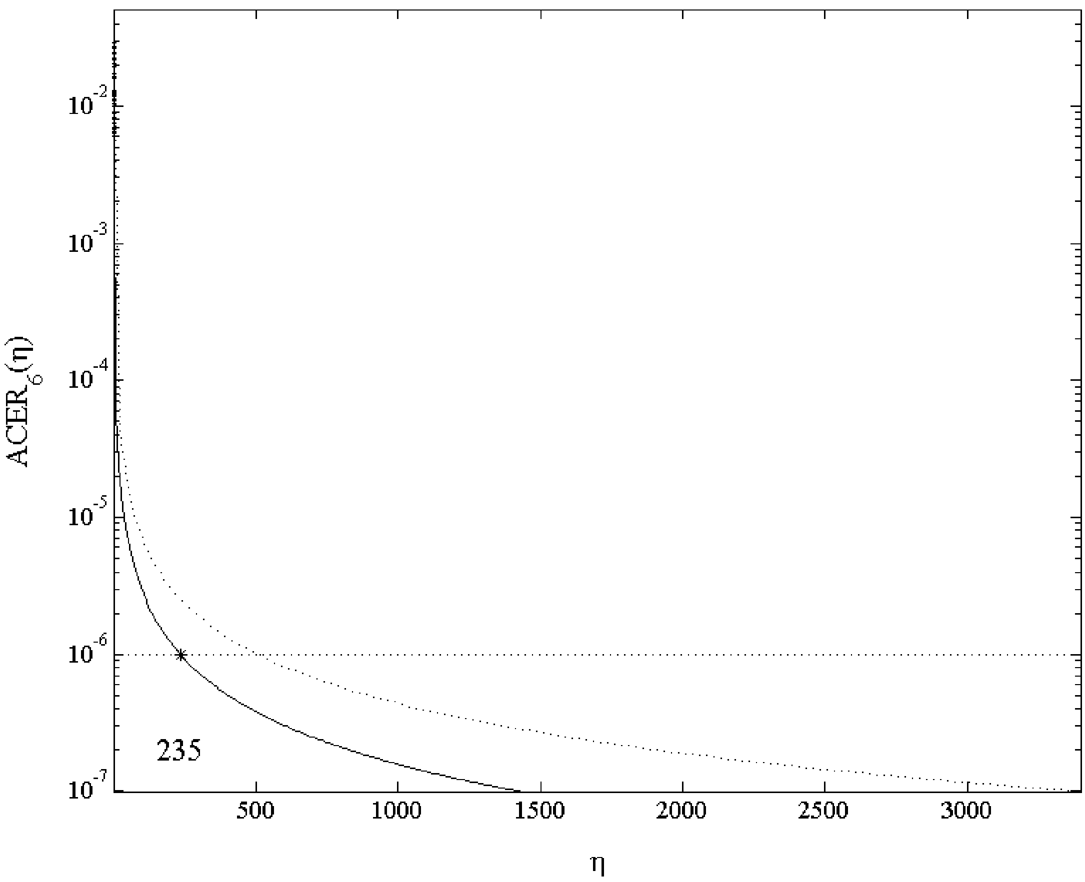


Figure 5.10: *Plot of extrapolated optimal curve and confidence bounds for* ACER_6

In Figure 5.12, The ACER function is plotted using the net return series. When compared to Figure 5.5, there are observable differences in the upper tails. The upper tail tends to hardly converge in Figure 5.12 but converges readily in Figure 5.5. With the former, we have been using the standardized residuals series whilst the raw simple net returns have been employed in the latter case. We then repeat Table 5.5 but this time using the simple net returns.

ACER k	η	q	b	a	c	γ	100yr Return Level
$\delta = 1$							
1	0.003	0.508	-0.02	3.148	0.832	4.681	6.82 [1.744, 16.07]
6	0.159	0.029	0.111	4.2e5	5	0.255	235 [2.75, 508.6]
21	-	-	-	-	-	-	-
61	-	-	-	-	-	-	-
δ=0.8							
1	0.003	0.508	-0.02	3.137	0.882	4.69	6.805 [1.79, 15.63]
6	-	-	-	-	-	-	-
21	-	-	-	-	-	-	-
61	-	-	-	-	-	-	-
δ=0.5							
1	0.003	0.574	-0.005	2.485	0.788	5.486	6.01 [2.043, 9.953]
6	-	-	-	-	-	-	-
21	-	-	-	-	-	-	-
61	-	-	-	-	-	-	-

Table 5.5: *Estimated parameters of the sub asymptotic form together with the return level using ACER*

