NTNU
Norwegian University of
Science and Technology

Comparison of ACER and POT Methods for estimation of Extreme Values

Kai Erik Dahlen

Master of Science in Physics and Mathematics
Submission date: June 2010
Supervisor: Arvid Næss, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Problem Description

The master thesis will consist of further generalize the ACER method to make it useable for estimation of extreme values from heavy tailed distributions. The performance of this modified ACER method will then be compared to the asymptotic POT method. This will be done by applying both methods to both generated synthetic data and real world data sets.

Assignment given: 26. January 2010
Supervisor: Arvid Næss, MATH

Preface

This is thesis is a continuation of my work on my project, 'Comparison of ACER and POT methods' , which was written during the fall semester of 2009. I have been working the the average conditional exceedance rate (ACER) and the peaks over threshold (POT) methods, and with the comparison of these two. I would like to express my thanks to professor Arvid Næss for being my supervisor for the last year, and for help and encouragements along the way. I also would like to thank associate professor Sjur Westgaard for providing some of the data used in this thesis, and Olexandr Batsevych for providing some of the Matlab programs used for the ACER method, and for giving a great introduction to the use of the program. At last I would like to thank my girlfriend for being supportive and encouraging during my work on this thesis.

Kai Erik Dahlen NTNU, Trondheim 22nd June, 2010

Abstract

In this thesis I have been working on comparing the performance of the ACER and POT methods for prediction of extreme values from heavy tailed distributions. To be able to apply the ACER method to heavy tailed data the ACER method was first modified to assume that the underlying extreme value distribution would be a Fréchet distribution, not a Gumbel distribution as assumed earlier. These two methods have then been tested with a wide range of synthetic and real world data sets to compare their preformance in estimation of these extreme values. I have found the ACER method seem to consistently perform better in the terms of accuracy compared to the asymptotic POT method.

Contents

Preface **i**

Abstract **iii**

1 Introduction **1**

2 Theory **3**
2.1 Return period 4
2.2 Peaks over threshold (POT) 5
2.2.1 Threshold selection 5
2.2.2 Dependent sequences 6
2.3 Average conditional exceedance rate (ACER) 6
2.3.1 Confidence intervals for return levels 10
2.4 Heavy tailed distributions 10
2.5 AR-GARCH model 11

3 Data **13**
3.1 Synthetic data 13
3.1.1 Pareto distribution 13
3.1.2 Student's t-distribution 15
3.1.3 ARCH/GARCH models 16
3.2 Data series 16
3.2.1 Electric market spot price data 16
3.2.2 Dow Jones Index 18

4 Analysis of data and results **21**
4.1 Analysis of synthetic data 21
4.1.1 Pareto distribution 21
4.1.2 Student's t-distribution 23
4.1.3 ARCH/GARCH model 31
4.2 Analysis of data 39
4.2.1 Electricity market spot price data 39
4.2.2 Dow Jones Index 46

5 Conclusion **51**
5.1 Further work 51

Biliography **53**

Chapter 1

Introduction

Extreme value theory (EVT) is a branch of statistic dealing with the extremes of a distribution, being the extremely small or extreme large observations. Important applications of extreme value theory is i.e. estimating return levels for a process for a given time period, the levels expected to be exceeded during this time period. By being able to accurately predicting this return levels we are able to estimate risk or being able to correctly dimension structures.

Since we are working with the extreme values from a distribution we are operating in the very tail of the distribution. Considering this it is clear that when working with extreme value statistic, the amount of data, or at least extreme observations, will be sparse. In classical EVT one is working with block maxima (or minima), fitting these observations to a distribution of the generalised extreme value family. This procedure is regarded as a wasteful approach as many observations are discarded. Further the peaks over threshold (POT) method deals with, as it's name implies, exceedances over a given threshold. This makes better use of the data as we are no longer looking at the distribution of block maxima, but the distribution of threshold exceedances. Though this method is not without flaws as it's results is based on a limiting distribution as the threshold increases. In this thesis we are going to compare this POT method to a method called average conditional exceedance rate (ACER) method. This method does not rely on asymptotic behaviour of data, but rather uses the sub asymptotic form of the generalised extreme value distribution as an assumption of the behaviour in the tail of a distribution. Earlier the ACER method has been using the sub asymptotic form of the Gumbel distribution, and has been applied to data that can be assumed has the Gumbel distribution as extreme value distribution with great results. The ACER method has it's advantage over these asymptotic methods that it is able to utilize the available data in a better manner, no having to rely on block maxima or exceedances over a given threshold. We are now going to further try to generalise the ACER method to by looking at data generated by heavy tailed processes. This will be done by assuming another form of the sub asymptotic extreme value distribution.

Chapter 2

Theory

In the following chapter we shall consider a series of independent stochastic variables $X = (X_1, ..., X_N)$, with an unknown common distribution. The extreme value of these stochastic variables is defined as

$$M_N = \max(X_1, ..., X_N), \tag{2.1}$$

with the exact distribution of M_N given by

$$\mathrm{P}(M_N \leq \eta) = \mathrm{P}(X_1 \leq \eta) \cdots \mathrm{P}(X_N \leq \eta) = F_X^N(\eta), \tag{2.2}$$

since all observations are iid. If one where to estimate the distribution of F_X, small errors in this estimate would lead to great errors in F_X^N. Instead we look at the behaviour of F_X^N when $N \Rightarrow \infty$, and it can then be shown that if there exists a sequence of constants $\{a_N > 0\}$ and $\{b_N\}$ such that the distribution of $M_N^* = (M_N - b_N)/a_N$ is non-degenerate, the distribution of M_N^* will be of one of three families:

$$\begin{aligned}
\mathrm{I} : G(\eta) &= \exp\Big\{ -\exp\Big[\Big(\frac{\eta - b}{a}\Big)\Big]\Big\}, -\infty < \eta < \infty \\
\mathrm{II} : G(\eta) &= \exp\Big\{ -\Big(\frac{\eta - b}{a}\Big)^{-\alpha}\Big\}, \eta > b \\
\mathrm{III} : G(\eta) &= \exp\Big\{ -\Big[-\Big(\frac{\eta - b}{a}\Big)\Big]^{-\alpha}\Big\}, \eta < b.
\end{aligned}$$

The distributions I, II and III is the Gumbel, the Fréchet and the Weibull distribution respectively. This is known as the Extremal Types Theorem, [4]. This three types can also be written as

$$\mathrm{Prob}(M_N \leq \eta) \longrightarrow G(\eta) = \exp\left\{ -\Big[1 + \xi\Big(\frac{\eta - \mu}{\sigma}\Big)\Big]_+^{1/\xi} \right\}, N \to \infty, \tag{2.3}$$

where $[z]_+ = \min(0, z)$. This distribution is the generalised extreme value (GEV) distribution with location parameter μ, scale parameter σ and shape parameter ξ. For values of the shape parameter $\xi > 0$, $\xi < 0$ and $\xi \to 0$ the general extreme value distribution equals the Fréchet, the Weibull and the Gumbel distribution respectively.

2.1 Return period

The return period of a level η for a random variable X is defined as

$$R = \frac{1}{P(X > \eta)} = \frac{1}{1 - F_X(\eta)}. \tag{2.4}$$

This means that the return rate R for η is the mean number of trials that must be done for X to exceed η.

When modeling POT the return period for a level $\eta_R = u + y$, where u is the threshold, is given by

$$R = \frac{1}{\lambda P(X > \eta_R)} = \frac{1}{\lambda P(Y > y)}. \tag{2.5}$$

Here λ is the mean crossing rate of the threshold per block (i.e. per year, month etc.), or the average proportion of observations that fall over the threshold. From (2.5) is follows that

$$P(Y \leq y) = 1 - \frac{1}{\lambda R}, \tag{2.6}$$

and since the distribution of Y is know, we have from (2.11) that

$$\eta_R = u - \frac{\tilde{\sigma}}{\xi}(1 - (\lambda R)^{\xi}), \tag{2.7}$$

for $\xi \neq 0$, and

$$\eta_R = u + \tilde{\sigma} \ln(\lambda R), \tag{2.8}$$

for $\xi = 0$.

Confidence intervals for the return level η_R is computed using the delta method, that is assuming that the maximum likelihood estimator is multi normal distributed with expectation equal to the real parameter value and variance covariance matrix V. The variance of the return level η_R can then be estimated by the delta method as

$$\mathrm{Var}(\eta_R) \approx \nabla \eta_R^T V \nabla \eta_R, \tag{2.9}$$

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

$$\nabla \eta_R = \Big[\frac{\partial \eta_R}{\partial \lambda}, \frac{\partial \eta_R}{\partial \sigma}, \frac{\partial \eta_r}{\partial \xi}\Big]^T. \tag{2.10}$$

For the ACER method the return levels are estimated by inverting (2.28) with the values for a, b, c and q, found by the least square routine, for the exceedance rate of interest.

2.2 Peaks over threshold (POT)

When fitting the GEV distribution to a set of data only block maxima (or minima) can be used. This is of course very wasteful when considering that a block may be a whole year, or a single block can contain several observations more extreme than observations of another block. To be able to use more of the data it can be useful to look at exceedances of a high threshold instead of block maxima. This describes the peaks over threshold (POT) method and is based on that if M_N is distributed according to (2.3), then for large enough values for a threshold u, the threshold excess $X - u$, conditional on $X > u$, can be approximated with

$$\text{Prob}(X - u > y | X > u) \approx H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)_+^{-1/\xi}, \tag{2.11}$$

where $\tilde{\sigma} = \sigma + \xi(u - \mu)$. This distribution is the generalized Pareto (GP) distribution with parameters $\tilde{\sigma}$ and ξ, where μ, σ and ξ is the same for both the GEV and GP distributions. When modeling POT the parameters of the distribution is estimated by i.e. maximum likelihood.

2.2.1 Threshold selection

As mentioned above the GP distribution is only a valid approximation to the threshold excess if the threshold is large enough. In practice that means that the threshold must be decided on by the user, a task which is not always trivial. Choosing a large value for the threshold may ensure that the asymptotic assumptions made are indeed satisfied, but choosing too large of a threshold will see much of the data discarded and hence the estimation of the parameters in the GP distribution may be poor. So the desired threshold is the lowest value for which the threshold excess fits the GP distribution. As a start it can be informative to look at a mean residual life plot of the data. The idea behind the mean residual life plot is to look at the first moment of the GP distribution

$$\text{E}X = \frac{\tilde{\sigma}}{1 - \xi} = \frac{\sigma + \xi(u - \mu)}{1 - \xi}, \ \xi < 1, \tag{2.12}$$

which is linear in u. The mean residual life plot is obtained by plotting empirical mean exceedance of the threshold against the threshold, that is plotting

$$\left\{ \left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) : u < x_{\max} \right) \right\}, \tag{2.13}$$

where $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{n_u}$ is the ordered observations above u. This plot should then be linear above a threshold u_0 for which the GP distribution is a valid approximation. In practice it can be difficult to interpret where the plot actually becomes linear since as the threshold increases so does the confidence bounds, which is obtained by the central limiting theorem for sample means.

Another method to aid in the selection of threshold is obtained by looking at the estimated shape and scale parameters. The shape parameter ξ is the same as for the GEV distribution. That means that above a threshold for which (2.11) is valid the estimate of the shape parameter should be constant. As for the scale parameter $\tilde{\sigma} = \sigma + \xi(u - \mu)$, the modified scale parameter $\sigma^* = \tilde{\sigma} - \xi u$ is independent of u and should be constant above the chosen threshold. Plotting these estimates against a range of thresholds and observing when the estimates become approximately constant will give a good indication of what threshold to choose.

2.2.2 Dependent sequences

When introduction the POT method it was specified that the observations needed to be independent. This means that when modeling POT all threshold exceedances should be independent. In practice this is obviously rarely the case. The most common way to solve this, as stated by [4], is to decluster the data, that is identifying clusters of dependent observations by an empirical rule and keeping only each clusters maxima. An appropriate choice of the empirical rule will ensure that the set of threshold excesses can be assumed independent. The use of declustering is considered to be a rather wasteful approach as all the data except the cluster maxima is discarded.

All computations regarding the use of the POT method is carried out using the R package POT, [7].

2.3 Average conditional exceedance rate (ACER)

For the ACER method let M_N be defined as in (2.1), but now no independence between the X_i's is assumed. The exact distribution of M_N will then be

$$\begin{aligned} \text{Prob}(M_N \leq \eta) &= \text{Prob}(X_1 \leq \eta, ..., X_N \leq \eta) \\ &= \text{Prob}(X_N \leq \eta | X_{N-1} \leq \eta, ..., X_1 \leq \eta)\text{Prob}(X_1 \leq \eta, ..., X_n \leq \eta) \\ &= \prod_{j=2}^{N} \text{Prob}(X_j \leq \eta | X_1 \leq \eta, ..., X_{j-1} \leq \eta) P(X_1 \leq \eta). \end{aligned} \tag{2.14}$$

Instead of assuming independence between the X_i's we can now assume a Markov-like property, or a k-step memory. With this assumption we can approximate (2.14) with

$$\begin{aligned} \text{Prob}(M_N \leq \eta) &\approx \prod_{j=k}^{N} \text{Prob}(X_j \leq \eta | X_{j-k+1} \leq \eta, ..., X_{j-1} \leq \eta) \\ &\quad \cdot \text{Prob}(X_{k-1} \leq \eta | X_1 \leq \eta, ..., X_{k-2} \leq \eta) \\ &\quad \cdot \text{Prob}(X_2 \leq \eta | X_1 \leq \eta) P(X_1 \leq \eta) \\ &= \prod_{j=k}^{n} (1 - \alpha_{kj}(\eta))(1 - \alpha_{k-1,k-1}(\eta)) \cdots (1 - \alpha_{11}(\eta)), \end{aligned} \tag{2.15}$$

where

$$\alpha_{kj}(\eta) = \text{Prob}(X_j > \eta | X_{j-k} \leq \eta, ..., X_{j-k+1} \leq \eta) \tag{2.16}$$

for $j \geq k \geq 2$. In the case where $k = 1$ we have $\alpha_{1j}(\eta) = \text{Prob}(X_j > \eta)$. Using that $(1 + x) \approx e^x$ if $|x| << 1$, equation (2.15) becomes

$$\text{Prob}(M_N \leq \eta) \approx P_k(\eta) = \exp\left(-\sum_{j=k}^{N} \alpha_{kj}(\eta) - \alpha_{k-1,k-1}(\eta) - \cdots - \alpha_{11}(\eta) \right). \tag{2.17}$$

In most cases $n >> 1$, so we can approximate (2.17) with

$$P_k(\eta) \approx \exp\left(-\sum_{j=k}^{N} \alpha_{kj}(\eta) \right). \tag{2.18}$$