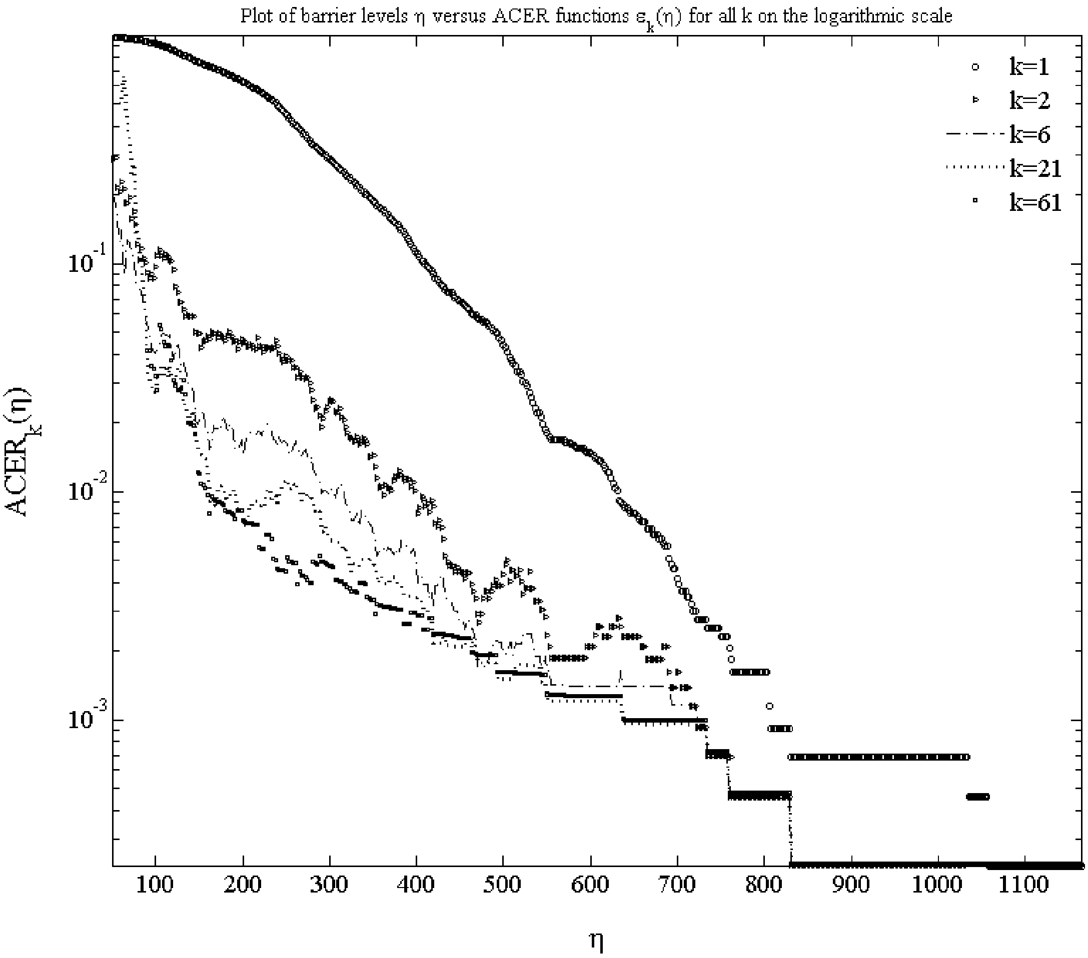


Figure 5.11: *A selection of ACER functions for the AR-GARCH fit of the original data set*

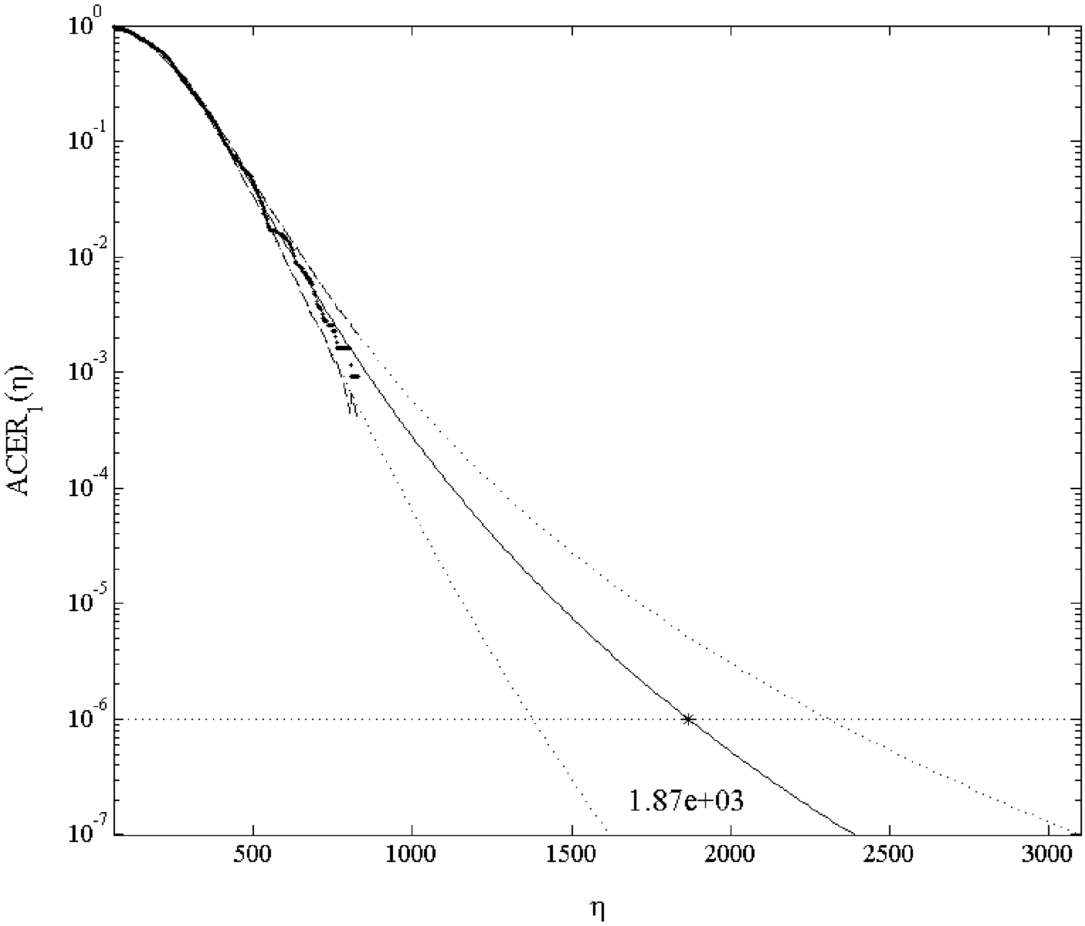


Figure 5.12: *A selection of ACER functions for the AR-GARCH fit of the electricity price data,* $\delta = 1$