Since $\alpha_{kj}(\eta)$ is defined as the probability that the j'th observation exceeds η conditional on that the $k-1$ previous observations was below η, it is reasonable to interpret $\sum_{j=k}^{N} \alpha_{kj}(\eta)$ as the expected number of independent exceedances of the level η conditioning on the $k-1$ previous observations. We can now introduce the ACER functions, defined as

$$\bar{\epsilon}_k(\eta) = \frac{1}{N-k+1} \sum_{j=k}^{N} \alpha_{kj}(\eta), \tag{2.19}$$

which then will be a representation of the average conditional exceedance rate (ACER). To estimate $\bar{\epsilon_k}$ the following random functions is introduced:

$$A_{kj}(\eta) = \text{I}(X_j > \eta, X_{j-1} \leq \eta, ..., X_{j-k+1} \leq \eta),\ j = k, ...N,\ k = 2, 3, ... \tag{2.20}$$

and

$$B_{kj}(\eta) = \text{I}(X_j \leq \eta, ..., X_{j-k+1} \leq \eta),\ j = k, ..., N,\ k = 2, 3... \tag{2.21}$$

where $\text{I}(A)$ is the indicator function for the event A. It is now possible to write $\bar{\epsilon}_k$ terms of A_{kj} and B_{kj} instead of α_{kj} obtaining

$$\bar{\epsilon}_k(\eta) = \lim_{N \to \infty} \frac{\sum_{j=k}^{N} A_{kj}(\eta)}{\sum_{j=k}^{N} B_{kj}(\eta)}. \tag{2.22}$$

The sample estimation of $\bar{\epsilon}_k$ will then be

$$\hat{\epsilon}_k(\eta) = \frac{1}{R} \sum_{r=1}^{R} \hat{\epsilon}_k^{(r)}(\eta), \tag{2.23}$$

where R is the number of realizations (i.e. number of time series) and

$$\hat{\epsilon}_k^{(r)}(\eta) = \frac{\sum_{j=k}^{N} A_{kj}^{(r)}(\eta)}{\sum_{j=k}^{N} B_{kj}^{(r)}(\eta)}. \tag{2.24}$$

For construction of confidence intervals for the ACER functions the standard deviation $\hat{s}_k(\eta)$ is estimated by

$$\hat{s}_k(\eta)^2 = \frac{1}{R-1}\sum_{r=1}^{R}(\hat{\epsilon}_k^{(r)} - \hat{\epsilon}_k)^2. \tag{2.25}$$

If it's assumed that the realizations are independent the 95 % confidence interval can be computed as

$$\text{CI} = (\hat{\epsilon}_k - 1.96\hat{s}_k(\eta)/\sqrt{R}, \hat{\epsilon}_k + 1.96\hat{s}_k(\eta)/\sqrt{R}). \tag{2.26}$$

Originally the underlying asymptotic extreme value distribution is assumed to be of Gumbel type. With this assumption the asymptotic form of the Gumbel distribution is used as a guide to the sub asymptotic form

$$\bar{\epsilon}_k \approx q_k(\eta)\exp\{-a_k(\eta - b_k)^{c_k}\}, \tag{2.27}$$

where $q_k(\eta)$ is slow varying compared to the exponential function. Now it is assumed that for $\eta > \eta_1$ the function q_k varies so slow compared to the exponential function that can be assumed to constant. So the form of the tail becomes

$$\bar{\epsilon}_k \approx q_k \exp\{-a_k(\eta - b_k)^{c_k}\},\ \eta > \eta_1. \tag{2.28}$$

The value η_1 is called a tail marked and is chosen by the user for a plot of $(\eta, \log\tilde{\epsilon}_k)$ as the marker for regular tail behavior. Even though the tail marker is a value specified by the user it is from [1] known that it does not greatly influence the estimated return level.

This sub asymptotic assumption is from [1] shown to be valid for a range of data, including wind speed data. Problems arise when the data considered does not have an underlying asymptotic extreme value distribution of Gumbel type. If the data considered is generated by a i.e. a heavy tailed process, the underlying extreme value distribution will not be of Gumbel type, which decays exponentially, and the ACER method will generally lead to a bad fit. We can now, instead of assuming an underlying Gumbel distribution, assume that the underlying extreme value distribution is of Fréchet type, which decays polynomially [4]. From this we can use the asymptotic form, the Fréchet distribution, as a guide for the sub asymptotic form. From this we get the sub asymptotic generalised extreme value (SGEV) distribution

$$G(\eta) = \exp\Big\{ - q\Big(1 + a(\eta - b)^c\Big)^{-1/\xi}\Big\} \tag{2.29}$$

for shape parameter $\xi \neq 0$. We are here interested in the case were $\xi > 0$, the Fréchet distribution. The case when $\xi < 0$ is not of much interest for us, as the Weibull distribution has a upper limit which will not be the case for the data we are considering. From this assumption the parametric form for the ACER function can be approximated with

$$\begin{aligned}\tilde{\epsilon}_k(\eta) &\approx q_k(1 + \xi a(\eta - b_k)^{c_k})^{-1/\xi} \\ &= q_k(1 - \tilde{a}_k(\eta - b_k)^{c_k})^{\gamma}\ \eta > \eta_1,\end{aligned} \tag{2.30}$$

where $\gamma = -1/\xi$, $\tilde{a}_k = a_k/\xi$ and η_1 is again a chosen tail marker. Here, as was the case in (2.28), it is assumed that q_k varies so slow in the tail that it can be assumed constant. Using the empirical values for the ACER function, and taking the logarithm on both sides we obtain

$$\log \hat{\epsilon}_k = \log q_k + \xi \log\left(1 - \tilde{a}_k(\eta - b_k)^{c_k}\right). \tag{2.31}$$

It is now possible to obtain parameter estimation by minimising the weighted square error function

$$\begin{aligned} F(\tilde{a}, b, c, q, \xi) &= \sum_{i=1}^{n} w_i |\log \hat{\epsilon}(\eta_i) - \log q - \gamma \log\left(1 - \tilde{a}(\eta_i - b)^c\right)|^2 \\ &= \sum_{i=1}^{n} w_i (y_i - \log q - \xi x_i)^2, \end{aligned} \tag{2.32}$$

where $y_i = \log \hat{\epsilon}(\eta_i)$ and $x_i = \log(1 - \tilde{a}(\eta_i - b)^c)$. The weights, w_i, are calculated as

$$w_i = \left(\ln \mathrm{CI}^+(\eta_i) - \ln \mathrm{CI}^-(\eta_i)\right)^{-2}, \tag{2.33}$$

where $\mathrm{CI}^+(\eta_i)$ and $\mathrm{CI}^-(\eta_i)$ is the upper and lower bounds of the 95% confidence interval at level η_i. These weights are used in order to put less emphasis on the more uncertain values for the ACER function. Also, it is possible to change the exponent in (2.33) to put more or even less emphasis on the uncertain values, though the exponent has not been changed in the parameter estimation here. A problem is that the function which should be minimised (2.32) is a function of five variables, which will not be trivial to minimise. To make things easier we have decided to use

$$\gamma^* = \frac{\sum_{i=1}^{n} w_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_{j=1}^{n} w_j (x_j - \bar{x})^2} \tag{2.34}$$

as an estimator for γ, and

$$\log q^* = \bar{y} - \gamma^* \bar{x} \tag{2.35}$$

as an estimator for $\log q$. Substituting these estimators back into (2.32) the function which should be minimised can be represented as a function of three variables instead of five. To minimize (2.32) the e04wd function of the NAG[1] toolbox for Matlab is used. This function uses the sequential quadratic programming algorithm for the optimisation, [5], with respect to the three parameters $\tilde{a}, b$ and c. For more accurate results the gradient of (2.32) with respect to these parameters is also specified. When running the optimisation algorithm the initial values for the parameters $\tilde{a}$ and b is set to the respective parameters found by the POT method, while the initial value for c is set as 1, which is the value for the asymptotic case. It has been observed that the initial values do not influence the estimated parameters as long as they are set with some caution.

[1] Numerical Algorithms Group

2.3.1 Confidence intervals for return levels

By inverting the exceedance rate function, with the parameters for the minimisation of the square error function (2.32), the estimated return level $\hat{\eta}_m$ for η_m has been found. Further we are going the estimate the 95% confidence intervals for this return level. This confidence intervals will be estimated by finding other curves of the form as the exceedance rate with other parameters $\tilde{a}, b, c, q$ and ξ. By inverting these functions in the same way the estimated exceedance rate function was, these curves will represent other estimates for the return level. Restricting to only those curves that falls within the bounds of the confidence intervals defined by (2.26), the maximum and minimum return levels estimate provided by these curves will provide an 95% confidence interval for η_m.

So to estimate the confidence intervals it will be necessary to find the curves $\hat{\epsilon}^{min}(\eta)$ and $\hat{\epsilon}^{max}(\eta)$, which provides the lower and upper bound of the return level confidence interval respectively. To find these curves, the parameters found by the optimisation will be slowly varied, keeping only the parameters for the curves which falls within the confidence intervals. Since the confidence intervals for $\hat{\epsilon}_k$ found by (2.26) will not be smooth, the confidence intervals will be approximated with

$$\text{CI} \approx (\hat{q}_k(1-\hat{\tilde{a}}_k(\eta-\hat{b}_k)^{\hat{c_k}})^{\hat{\gamma}} - 1.96\hat{s}_k(\eta)/\sqrt{R}, \hat{q}_k(1-\hat{\tilde{a}}_k(\eta-\hat{b}_k)^{\hat{c_k}})^{\hat{\gamma}} + 1.96\hat{s}_k(\eta)/\sqrt{R}). \quad (2.36)$$

These will be smooth curves which approximates the confidence intervals. The curves which falls within these confidence interval is then found by varying the parameters a certain percentage, which is an input for the routine, from the parameters found by optimisation routine. By this variation of the parameters $11^5 - 1$ curves are fitted, and the curves, which falls within these confidence interval, providing the maximum and minimum return level estimates, $\hat{\epsilon}^{max}$ and $\hat{\epsilon}^{min}$ respectively, is kept and provides an estimate for the 95 % confidence interval for the return level.

2.4 Heavy tailed distributions

A heavy tailed distribution is a distribution with tail that are not exponentially bound, i.e. a distribution with heavier tails than the exponential distribution. The definition of a heavy (right) tailed distribution is, as stated in [2],

$$\lim_{x\to\infty} e^{\lambda x}\Pr(X > x) = \infty \quad \forall \quad \lambda > 0. \quad (2.37)$$

This means that a distribution has a heavy (right) tail if it has a tail which is heavier than the ones for the exponential distribution.

2.5 AR-GARCH model

An AR(k)-GARCH(p,q) is an autoregressive model of order k with GARCH noise of order p, q, that is a model of the form

$$\begin{aligned} r_t &= a_0 + \sum_{i=1}^{k} a_i r_{t-i} + \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2, \end{aligned} \tag{2.38}$$

where $\epsilon_t = \sigma_t z_t$ and $z_t \sim \text{IID}(0, 1)$. The distribution of the z_t's will here be either Gaussian or Student's t distribution with ν degrees of freedom, and scaled to variance 1. When fitting the AR-GARCH model to a data set it is possible to estimate conditional tail quantiles. As shown in [3], the tail quantiles can be estimated by assuming either Gaussian or t-distribution, and multiplying the estimates of conditional volatility σ_t with the quantiles in the chosen distribution and adding the conditional mean, $a_0 + \sum_{i=1}^{k} a_i r_{t-i}$ for each t.

Chapter 3

Data

In this chapter the data, which later will be analysed, is introduced. We will look at both synthetic time series and real world data sets.

3.1 Synthetic data

To investigate the performance of the ACER method with the Fréchet assumption against the performance of the common asymptotic methods we are going simulate synthetic time series and applying both methods to the same data sets. This will give an indication of prediction accuracy in terms of both return level estimates and most important the length of the estimated confidence intervals for these return levels.

3.1.1 Pareto distribution

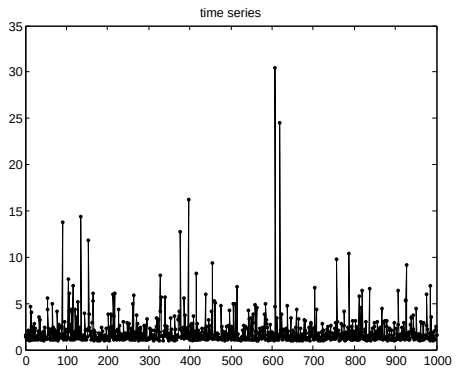
For the first time series we are using independent realisations from a Pareto distribution. This gives a time series $\{X_t\}$ where each X_t is an independent realisation from the Pareto distribution, defined by the CDF

$$F_X(x) = \begin{cases} 1 - 1/x^\alpha & \text{for } x \geq 0 \\ 0 & \text{for } x < 0, \end{cases} \tag{3.1}$$

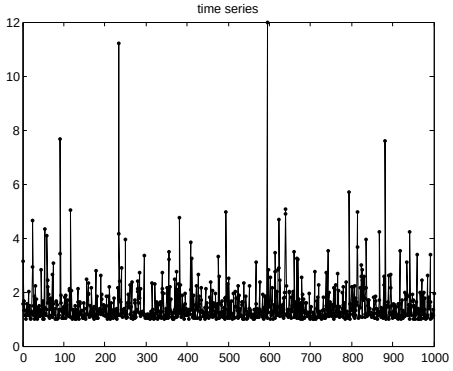
where α is a positive parameter. Small values for α will produce very extreme observations, so to begin with only values between 2 and 5 will be considered for α. The Pareto distribution is by the definition stated in (2.37) a heavy tailed distribution. To generate synthetic time series from this distribution we are going to use the inverse transformation rule. So we have if $U \sim \text{Unif}[0, 1]$ and

$$X = \frac{1}{U^{1/\alpha}}, \tag{3.2}$$

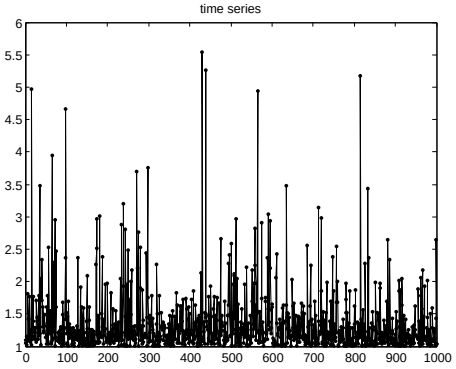
then $X \sim \text{Pareto}(\alpha)$. Samples, for different values of α, is plotted in 3.1(a)-3.1(d). From this plots it is clear that it is for the small values of the parameter α we observe the most extreme deviations from the mean, which for a Pareto distribution is $\alpha/(\alpha - 1)$ when $\alpha > 1$ (The mean does not exist for $\alpha \in (0, 1]$).



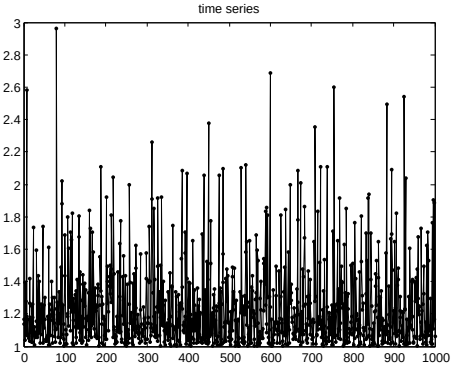
(a) Time series generated from the Pareto distribution with $\alpha = 2$.



(b) Time series generated from the Pareto distribution with $\alpha = 3$.



(c) Time series generated from the Pareto distribution with $\alpha = 4$.



(d) Time series generated from the Pareto distribution with $\alpha = 5$.

Figure 3.1: Time series generated for the Pareto distribution.

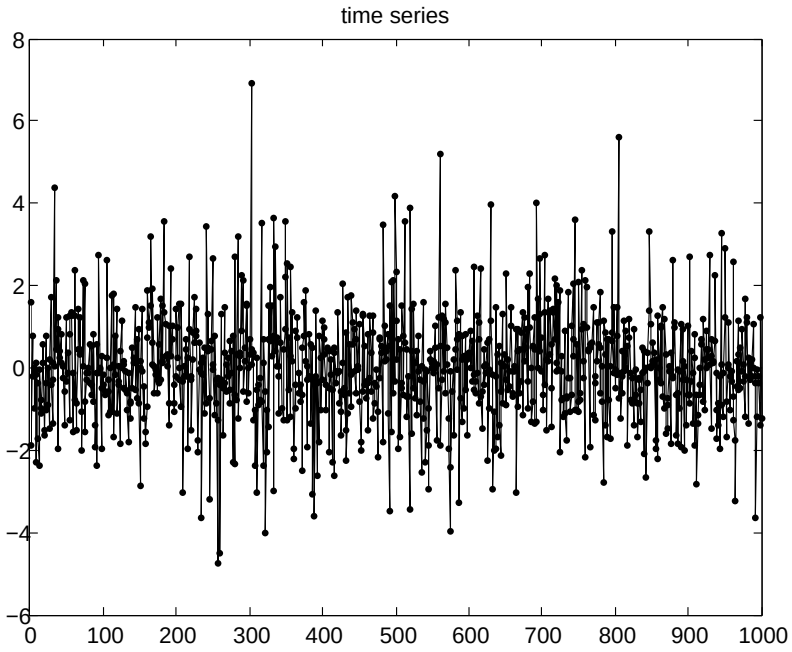


Figure 3.2: First 1000 observations from the student's t-model with $\nu = 4$.

3.1.2 Student's t-distribution

Another common heavy tailed distribution is the Student's t-distribution with ν degrees of freedom. This distribution is defined by the pdf

$$f_X(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}. \tag{3.3}$$

The time series $\{X_t\}$ will be constructed as

$$X_t \overset{\text{iid}}{\sim} t(\nu).$$

The first 1000 observations from a time series, here with degrees of freedom $\nu = 4$, can be seen in figure 3.2.

3.1.3 ARCH/GARCH models

An ARCH[1](p) model is a model where the error term Z_t is assumed to be $Z_t = \sigma_t e_t$, where $e_t \sim \text{IID}(0,1)$ and σ_t is given by the recursion equation

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2, \tag{3.4}$$

where $\alpha_0 > 0$ and $\alpha_i \geq 0$. A GARCH(p,q), generalised ARCH, model is an ARCH model for which an ARMA model is assumed for the error variance. Hence, the error variance term is

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{3.5}$$

with $\alpha_0 > 0$ and $\alpha_i, \beta_j \geq 0$. The distribution of $\{Z_t\}$ is

$$Z_t = \sigma_t e_t, \tag{3.6}$$

where $e_t \sim \text{IID}(0,1)$. In this section it is used that

$$\sqrt{\frac{\nu}{\nu-2}} e_t \sim t_\nu, \ \nu > 2, \tag{3.7}$$

where t_ν is a Student's t-distribution with ν degrees of freedom. The distribution of e_t in (3.7) is scaled to make the variance of e_t equal to 1. The Student's t-distribution is used here over the Gaussian distribution because it is a heavy tailed distribution. It is also found that the GARCH models can be better fitted to heavy tailed financial data when using the Student's t-distribution over the Gaussian distribution [6]. In figure 3.3 1000 observations from a GARCH(2,2) process, with $\alpha = [0.5, 0.3, 0.067]^T$ and $\beta = [0.5, 0.1]^T$, is observed. Also, the underlying Student's t-distribution with $\nu = 8$ degrees of freedom.

3.2 Data series

In this section we are going to look at a couple of real data sets, first electric market spot price data from the Nordic Power Exchange and then closing index of the Dow Jones stock exchange.

3.2.1 Electric market spot price data

The Nordic Power Exchange, Nord Pool, data is electric market spot price data. These data is provided by Sjur Westgaard. The data consists of hourly spot price values for a period of ten years, with a time series consisting of the daily ith hour measurement. In

[1] ARCH is an abbreviation of autoregressive conditional heteroskedasticity

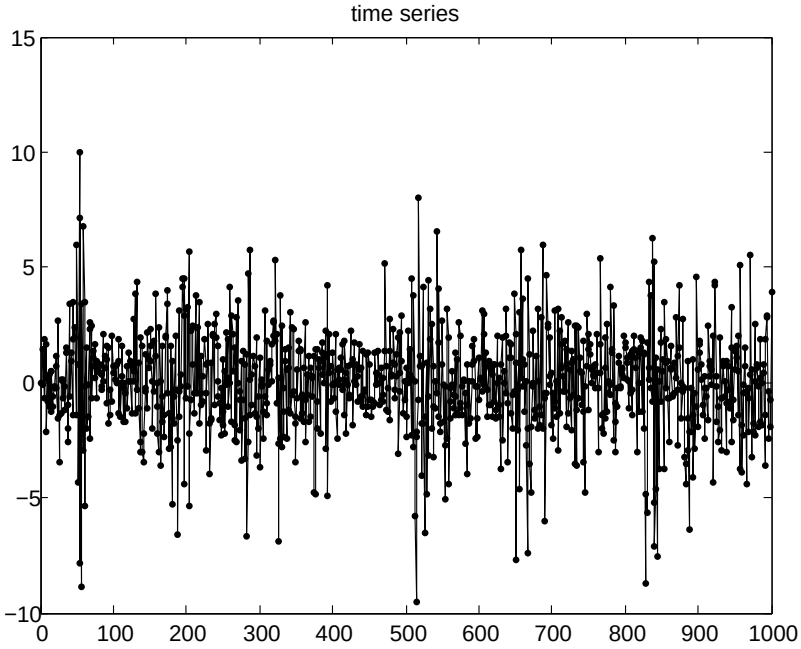


Figure 3.3: Time series generated from a GARCH process.

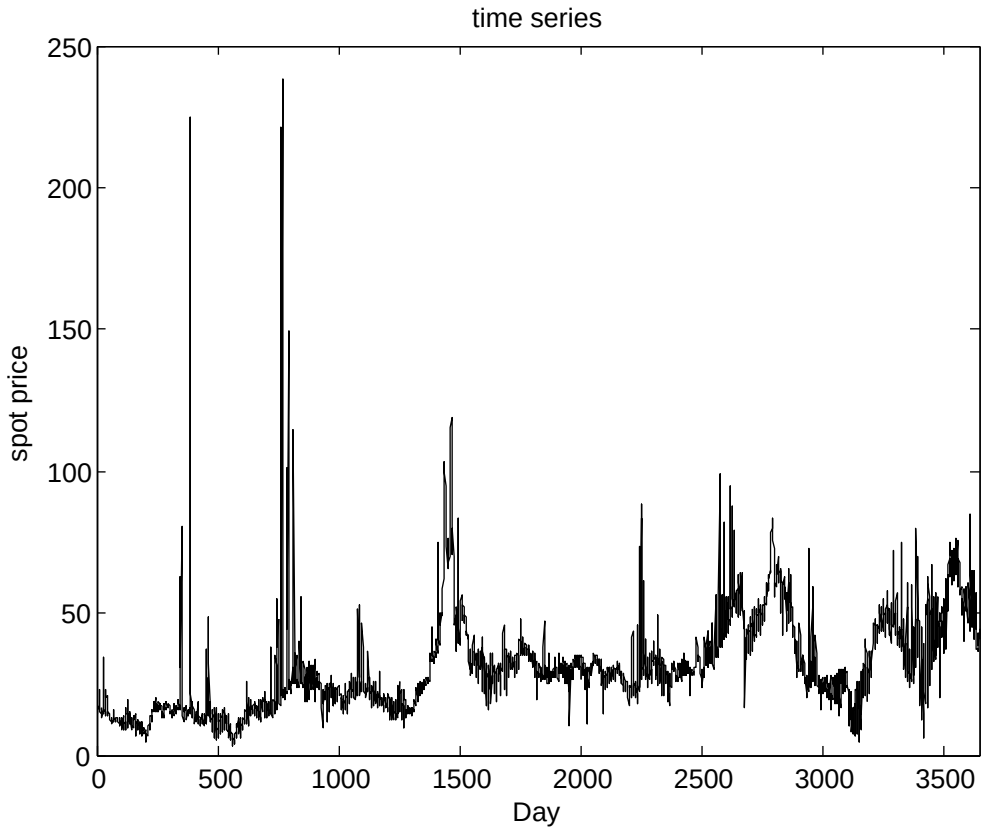


Figure 3.4: Spot price for the 9th hour for the then year period.

table 3.4 the spot price for the 9th hour for this ten year period is plotted. The data will be transformed to log-daily return rates, defined by

$$r_t = \log\Big(\frac{x_t}{x_{t-1}}\Big), \tag{3.8}$$

where $\{x_t\}$ is the original time series and $\{r_t\}$ is the transformed time series. In figure 3.5 the log-daily returns is plotted for the 9th hour for the ten year period.

3.2.2 Dow Jones Index

The Dow Jones closing index data consist of the closing index for this stock exchange for the time period 1st of January, 1980 to 29th of December, 2000. These data will, as with the spot price data, also be transformed to log-daily returns. In figure 3.6 the transformed data is plotted. From the plotted data we observe that there are one observation, or increase in the index, which is far greater than all other changes in the index. For the ACER method this should not produce any problems since the method

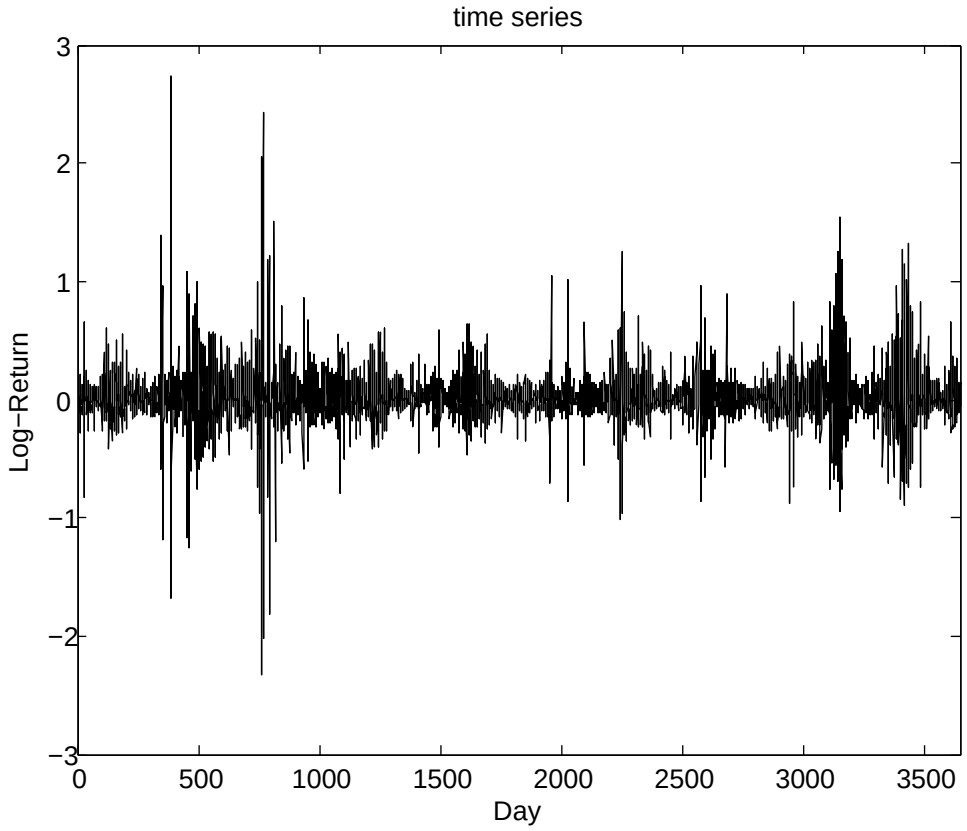


Figure 3.5: Log-daily returns for the 9th hour for the ten year period.

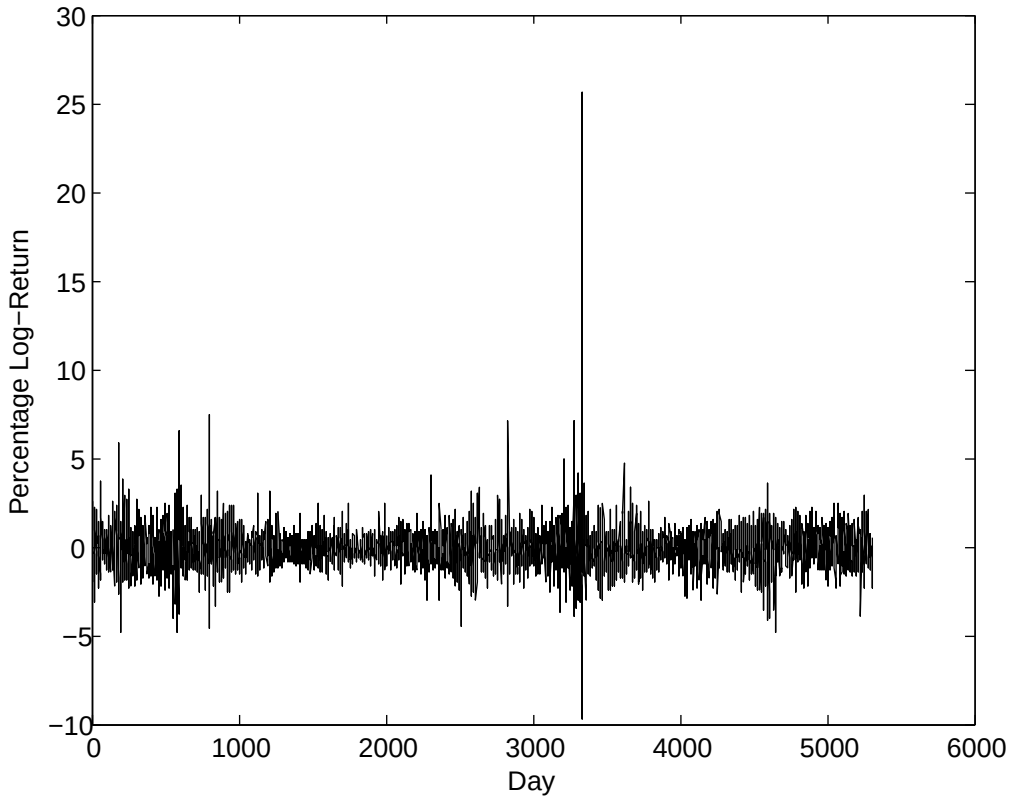


Figure 3.6: Log-daily returns for the Dow Jones index.

neglects observations when the confidence interval for the empirical ACER function is too great.