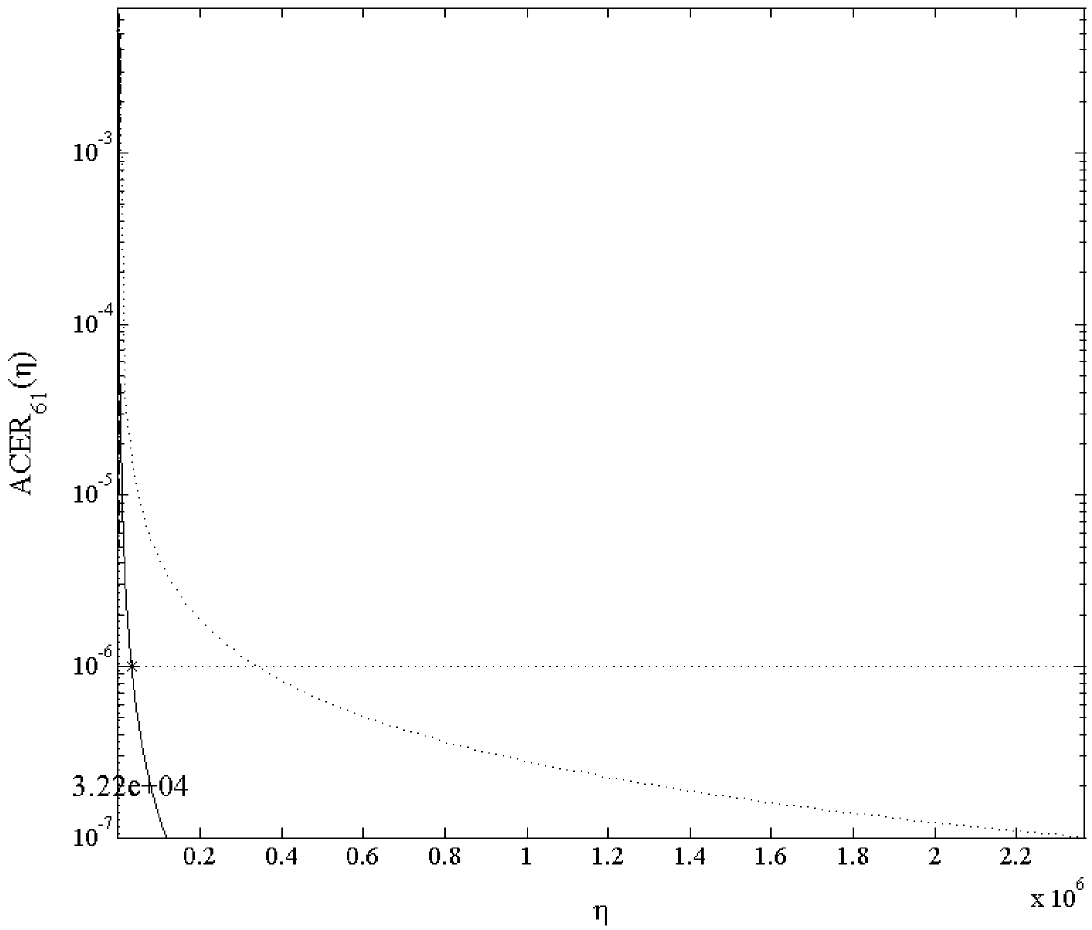


Figure 5.13: *A selection of ACER functions for the AR-GARCH fit of daily price data,* $\delta = 1$

The inverse of the parameter γ describes the shape (ξ) of the distribution. As said earlier, ξ explicitly determines the type of extreme value limit law we are considering. Since we are working with a heavy-tailed Frechét distribution, we expect $\xi > 0$ in our analysis. In Table 5.4, The values obtained for ξ were reasonable for our purpose when $\delta = 1$ and $\delta = 2$. However, cutting the most extreme observations at $\delta = 0.8$ and $\delta = 0.5$, ξ almost approached 0 when $k = 1$ and $k = 6$ for the former and when $k = 6$ for the latter. As a result, the underlying extreme value distribution is suspected to be asymptotically Gumbel. But from Table 5.5, (table 5.5 constructed with the simple net returns data), $\delta = 0.8$ and $k = 1$, $\xi = 0.213$ clearly indicates an underlying Frechét distribution.

ACER k	η	q	b	a	c	γ	100yr Return Level
$\delta = 1$							
1	100	1.001	-0.760	2.544e-7	2.486	3.866	1868 [1378, 2309]
6	260.679	0.017	228.6	0.00014	5	2.057	5.102e4 [986.2, 3.644e5]
21	260.68	0.028	190.1	0.0034	1.416	1.086	4.291e4 [951.2, 4.28e5]
61	288.26	0.135	17.03	0.00016	2.324	0.767	3.22e4 [942.4, 3.4e5]

Table 5.6: *Estimated parameters of the sub asymptotic form together with the return level using ACER*

Figure 5.6, higher return level estimates were characterised by wide confidence limits. From the output, the c values were quite far from 1 and $\gamma's$ supports the heavy-tailed Frechét distribution.

6 EVT Application to Managing Price Risk

The traditional time series models: both the parametric and non parametric though they provide good fit to the empirical distribution in areas where many observations are present, they work very poor in providing fit to the extreme tails. This is the major drawback since in risk management, the concern mainly lies in the tail probabilities and quantile estimation that in most cases are directly not observed from the data. Instead of using the whole price (return) observations present, EVT rather focuses on tail behaviour modeling using only extreme values.

For the purposes of risk management, the EVT foundation is based on the Fisher-Tippett theorem which aims in describing the limiting behaviour of sample maxima, the POT method. Unlike the POT method which is a GPD based, the ACER method is based on the Generalized Extreme Value distribution. Broadly speaking, the EVT based Value at Risk and Expected Shortfall are the most common tool for modeling risk. The ACER method functions in a different way of risk modeling by providing the exact extreme value distribution inherent in the data, in addition to an option of cutting the tail of highly extreme observations and finally predicting the return level together with its confidence limit.