Chapter 4

Analysis of data and results

In this chapter the data presented in chapter 3 will be analysed. Both the POT method and the ACER method where the underlying asymptotic extreme value distribution is assumed to be a Fréchet distribution. We will use both method to predict return levels, and the confidence intervals for these return levels.

4.1 Analysis of synthetic data

We will begin with the analysis of the synthetic time series introduced in 3.1. The analysis of these time series will let us use the ACER method on time series with know properties, such as exceedance rates and return levels, and compare the result to, where these are available, analytical results and the estimates provided by the asymptotic POT method.

4.1.1 Pareto distribution

First we are going to consider the model based on the Pareto distribution introduced in section 3.1.1. To predict return levels are we are to simulate a set of 10 time series, each consisting of 3650 realisations from a Pareto distribution. Since all realisations are independently drawn from a Pareto distribution there will be no need to use higher level ACER functions, and for the POT method the assumption of independent realisations is fulfilled, and therefore there are no need to decluster the data. The time series analysed will be generated from the Pareto distribution with the value for the parameter $\alpha = 2$ and $\alpha = 3$.

For a set of these time series generated from the Pareto, here with parameter $\alpha = 2$, the logarithm of the ACER function is plotted in figure 4.1. For the same model, only with $\alpha = 3$, the logarithm of the ACER function is plotted in figure 4.2.The parameters in (2.30) is then found by minimising (2.32) for the two different cases. For the Pareto distribution the exceedance rate of level η is easily derived as

$$P(X > \eta) = 1 - F_X(\eta) = 1 - (1 - \frac{1}{\eta^\alpha}) = \eta^{-\alpha}, \tag{4.1}$$

since all observations are independent. From this it is possible to plot both the extrapolated ACER function with confidence intervals and the actual exceedance rate for the Pareto distribution in the same plot, hence observing the actual fit of the model. In figure 4.3 and 4.4 the extrapolated ACER function with confidence intervals, the empirical ACER function and the analytical exceedance rate for the Pareto time series with parameter $\alpha = 2$ and $\alpha = 3$ respectively is plotted. From these plots we observe that the extrapolated ACER function is a good approximation to the analytical exceedance rate, and the analytical exceedance rate falls within the 95% confidence interval for the ACER function for all values of the exceedance. To compare the ACER method to the asymptotic POT method we need to fit a GPD to the exceedances. Though efore fitting a GPD to the data it should be noted that if $X \sim \text{Pareto}(\alpha)$, then the distribution of conditial threshold exceedances is

$$\begin{aligned} \text{Prob}(X > u + y | X > u) &= \frac{1 - F(u+y)}{1 - F(u)} = \frac{1/(u+y)^{\alpha}}{1/(u)^{\alpha}} \\ &= \left(\frac{u+y}{u}\right)^{-\alpha} = \left(1 + \frac{y}{u}\right)^{-\alpha}, \end{aligned} \tag{4.2}$$

which is the generalised Pareto distribution with shape parameter $\xi = 1/\alpha$ and scale parameter $\tilde{\sigma} = u/\alpha$. This is an exact result, and holds for all thresholds $u \geq 1$. Hence there is no asymptotic argument in the derivation of the distribution of the threshold exceedances. This means that the threshold can be set to $u = 1$ without any inspection of mrl plots or plots of the modified scale and shape parameters against the threshold, which in turn leads to the POT method being able to use all of the data in the time series. In figure 4.5 and 4.6 both the extrapolated ACER function and the exceedance rate for the GPD, for the Pareto model with $\alpha = 2$ and $\alpha = 3$ respectively, with 95% confidence interval is plotted. We observe from these plots that both methods produce results close to each other, but the acurracy, in terms of confidence interval width, is far greater for the POT method in both cases. This means that even though both methods are able to actually use all of the data in the time seires, the POT method performes much better the ACER method. This may be because the exceedances for the Pareto distribution is disitributed by GPD with no asymptotic argument, and for the ACER method it is necessary to estimate the extra parameters for the exceedance rate function, c and q for which the value of both should be, in this case, 1.

Now we are going to predict the 100 time series return level, that is the level which is expected to be exceeded once every $3650 \cdot 100$ observation. This will only be done for the Pareto model with $\alpha = 3$. For the Pareto model with $\alpha = 3$ it is easy to show that the value expected to be exceeded once for every 100 time series is

$$P(X > \eta^{100ts}) = 1/(\eta^{100ts})^3 = 1/(3650 \cdot 100) \Rightarrow \eta^{100ts} = 365000^{\frac{1}{3}} = 71.47. \tag{4.3}$$

The two methods will then be used in order to predict this value. Estimation of the 100 time series return level with confidence intervals with the two methods for several sets of time series can be observed in table 4.1. The values in parentheses is percentage deviation from real value. From this it is observed that both in terms of length of

$\eta^{100\text{ts}}_{ACER}$	CI_{ACER}	$\eta^{100\text{ts}}_{POT}$	CI_{POT}
76.27(6.7%)	[61.11, 94.91]	69.83(2.3%)	[61.15, 78.51]
65.93(7.8%)	[57.98, 79.30]	69.46(2.8%)	[60.76, 78.15]
72.36(1.2%)	[59.51, 84.64]	77.59(8.6%)	[67.70, 87.48]
70.11(1.9%)	[58.45, 81.88]	70.75(1.0%)	[61.88, 79.63]
71.39(0.1%)	[67.98, 86.45]	69.83(2.3%)	[61.15, 78.51]

Table 4.1: 100ts return level estimated with ACER and POT.

	1	2	3	4	5
$\tilde{a}$	-1.08	-0.96	-0.98	-1.03	-0.89
b	1.01	0.99	1.00	1.02	1.03
c	1.02	1.00	0.98	1.04	1.01
γ	-2.83	-3.07	-3.05	-2.86	-3.04
q	1.00	0.99	1.01	0.95	0.97

Table 4.2: Estimated parameters for the sub asymptotic extreme value distribution.

confidence intervals and deviation from real return value, the POT method seem to be the most accurate method. For the ACER method the predicted return level here deviates, on average, 3.52% from the actual return level of this model. For the POT method this average deviation is 3.40%. In terms of the confidence intervals, the length of the 95% confidence intervals for the predicted return level for the ACER method is on average 24.43 units, while for the POT method this figure is only 17.93. This means that the confidence intervals for the return levels predicted for by the ACER method is on average approximatly 36% longer than for the POT method. For the ACER method the estimated parameters is presented in table 4.2. Since there are no asymptotic argument in the derivation of the threshold exceedances the parameters c and q from the SGEV distribution should ideally be close to 1. From table 4.2 it is observed that the parameter c is for all five data sets close to the real value. The parameter q is also close to 1 for all of the five data sets. The estimates for the shape parameter $\gamma = -1/\xi$ is all close to the real parameter, which is, as seen from (4.2), $\gamma = -3$.

4.1.2 Student's t-distribution

Secondly are we going to consider the Student's t-model introduced in section 3.1.2. We will also here look at data sets consisting 10 time series each with 3650 realisations. Since all observations are independent the first ACER function will used. In figure 4.7 the extrapolated ACER function, the empirical ACER function and the exceedance rate for the Student's t-model is plotted. From this it seems that the ACER method gives a good fit to the actual exceedance rate, and we see that the exceedance rate for the t-model is inside the confidence bounds for the extrapolated ACER function for all values of the exceedances. When fitting the data to the GPD with the POT method there are no need, due to the independence between observations, to decluster the data, but the

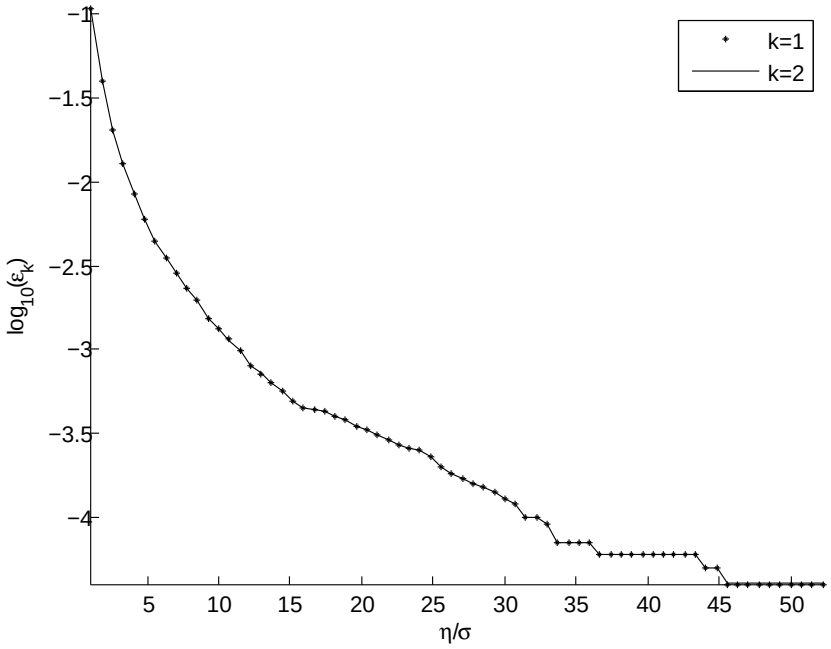


Figure 4.1: Logarithm ACER function for Pareto time series with $\alpha = 2$.

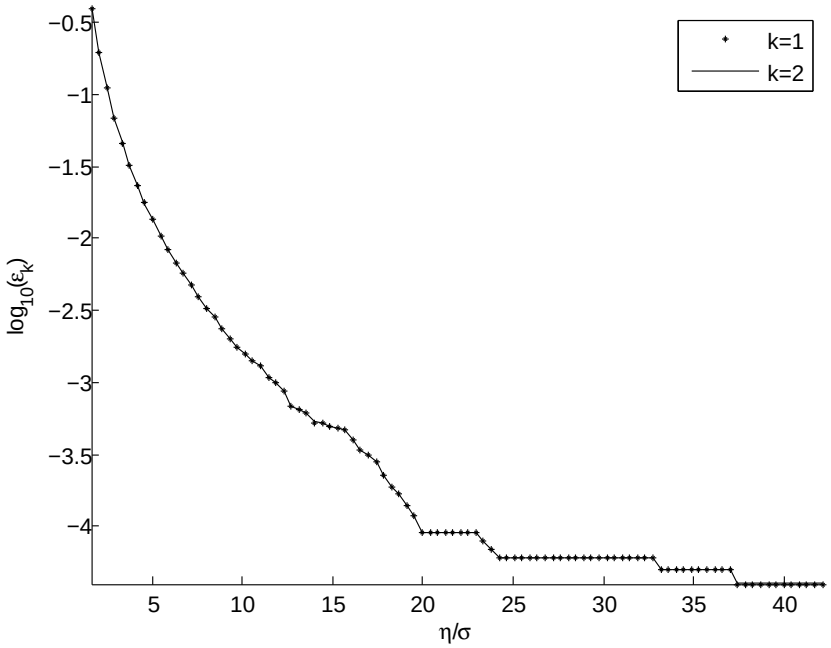


Figure 4.2: Logarithm ACER function for Pareto time series with $\alpha = 3$.

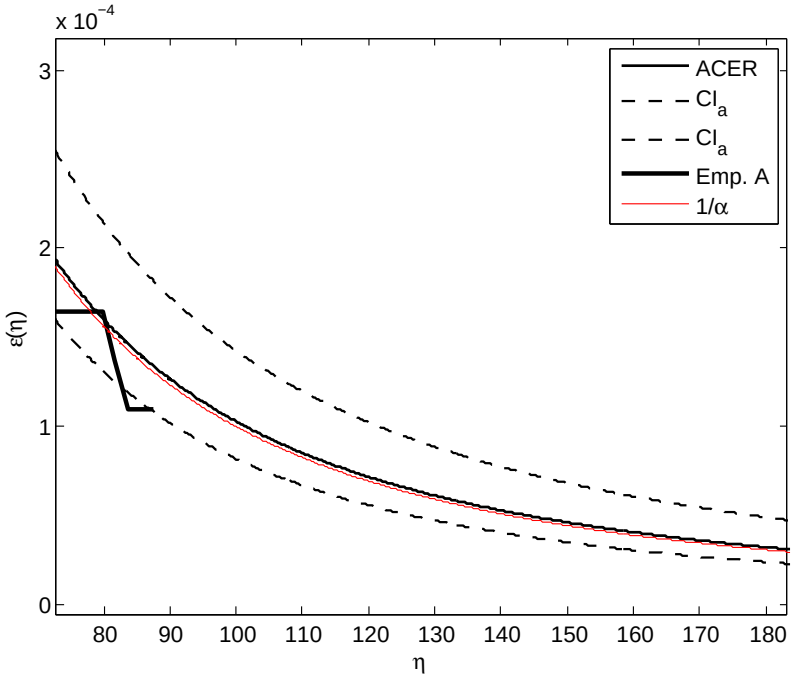


Figure 4.3: Extrapolated ACER function for Pareto time series with $\alpha = 2$.

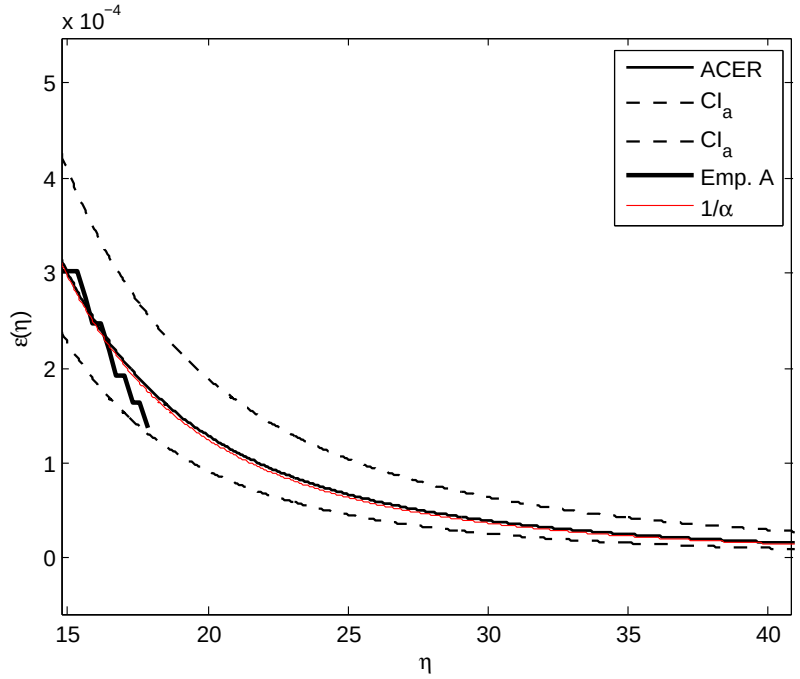


Figure 4.4: Extrapolated ACER function for Pareto time series with $\alpha = 3$.

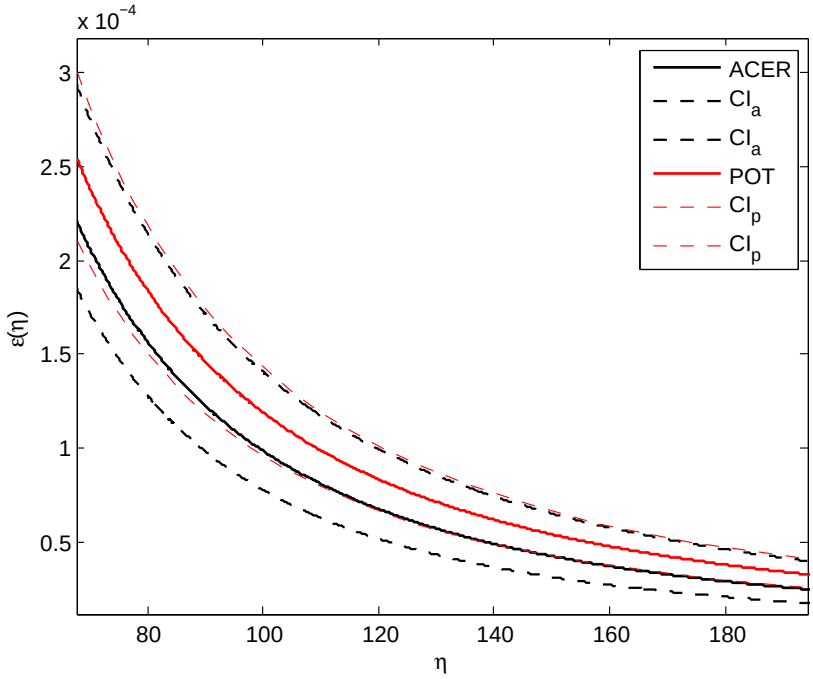


Figure 4.5: Extrapolated ACER function and POT fitted GPD with $\alpha = 2$.

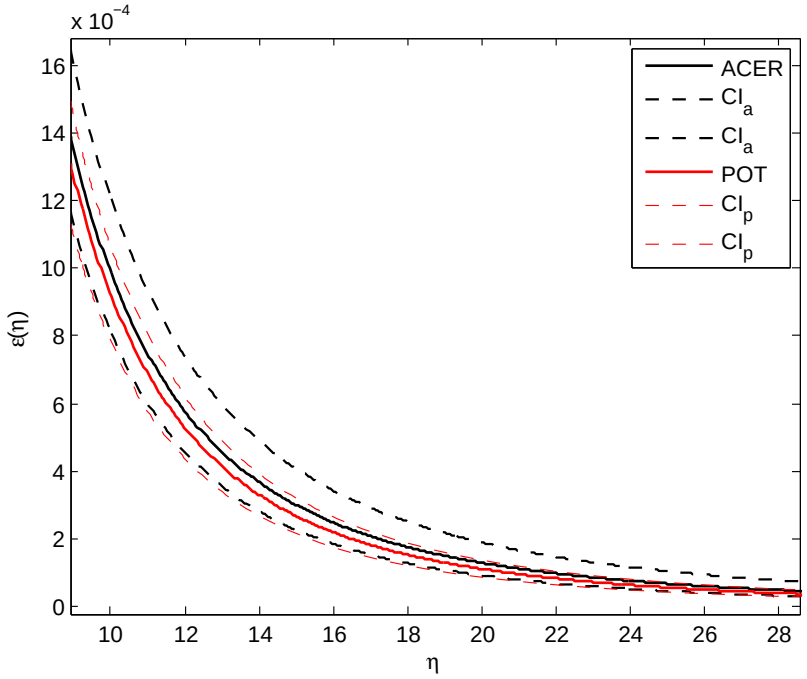


Figure 4.6: Extrapolated ACER function and POT fitted GPD with $\alpha = 3$.

η^{100ts}_{ACER}	CI_{ACER}	η^{100ts}_{POT}	CI_{POT}
32.99(2.1%)	[29.32, 37.80]	33.82(4.7%)	[23.75, 43.88]
27.78(14.0%)	[24.17, 31.59]	24.56(24.0%)	[18.35, 31.05]
31.68(1.9%)	[28.43, 34.72]	30.24(6.4%)	[21.87, 38.61]
29.76(7.7%)	[26.88, 33.31]	31.83(1.5%)	[22.46, 41.20]
31.80(1.5%)	[26.29, 34.25]	29.05(10.1%)	[21.15, 36.96]

Table 4.3: 100ts return level estimated with ACER and POT for the t-model with $\nu = 4$.

threshold will need to be decided. From inspection of the mean residual plot in figure 4.9(a) it is difficult to decide on a right level for the threshold, but from the plot of the modified scale and the shape parameters against the threshold in figure 4.9(b) these parameters seem to stabilise around $u = 2$, so this is the threshold we choose. When the threshold is set to $u = 2$, it is for this specific model expected that only 5.8% of all observations fall above this threshold. For the ACER method, for which in this case the tail marker is $\eta_1 = 0$, it is expected that half of all observations fall in this category. With this huge difference in usage of the available data, it is expected that the length of the estimated confidence intervals should be much greater for the POT method compared to those estimated by ACER. In figure 4.11 the extrapolated ACER function and the exceedance rate for the GPD is plotted with confidence intervals. From this plot it is observed the predicted exceedance rate for both methods is almost the same, but the width of the confidence intervals is far greater for the POT method.

Further we are going to use the two methods to predict the return level associated with the length of 100 time series. The actual return level associated with this period is found to be $\eta^{100ts} = 32.30$. For a total of five sets of time series the 100 time series return level with confidence intervals has been estimated using both methods, with the values found in table 4.3. From this table we observe that for the second data set both methods seriously under estimate the return level, and for both methods the actual return level is not even included in the 95% confidence intervals. For the four other there are not too much difference in the deviance of the estimated return levels compared to the real value, the ACER method seem to predict slightly closer to the real value. Considering the confidence intervals we see the same in this table as in figure 4.11, that the length of the confidence intervals is far greater for the return levels estimated with the POT method compared to those estimated with the ACER method. Considering the confidence intervals for the return levels estimated by the ACER method, the length of these is on average only apporixmatly 56% of the length of the confidence intervals esimated by the POT method. In table 4.4 the estimated parameters for the SGEV distribution is presented. We see that these parameters, as was also the case considering the Pareto model, do not vary much across different data sets.

The next model we will consider is another Student's t-model, this time with $\nu = 3$ degrees of freedom. As for the first Student's t-model the first ACER function will be the one analysed, and there will be no need for declustering of the data when fitting the GPD with the POT method. In figure 4.8 the extrapolated ACER function, the

	1	2	3	4	5
$\tilde{a}$	-0.41	-0.36	-0.39	-0.38	-0.39
b	0.01	0.00	0.00	0.11	0.01
c	1.49	1.50	1.46	1.46	1.48
γ	-2.77	-3.01	-2.93	-2.98	-2.88
q	0.48	0.47	0.49	0.43	0.47

Table 4.4: Estimated parameters for the sub asymptotic extreme value distribution for the t-model with $\nu = 4$.

empirical ACER function and the exceedance rate for this model is plotted. From this we see that the ACER method seem to predict close to the real exceedance rate. The real exceedance rate also fall inside the 95% confidence interval for the ACER function for the all values of the exceedance leve. When deciding upon a level for the threshold we star by inspecting the mean residual life plot in figure 4.10(a). Again it is very difficult to decide a threshold from this plot, so in figure 4.10(b) the modified scale and shape parameters is plotted against the threshold. Considering these plots the threshold is set at $u = 3$. While the ACER method is able to use all positive observations, the POT method will on an average only use about 2.9% of the data in this model when $u = 3$. In figure 4.12 both the extrapolated ACER function and the exceedance rate of the GPD is plotted. Again the ACER method seem to produce more accurate estimates in terms of confidence interval width.

At last we are going to estimate the return level for associated with the length of 100 time series. For the Student's t-model with $\nu = 3$ degrees of freedom the return level considered is $\nu^{100\text{ts}} = 73.82$. Estimated return levels and estimated confidence intervals for this return level is presented for both methods in table 4.5, with the percentage deviance from the actual return level in parenthesis. As we observed earlier in figure 4.12 both methods predict the actual return level, in terms of percentage deviance, rather well (except the POT method for the last time series). Also, we again observe that the length of the confidence intervals is far greater for the POT method than for the ACER method. This is again, at least partially, because of the asymptotic argument in the derivation of the POT method. In table 4.6 the parameters for the extrapolated ACER function is presented. We see from this that the parameters seem to vary a lot more for this model than for both the Pareto model with $\alpha = 3$ and for the Student's model with $\alpha = 4$. Still, as for the Student's t-model with $\nu = 4$ degrees of freedom, the sub asymptotic parameters c and q significantly differs from 1, which is the asymptotic case.

4.1.3 ARCH/GARCH model

First in this section we are going to look at a simple ARCH(2) model where Z_t is Student's t distributed with $\nu = 4$ degrees of freedom and scaled to variance 1 and $\alpha = [0.5, 0.2, 0.1]^T$. After inspection of the five first ACER function we conclude that there is no need to use a higher order ACER method for this model, and there are also no need for declusting for the POT method. For the ACER method the tail marker

$\eta^{100\text{ts}}_{ACER}$	CI_{ACER}	$\eta^{100\text{ts}}_{POT}$	CI_{POT}
70.68(4.25%)	[56.78, 91.17]	66.24(10.27%)	[36.50, 95.97]
71.87(2.64%)	[57.74, 93.38]	74.85(1.40%)	[39.76, 109.91]
69.72(5.55%)	[56.41, 82.91]	74.07(0.34%)	[41.37, 106.77]
68.83(6.76%)	[58.01, 87.19]	65.93(10.69%)	[36.06, 95.80]
67.84(8.10%)	[51.50, 86.62]	59.05(20.01%)	[33.89, 84.21]

Table 4.5: 100ts return level estimated with ACER and POT for the t-model with $\nu = 3$.

	1	2	3	4	5
$\tilde{a}$	-0.52	-0.43	-0.51	-0.55	-0.55
b	0.01	1.44	0.54	0.00	0.00
c	1.83	1.10	1.24	1.57	1.74
γ	-1.66	-2.76	-2.53	-1.98	-1.77
q	0.39	0.12	0.31	0.47	0.43

Table 4.6: Estimated parameters for the sub asymptotic extreme value distribution for the t-model with $\nu = 3$.

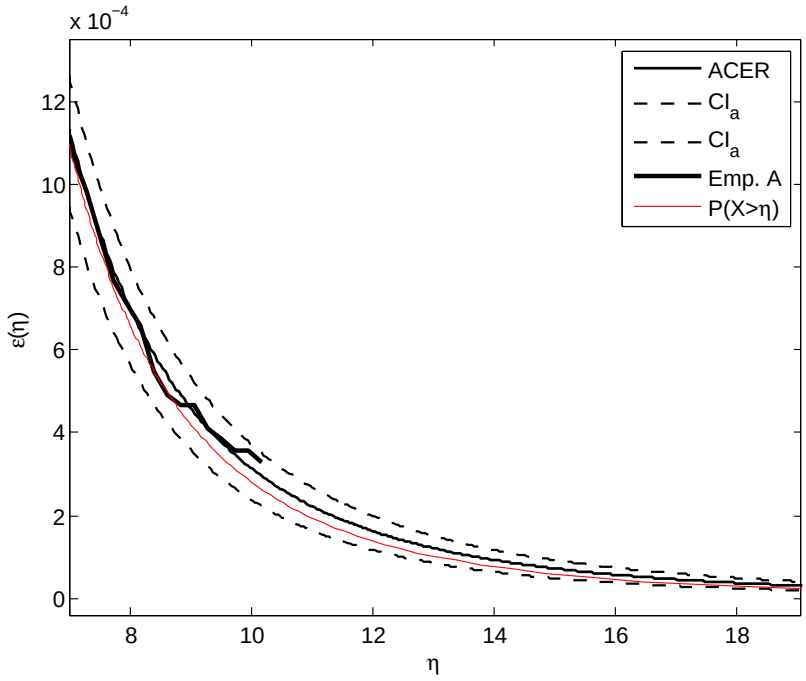


Figure 4.7: Extrapolated ACER function and exceedance rate for the t-model with $\nu = 4$.