7 Discussion

For the POT analysis, Byström paper has been used as a guide. We went further with the VaR and ES estimation and their respective confidence limits. In his time dependency modeling, AR part was modeled only to include the weekly seasonality. To gain more insight into seasonal variational effect, we model the AR to include monthly and quartely effect.
Table 5.2 happens to be one of the most important results in this thesis. It depicts how the traditional time series and EVT inspired models are used in describing tail quantiles. As mentioned earlier, the traditional time series performed poorly and with the two EVT based models considered, the ACER performed extremely well in describing the tail quantiles. The number of exceedances obtained using the ACER function was closer to the expected at all quantiles and even get better as we move into more extreme quantiles. Though the POT method requires the calculation of unconditional VaR before finally estimating the conditional GPD, the ACER method does so based on the exceedances of the observed data. We chose to present the result associated with the conditional case of the GPD because though the two coincide in iid case, the unconditional ignores the dependence structure of our data but under dependence there is divergence of which the conditional GPD is of much interest. The conditional VaR also provides a more reliable risk assessment through explicit modeling of the stochastic nature of conditional volatility. Generally, electricity prices as well as financial returns are conditional heteroscedastic and hence far from being an iid. In our analysis, we assumed iid for the EVT based GPD residuals. The ACER method provides generalization to data (dependent) with anti-clustering requirement on the extremes. The seasonal effect was effectively modeled with the AR filter but in practice, provided we have available amount of data, explicit modeling of seasonal effects are automatically accounted for by the ACER program. Moreover, In our analysis, there were situations where we had no enough tail data to estimate for the confidence interval. This was the case of the top k values in the ACER implementation but it's also a general drawback for extreme value statistics. We therefore utilised more data when lower values of k are used. We should also remember that the amount of extreme values to cut depends on how extreme the observations we have are. Cutting with $\delta = 0.8$ removed some of the most extreme observations and provided very good estimates and narrow confidence limits for our first two ACER functions. The ACER method is known to provide better estimate of confidence interval than the POT [9]. Moreover , risk management practices should based on confidence interval than tail probabilities [6]. Though the two are dual, the confidence interval based ones are easier to compute than using the tail probabilities. In our analysis, the EVT based ACER was used in estimating the confidence interval and return levels and it provided reasonable confidence limits. As a final point we must notice that, the drawbacks using the GPD based EVT method which basically include the problem of selecting suitable threshold and independence assumption, are no issues with the ACER method.

7.1 Conclusion

This thesis aims to provide clear picture of how risk can be managed in the electrical power market. It does so by applying and evaluating extreme value theory and value-at-risk analysis as risk management tools to a deregulated power market. As a result, the foundation of EVT was critical and considered in detail. EVT is a powerful tool to estimate the effects of extreme events in risky markets, based on sound statistical methodology and for modeling tail related risk measures. In modeling the tails of our returns (prices) the traditional time series was compared to the EVT based approaches. The traditional models, however, did not produce good enough results and accuracy to be considered as reliable risk management approaches. The reason for their poor performance was due to the extreme nature of prices observed in electricity markets.

Price forecasting models and accurate risk management methods are the essential tools required for risk management purposes [8]. Bystrom innovations which employs AR-GARCH filter with EVT further applied is found to produce reliable forecast when applied to our data.

In our analysis of extreme events, both the POT and ACER methods did very well in predicting the tail behavior. Our aim was not to compare the two methods but outline their usefulness in modeling tail behavior.

7.2 Further Work

The ACER method has been successful in describing the regular tail behavior of extreme events, providing good estimate of return level and confidence limits. However, in order for risk to be quantified and to determine the size of potential losses given occurrence of bigger loss, more work is needed to estimate the Value at Risk (VaR) and Expected Shortfall (ES) with this new method.

Though two EVT methods considered in this thesis work in completely different way and each has its own drawbacks, it could be interesting to combine those ideas with more detailed statistical information from financial time series data to precisely estimate risk measures.

Bibliography

[1] David E Allen, Abhay K Singh, and Robert Powell. Extreme market risk-an extreme value theory approach. *School of Accounting Finance & Economics Edith Cowan University*.

[2] Johnson B. and Barz G. Selecting stochastic processes for modelling electricity prices. *Energy modelling and the management of uncertainty*, 1999.

[3] Hans N.E. Byström. Extreme value theory and extremely large electricity price changes. *International Review of Economics and Finance*, 14:41–55, 2005.

[4] Klüppelberg Claudia, Meyer-Brandis Thilo, and Schmidt Andrea. Electricity spot price modeling with a view towards extreme spike risk. *Quantitative finance*, pages 1–10, 2009.

[5] Stuart Coles. *An Introduction to Statistical Models of Extreme Values*. Springer-Verlag London, 2001.

[6] Francis X Diebold, Til Schuermann, and John D. Stroughair. Pitfalls and opportunities in the use of extreme value theory in risk management. *Advances in Computational Finance*, pages 3–12, 1998.

[7] Oleh Karpa. Acer user guide. *Centre for Ships and Ocean Structures (CeSOS), NTNU*, 2012.

[8] Kimmo Lehikoinen. Extreme value theory in risk management for electricity market. 58016L:16–23, 2007.

[9] Arvid Næss and Oleg Gaidai. Estimation of extreme values from sampled time series. *Structural Safety*, 31(4):325 - 334, July 2009.

[10] Artzner Phillipe and Delbaen Freddy. Coherent measures of risk. *Mathematical Finance*, 9(3), 1998.