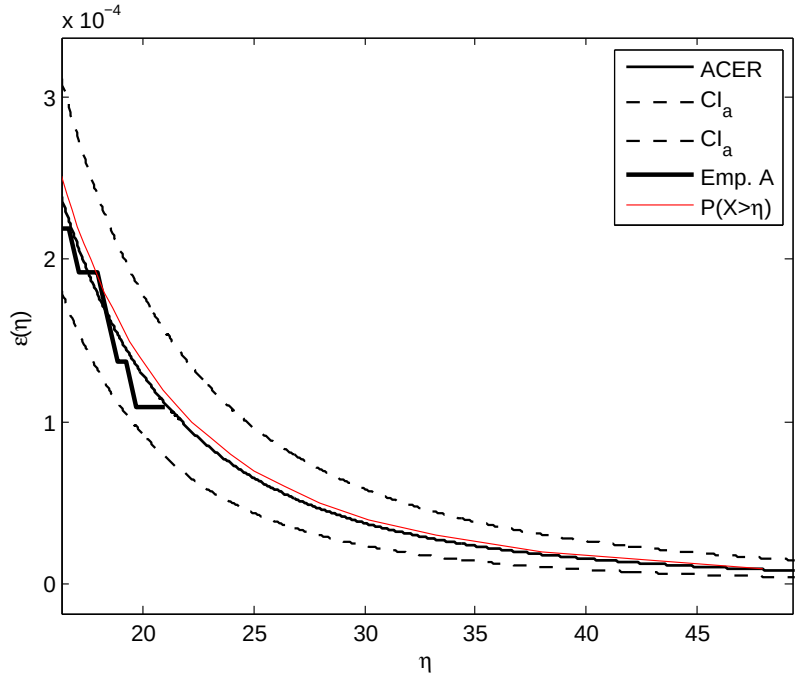
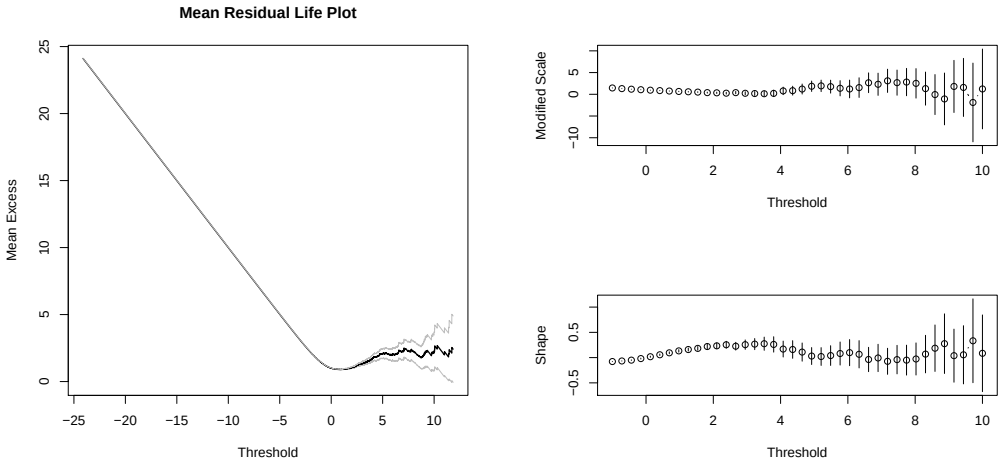


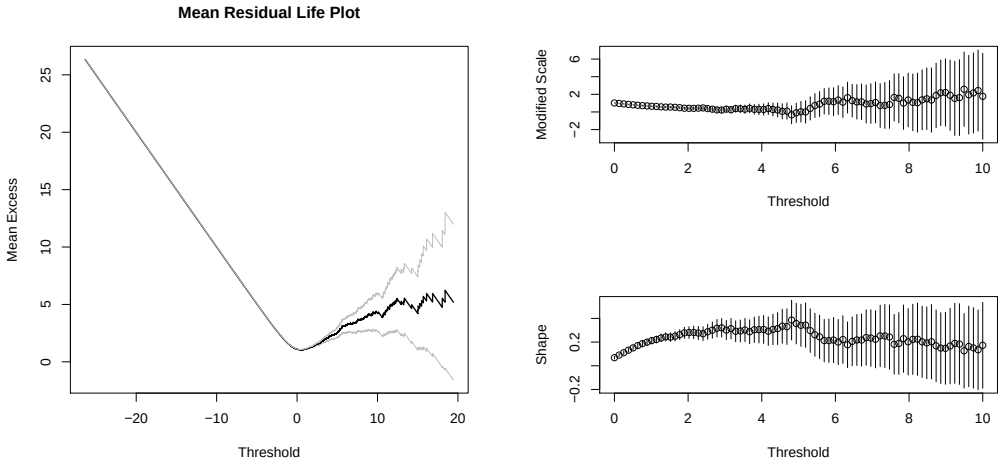
Figure 4.8: Extrapolated ACER function and exceedance rate for the t-model with $\nu = 3$.



(a) Mean residual life plot for the Student's t-model with $\nu = 4$.

(b) Shape and modified scale parameters plotted against the threshold for the Student's t-model with $\nu = 4$.

Figure 4.9: Plots for deciding upon the threshold for the Student's t-model with $\nu = 4$.



(a) Mean residual life plot for the Student's t-model with $\nu = 3$.

(b) Shape and modified scale parameters plotted against the threshold for the Student's t-model with $\nu = 3$.

Figure 4.10: Plots for deciding upon the threshold for the Student's t-model with $\nu = 3$.

is set at $\eta_1 = 0$, and for the POT method the threshold is set, after inspection of the mean residual life plot and a plot of the modified scale and shape parameter against the threshold, to $u = 2$. Both methods then have been applied to five sets of time series, each consisting of 10 time series with 3650 observations each, and used to try and predict the 100 time series return level. In table 4.7 the estimated return levels along with the estimated 95% confidence intervals for these return levels is presented, and in table 4.8 the estimated parameters for the SGEV distribution for the five data sets are available. From the return levels we see that both methods again predict return levels rather close to each other, though there seem to be greater variability in the return levels predicted by the POT method. As for the confidence intervals we again see that the ACER method performes with much more accuracy that the asymptotic method. For illustration purposes it can be mentioned that for the for the first data set 49.8% of the observations is used by the ACER method (Half of the observations is expected to be positive since $Z_t \sim \text{IID}(0, 1)$), while only 1.5% of the observations falls above the threshold selected for use with the POT method. This will of course influence the length of the confidence interval for the return levels, at least when predicting such far out in the tail of the distribution as done here. As for the parameters of the extrapolated ACER function there does not seem to vary a lot between the data sets.

In this section we are going to look at several GARCH models as introduced earlier in section 3.1.3. The first GARCH model analysed is a GARCH(1,1) model with parameters $\alpha = [0.5, 0.3]^T$ and $\beta = 0.3$ and the Student's t-distribution with $\nu = 4$ degrees of freedom. For this analysis only the upper tail of the distribution is considered. As used

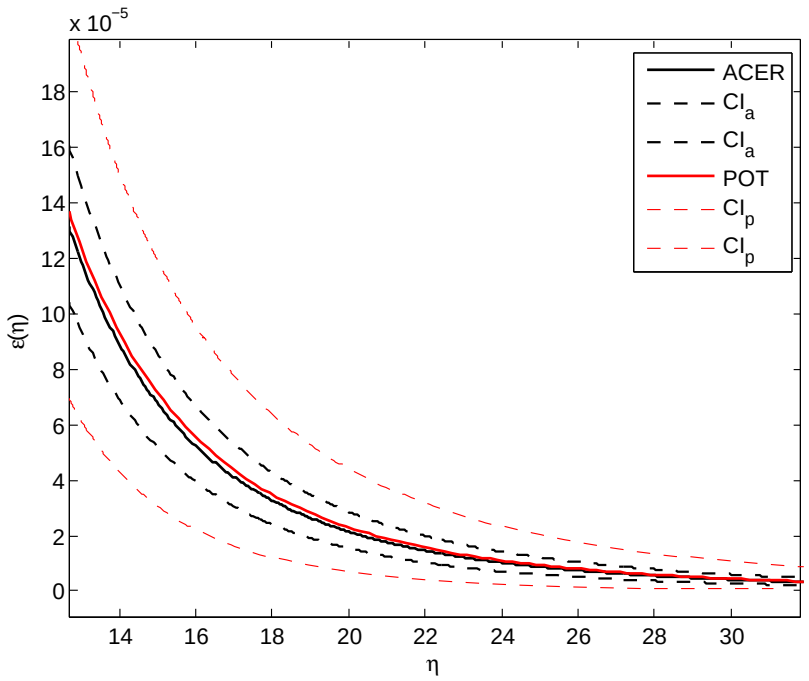


Figure 4.11: Extrapolated ACER function and POT fitted GPD with confidence intervals for the Student's t-model with $\nu = 4$.

$\eta^{100\text{ts}}_{ACER}$	CI_{ACER}	$\eta^{100\text{ts}}_{POT}$	CI_{POT}
18.88	[16.71, 21.08]	15.46	[9.09, 21.82]
19.76	[17.75, 23.22]	19.27	[9.54, 29.00]
19.57	[17.08, 21.65]	17.64	[10.14, 25.14]
20.62	[17.87, 23.49]	24.04	[12.47, 35.62]
22.66	[17.99, 27.10]	24.58	[10.11, 39.07]

Table 4.7: 100ts return level estimated with ACER and POT for the ARCH(2) model.

	1	2	3	4	5
$\tilde{a}$	-0.94	-1.06	-0.98	-0.85	-1.07
b	0.01	0.00	0.01	0.21	0.00
c	1.49	1.50	1.46	1.24	1.49
γ	-2.77	-2.65	-2.79	-3.27	-2.54
q	0.46	0.49	0.48	0.36	0.47

Table 4.8: Estimated parameters for the sub asymptotic extreme value distribution for the first ARCH(2) model.

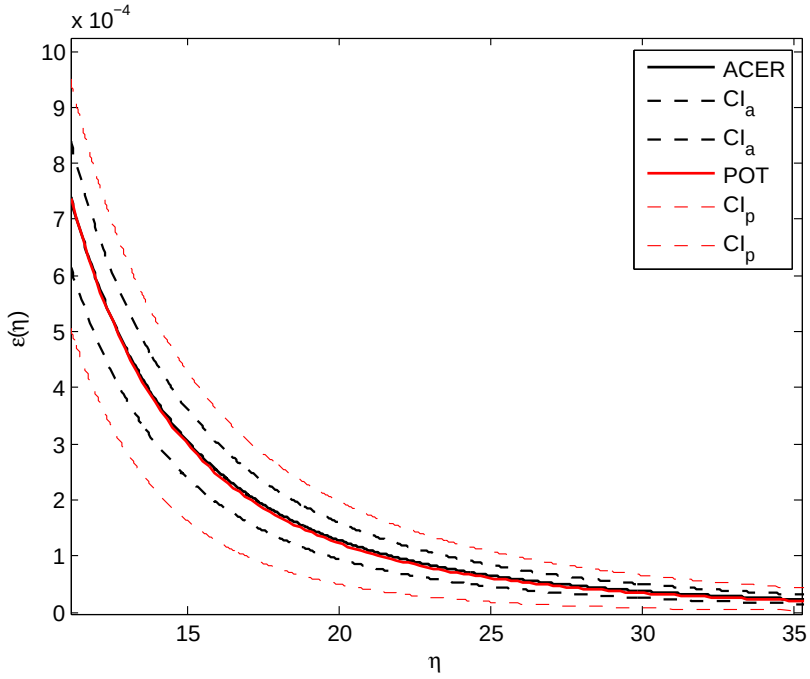


Figure 4.12: Extrapolated ACER function and POT fitted GPD with confidence intervals for the Student's t-model with $\nu = 3$.

η^{100ts}_{ACER}	CI_{ACER}	η^{100ts}_{POT}	CI_{POT}
41.60	[34.99, 51.00]	36.33	[20.48, 52.19]
36.04	[28.39, 48.41]	33.58	[19.74, 47.41]
43.82	[35.25, 52.85]	38.98	[21.15, 56.81]
37.37	[29.74, 46.26]	33.75	[18.74, 48.76]
45.19	[37.21, 53.28]	44.66	[22.39, 65.94]

Table 4.9: 100ts return level estimated with ACER and POT for the first GARCH model.

earlier the data sets will consist of 10 time series, each with 3650 observations. For the models considered earlier all observations have been independent, this is not the case now. In figure 4.13 the four first ACER functions is plotted against scaled exceedance. From this plot it is shown that the second ACER function coincides with the higher order ACER functions in the tail, which is where we are interested in the exceedances. Hence, the second ACER function should be used in the estimation of the parameters in the SGEV. For the POT method it is clear the data no longer for fill the assumption of iid observations. This means that it is now necessary to declust the data to ensure that all threshold exceedances are independent. The declusering is preformed with the following empirical rule:

- First observation above the threshold marks the start of a cluster.
- Last observation above the threshold , followed by M observations below the threshold, marks the end of a cluster.
- All data points in the cluster, except the cluster maxima, is discarded.

For this model the declustering threshold is set at 0 and $M = 2$. The declustering is performed by the clust function from the POT R package. After the declustering the selection of threshold is made from figure 4.14(a) and 4.14(b). From these plots the threshold is selected as $u = 2$. Further the two methods are used to predict a 100 time series return rate from five different sets of generated time series. The estimated return levels are presented in table 4.9. It is observed from the table that the estimated return levels does not greatly differ, though the ACER method consistently predict return levels above those predicted by the POT method. Also the accuracy of the prediction in term of confidence interval length is again much greater for the ACER method. In table 4.10 the estimated parameters for the SGEV distribution for the five data sets. We see from this that also here the sub asymptotic parameters c and q differs from 1, which is the asymptotic case. Again the parameters differ slightly between data sets, but not as much as for the second Student's t-model.

The second GARCH model will be a GARCH(2,2) model with the same parameters as used in figure 3.3. That is $\alpha = [0.5, 0.3, 0.067]^T$ and $\beta = [0.5, 0.1]^T$ and the Student's t-distribution with $\nu = 8$ degrees of freedom. Again, as was the case with the first GARCH model, only the upper tail of the distribution will be analysed. In figure 4.15

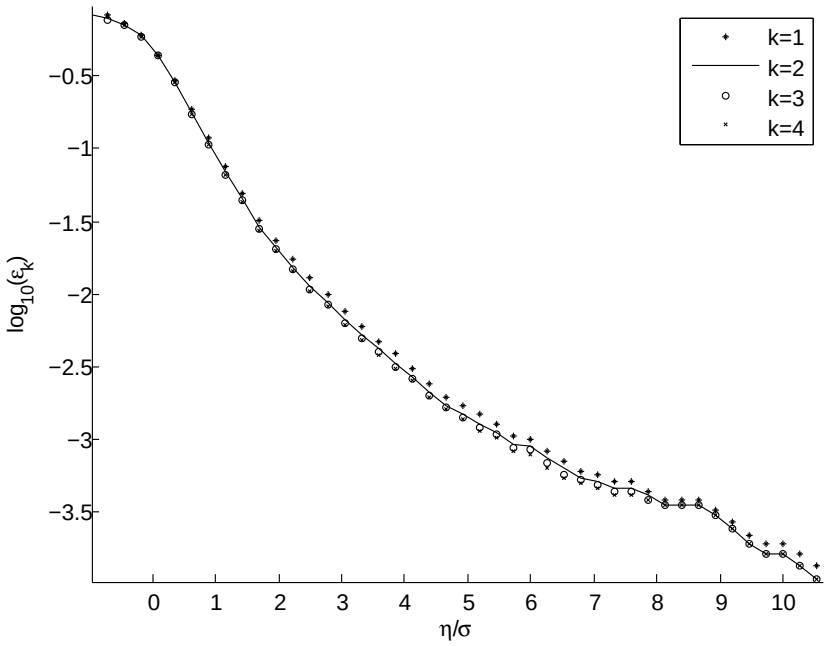
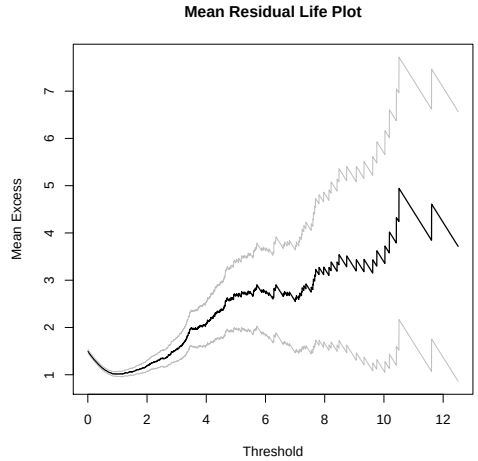


Figure 4.13: Empirical ACER functions plotted against scaled exceedance.

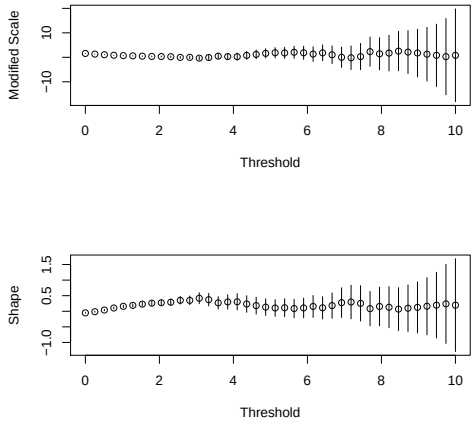
the logarithm of the ACER function is plotted against the scaled exceedance level. From this it is observed that the third ACER function coincides with the higher order ACER functions in the tail, while the lower order ACER functions differ. This leads to the use of the third ACER function when estimating the parameters for the SGEV for this model. For the POT method it will now be necessary to declust the data to ensure independence between observations before fitting. Since it was necessary to use the third ACER function, the data is declustered with the same rule as used for the first GARCH model, but with $M = 3$. To select an appropriate threshold the mrl plot in figure 4.16(a) and the shape and modified scale plot in figure 4.16(b) are considered.

	1	2	3	4	5
$\tilde{a}$	-1.01	-0.79	-1.04	-0.92	-0.95
b	0.00	0.32	0.00	0.00	0.01
c	1.56	1.28	1.65	1.51	1.46
γ	-2.05	-2.68	-1.91	-2.23	-2.19
q	0.46	0.33	0.45	0.48	0.49

Table 4.10: Estimated parameters for the sub asymptotic extreme value distribution for the first GARCH model.



(a) Mean residual life plot for the GARCH(1,1) model.



(b) Shape and modified scale parameters plotted against the threshold for the GARCH(1,1) model.

Figure 4.14: Plots for deciding upon the threshold for the GARCH(1,1) model.

From this the threshold is selected to be $u = 3$. In figure 4.17 both the extrapolated ACER and the exceedance rate of the GPD is plotted with confidence intervals. From this it is observed that while the estimates is fairly equal the accuracy, in terms of the plotted confidence intervals, is far greater for the ACER method. One of the reasons for this is that the POT method in this case only uses 1246 of the observations, or about 3%, while the ACER method is able to use 9249 of the observations, which is about 25% of the observations. As apposed to the Pareto model the return levels of the GARCH model cannot be expressed analytically. It is still interesting to look at the return levels with confidence intervals and, for the SGEV distriubtion, the estimated parameters.

4.2 Analysis of data

In this section the data introduced in section 3.2 will be analysed. As for the synthetic data we will use both the ACER and POT method, and compare their results.

4.2.1 Electricity market spot price data

The log-daily return transformed electric market spot price data introduced earlier will now be analysed. The data analysed will be the 9th hour spot price for the ten year period. As for the GARCH model only the right (upper) tail of the distribution will be analysed. In table 4.11 the yearly maxima of the log-daily return can be observed. We

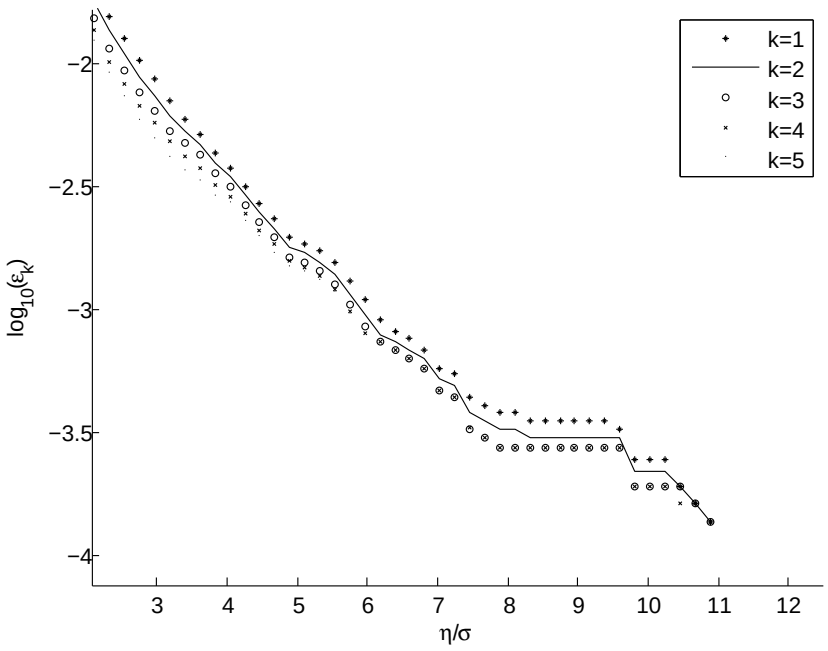
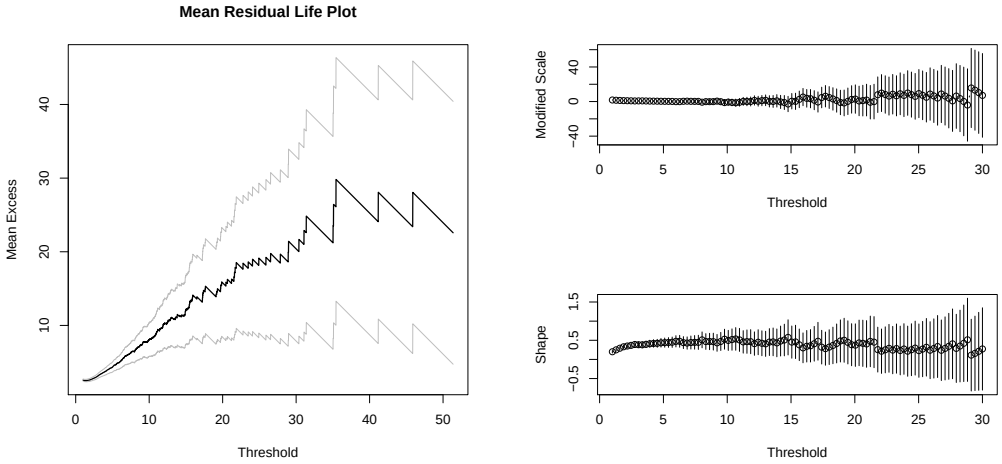


Figure 4.15: Logarithm of ACER functions plotted against scaled exceedance level for the GARCH(2,2) model.



(a) Mean residual life plot for the GARCH(2,2) model.

(b) Shape and modified scale parameters plotted against the threshold for the GARCH(2,2) model.

Figure 4.16: Plots for deciding upon the threshold for the GARCH(2,2) model.

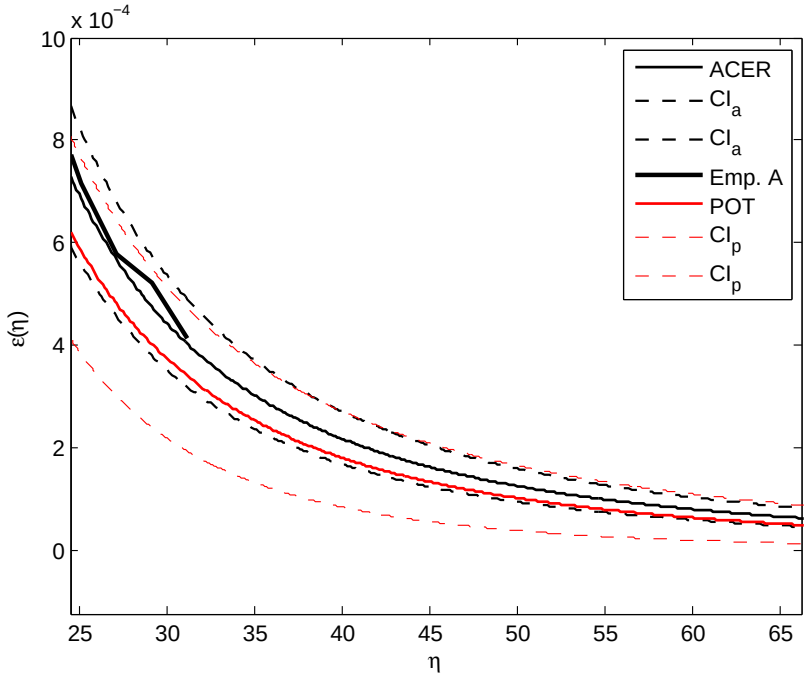


Figure 4.17: Extrapolated ACER function and POT fitted GPD for the GARCH model.

Year	Max increase	Month observed
1	1.392	December
2	2.733	January
3	2.423	February
4	0.6006	June
5	0.6360	June
6	1.056	May
7	1.257	February
8	0.958	January
9	1.545	August
10	1.317	May

Table 4.11: Maximum log-return value each year of the spot price data.

see from this table the two years with by far greatest observed increase in the electricity spot price, year two and three, this increase is found in two of the coldest months of the year, January and February. This is a trend observed through out the data set, making the assumption that all observations come from the same distribution dubious. Even though this trend is observed in the data we are going to try to use both methods to predict extreme values. In figure 4.18 the empirical ACER functions for $k = 1, 2, 4, 5, 6, 7$ is plotted. From the plot of the empirical ACER functions it is observed that it is necessary to use the fifth ACER function when estimating the parameters in (2.32). The tale marker for the ACER method is set, after inspection, to $\eta_1 = 0$. For the POT method the data will be declustered by the same rule as above with $M = 5$. Choice of threshold is made after we have inspected figure 4.19(a) and 4.19(b), which lead us to set the threshold as $u = 0.2$. After declustering of the data only 183 observations fall above the threshold. In figure 4.19 the extrapolated ACER with confidence intervals, the empirical ACER and the exceedances rate for the GPD fitted by POT is plotted. From the plot we observe that the exceedance rate estimated by both the ACER and the POT method is rather close, though the ACER method has even heavier tail than the POT method. We also observe that the exceedance rate estimated by the POT method falls within the confidence intervals for the ACER method for all values of the exceedance level. The confidence intervals for the POT method is not plotted since the scarce amount of data points above the threshold leads to very wide confidence intervals.

When we are modeling the tail of the distribution of the electricity market sport price data with the ACER and POT method in the way we have done above, we do not take into account any seasonality, which is highly present in the data, or any other trends in the data. It is also not possible for the tail quantiles for the distribution to change with time, which is the case for such data. In [3] the POT method is applied to the residuals after the data has been fitted to an AR-GARCH model. It is found that by applying the POT method to the residuals the estimates of the extreme tail quantiles gives considerably greater accuracy than from AR-GARCH models with Gaussian or t-distributed errors. So with the use of this we are going to fit the data to an AR-GARCH

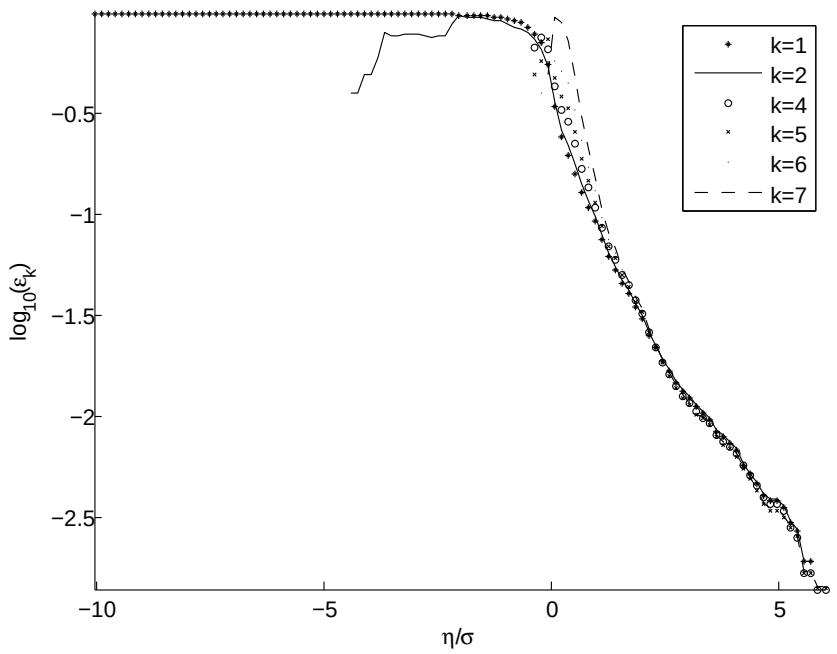
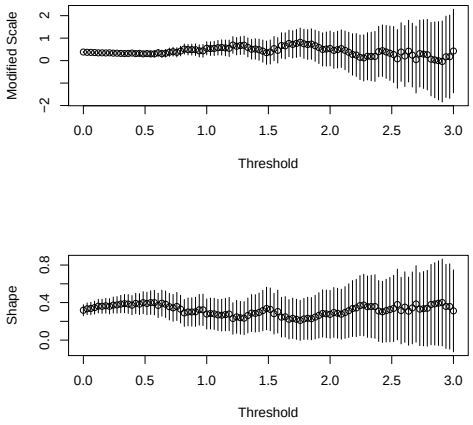


Figure 4.18: Empirical ACER functions plotted against scaled exceedance level.

(a) Mean residual life plot for the spot price data.



(b) Shape and modified scale parameters plotted against the threshold for spot price data.

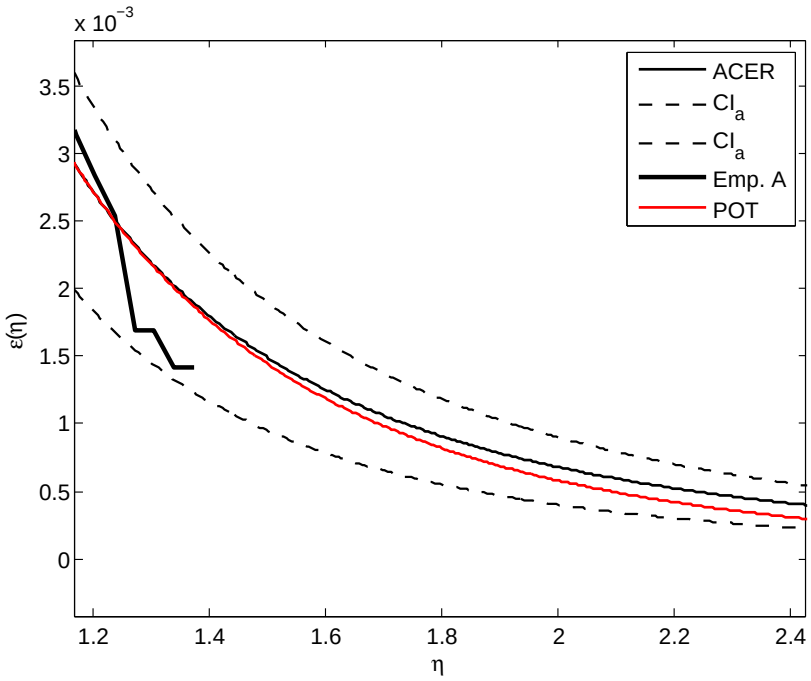


Figure 4.19: Empirical ACER functions plotted against scaled exceedance level for the spot price data.

model and apply both POT and ACER method to residuals. The AR-GARCH model which we are going to fit the data, which will be a combination of an AR(7) model and an GARCH(1,1) model, will be of the form

$$\begin{aligned} r_t &= a_0 + \sum_{i=1}^{7} a_i r_{t-i} + \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \end{aligned} \tag{4.4}$$

where $\epsilon_t = \sigma^t z_t$ with $z_t \sim \text{IID}(0,1)$ which is this case will be a standard Gaussian distribution or a Student's t-distribution scaled to have variance equal to 1, and as before ν degrees of freedom. So we are now going to fit two AR-GARCH models to the data, one with the Gaussian distribution and one with the Student's t-distribution. The ACER and POT method will be applied to the residuals from the AR-GARCH model with the Gaussian distribution. In table 4.12 the estimated parameters for the AR-GARCH model, for both the Gaussian and the Student's t distribution, is presented. The data is fitted to the AR-GARCH model with the function garchFit from the fGarch package for R, [8]. As shown in 2.5 the conditional tail quantiles for the AR-GARCH model when assuming either Gaussian or t-distribution. Instead of assuming these distribution it is also possible, as show in [3], to fit a GPD with the POT method to the residuals from the model, and estimating the conditional quantiles from the estimated GPD. We now also want to use the ACER method to estimate exceedance rates, and from this estimate the tail quantiles for this model. When fitting either the POT method or the ACER method to this model, we are first going to fit the residuals from the Gaussian AR-GARCH model with these two methods. When we are using the ACER method to model the the exceedance rate for the residuals it is sufficient to use the second ACER function. This is a real improvement over the straight forward method of fitting the ACER method to the data set, when it was necessary fifth ACER function. Since we are able to use the second ACER function instead of the fifth ACER function more of the data will be used, and the estimates will have better accuracy. This is the case since the AR-GARCH process actually remove much of the autocorrelation in the data, but according the Ljung–Box test, for which the values is provided for both model in 4.12, for the standardised residuals there are still some autocorrelation left. The tail marker is in this case set to $\nu_1 = 0$. For the POT method the the threshold is set after inspection of figure 4.20(a) and 4.20(b) to $u = 0.1$. So now that the residuals have been fitted to the GPD and the SGEV with the POT and ACER method respectively, the conditional tail quantiles can be calculated. The conditional tail quantiles will be calculated as

$$\alpha_{p,t} = a_0 + \sum_{i=1}^{7} a_i r_{t-i} + \sigma_t \alpha_p, \tag{4.5}$$

where $a_0 + \sum_{i=0}^{7} a_0 r_{t-1}$ is the conditional mean and σ_t is the conditional volatility and α_p is the quantile of the residual distribution, which either is fitted with POT or ACER, associated with probability p. After estimating the tail quantiles for all four models, we can use the data to observe the number of observations which falls above each quantile.

Parameter	Gaussian	Student's t
a_0	$2.998 \cdot 10^{-3}$	$-1.628 \cdot 10^{-3}$
a_1	-0.338	-0.283
a_2	-0.374	-0.309
a_3	-0.306	-0.256
a_4	-0.274	-0.236
a_5	-0.309	-0.269
a_6	-0.221	-0.181
a_7	0.388	0.425
α_0	$2.99 \cdot 10^{-4}$	$8.52 \cdot 10^{-4}$
α_1	$6.36 \cdot 10^{-2}$	0.455
β_1	0.93	0.74
ν	–	2.58
$Q(10)$	57.84	119.48
$Q(15)$	79.89	145.51
$Q(20)$	90.62	161.69

Table 4.12: Parameters for the AR-GARCH models.

In table 4.13 the number of observations above each quantile for the four models, with the expected number of exceedances in the second column from the left, is presented. From [3] it was expected that the Gaussian AR-GARCH model underestimates the quantiles, the Student's t AR-GARCH model overestimates the quantiles and the POT fit the quantiles rather good. We can also observe that the ACER method actually seem to predict the quantiles, based on exceedances of the observed data, even better than the POT method. In figure 4.20 the extrapolated ACER function is plotted alongside the exceedance rate for the GPD fitted with POT. If we compare this case to when we directly fitted the log-daily returns with these methods, it is clear that it is the POT method that gains the most, in terms of confidence interval width, from this procedure. When we did fit the log-daily returns directly, the declustering of the data resulted in a much greater loss of observations compared to the fitting of the residuals. The threshold could also be set lower when fitting the residuals, resulting ultimately in better use of the data compared to the direct fit of the log-daily returns. Even though the POT method perform es much better when fitting the residuals, the ACER method still is more accurate considering the confidence interval width and tail quantile estimates presented earlier.

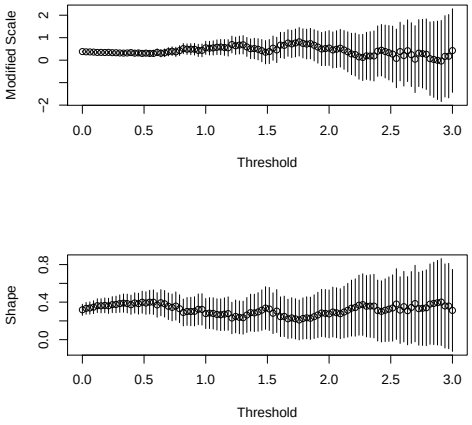
4.2.2 Dow Jones Index

In this section we are going to use the ACER and POT method for return level estimation in the same way as done with the synthetic data. For the ACER method we need to use the third ACER function, and for the POT method we declust the data with $M = 3$ and the threshold is set to $u = 2$ after inspection of 4.21(a) and 4.21(b). In figure 4.21 the extrapolated ACER function with confidence intervals, the empirical ACER function and

Probability	Expected	Gaussian	Student's t	POT	ACER
0.95	182	189	37	145	209
0.99	37	75	7	17	40
0.995	18	75	4	7	16
0.999	4	37	2	2	2
0.9995	2	32	1	2	2

Table 4.13: Number of exceedances observed over estimated tail-quantiles.

(a) Mean residual life plot for the residuals of the Gaussian AR-GARCH model.



(b) Shape and modified scale parameters plotted against the threshold for the AR-GARCH model.

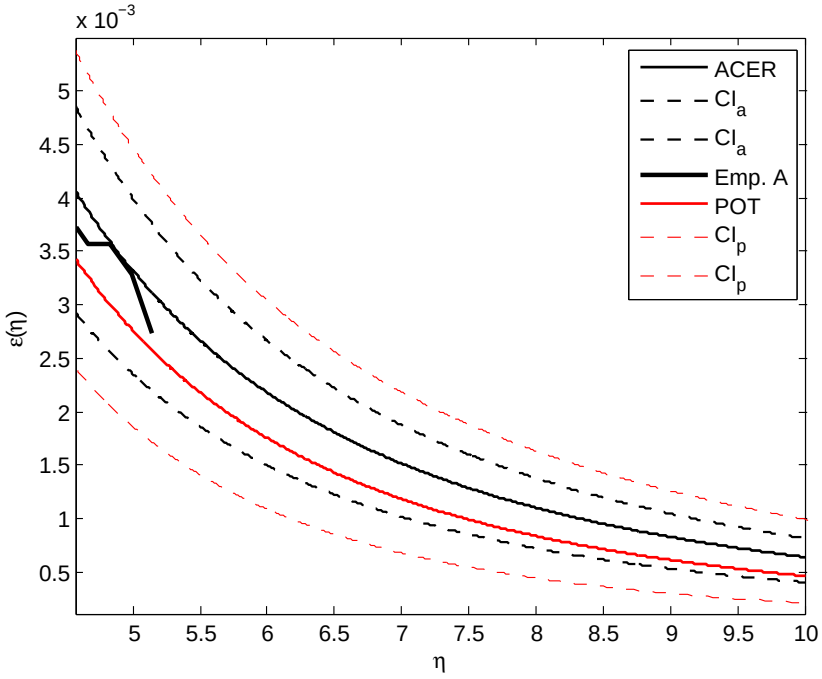
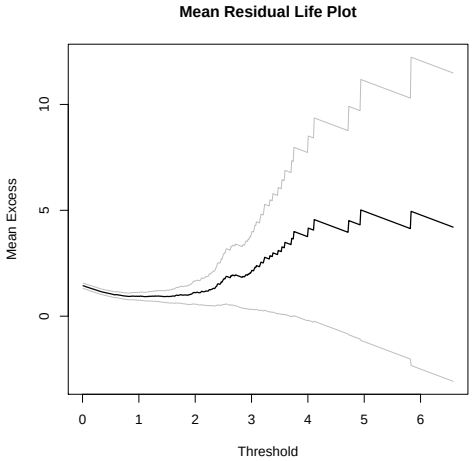
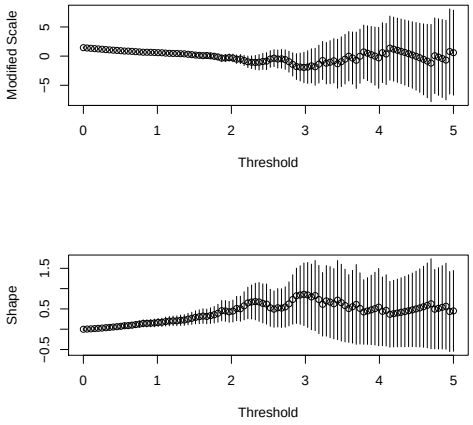


Figure 4.20: Extrapolated ACER function and exceedance rate for GPD fitted with POT for residuals of the Gaussian AR-GARCH model.



(a) Mean residual life plot for the Dow Jones index data.



(b) Shape and modified scale parameters plotted against the threshold for the Dow Jones index data.

the exceedance rate for the GPD fitted with POT is plotted. The confidence intervals for the POT method is very wide, and because this omitted in this figure. From this figure we observe that while both methods predict exceedances close to each other, the predicted exceedance rate of the POT method reveals that this method predict a heavier tail than the ACER method for the distribution of log-returns. The reason for this may be that while the ACER method, by the choice of weights, put less emphasis on more insecure observations (observations where the 95% confidence interval for the empirical ACER function is wide), the POT method does no such thing. As we observed in section 3.2.2 there seem to be some observations, which is much more extreme than

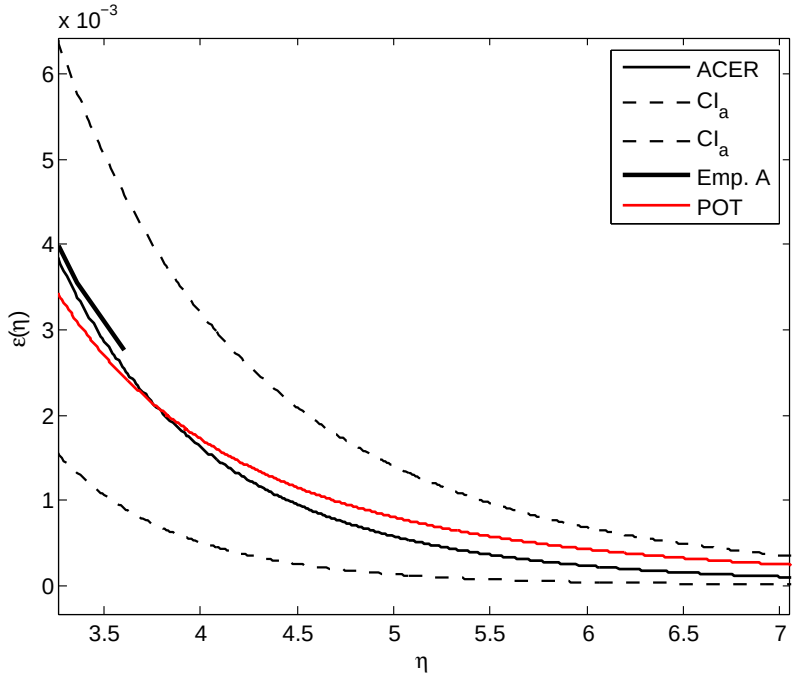


Figure 4.21: Extrapolated ACER function and exceedance rate for GPD fitted with POT for residuals of the Dow Jones index data.

Chapter 5

Conclusion

After comparing the performance of the ACER and POT methods in estimation of extreme values it seem like the ACER method generally perform better the POT method. The one clear exception is when there is no asymptotic argument in the derivation of the threshold exceedances, as was the case for the Pareto distribution. Further the ACER and POT methods seem to generally predict return levels which is close, but the accuracy, when considering the confidence intervals for these return levels, tend to be far greater when using the ACER method. This is most likely due to the fact that the POT method only uses observations above a given threshold when fitting the GPD. The method for estimating the parameters for the ACER function seem robust in that the parameters do not vary a lot between samples from the same model. Though it is observed that when the tails of the distribution become too heavy there may be problems estimating the parameters (i.e. a Pareto distribution with $\alpha \in (0, 1)$ or Student's t-distribution with $\nu \in (1, 2]$). This is not a great flaw as these distributions do not represent the phenomena we would want to apply this method to very well.

5.1 Further work

When modeling the residuals for the AR-GARCH model for the Nord Pool data, it could be interesting to use the combination of the AR-GARCH model and the ACER method to predict tail quantiles in the future. This could be done by for each data point (or in this case, day) fitting an AR-GARCH model to the data and using the ACER method to predict the tail quantiles for the residuals. This is not done in this thesis since the GARCH fitting is done with the fGarch package for R, while all implementation of the ACER method is done in Matlab.

Bibliography

[1] Naess A. and Gaidai O. Estimation of extreme values from sampled time series. *Structural Safety*, 31(4):325–334, July 2009.

[2] Søren Asmussen. *Applied probability and queues*. Springer-Verlag New York, 2003.

[3] Hans N. E. Byström. Extreme value theory and extremely large electricity price changes. *International Review of Economics and Finance*, 14(1):41 – 55, 2005.

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

[5] NAG. NAG toolbox for MATLAB, e04wd. Available from: `http://www.nag.co.uk/numeric/MB/manual_21_1/pdf/E04/e04wd.pdf`.

[6] Richard A. Davis Peter J. Brockwell. *Introduction to Time Series and Forecasting*. Springer Science, 2002.

[7] M. A. Ribatet. A User's Guide to the POT Package (version 1.0), August 2006. Available from: `http://cran.r-project.org/`.

[8] Diethelm Wuertz and Yohan Chalabi. fGarch package for R, November 2009. Available from: `http://cran.r-project.org/`.