

Estimating Value at Risk using some estimators of the generalized Pareto distribution under the peaks over threshold approach

Osama Abdelaziz Hussien
*Professor of Statistics
Department of Statistics, Mathematics
and Insurance
Alexandria University*

Mohammad Ibrahim Ahmmad
*Assistant professor of Statistics
Department of Statistics, Mathematics
and Insurance
Alexandria University*

Eman Ashraf Mohamed
*Graduate student
Department of Statistics, Mathematics
and Insurance
Alexandria University*

Abstract

Extreme value theory (EVT) has two main approaches for dealing with extremes; the block maxima (BM) and the peaks over threshold approach (POT). This paper focuses on the POT approach where the generalized Pareto distribution (GPD) is the limiting distribution for exceedances. The GPD has many estimators for the scale and shape parameters. A simulation study for comparing the performance of three new estimators, nonlinear least square (NLS), POT nonlinear least square (POT-NLS) and weighted nonlinear least square (WNLS), of the GPD under the peaks over threshold (POT) approach was conducted. Different distributions which belong to the maximum domain of attraction of the GPD were considered for generating the data to study the effect of the deviation from the GPD and the effect of different sample sizes, threshold values (u) and quantile levels. Then those estimates were used to calculate one of the common tail risk measures, the value at risk (VaR), for a heavy-tailed dataset.

1. Introduction

When natural disasters happen, we are interested in studying their occurrence and frequency, and whether anything could have been done to prevent them or at least to be prepared for them. In some applications we are not interested in estimating population central characteristics (e.g., the average temperature, the median income, the mean rainfall, etc.) depending on random samples selected from the population, but we are interested in estimating the maximum or the minimum, see (E Castillo & Sarabia, 1994). For example, in designing a dam, engineers are interested in both the average flood to estimate the total amount of water and the maximum flood to be prepared against it. These natural disasters could also include, earthquakes, massive snowfalls, massive forest fires and floods. Not only natural phenomena, but also many other practical situations wherein we are interested in extremes. These include: strength of material as a sheet of metal breaks at its weakest point so that a minimum strength determines the quality of the entire sheet, corrosion analysis as pitting corrosion can lead to the failure of metal structures such as tanks and tubes, telecommunication and insurance as pricing a high excess loss layer is connected to extreme losses. Thus, the determination of the

distribution of extremes (minima or maxima) of the phenomena at hand is very important to obtain a good solutions to the devastating disasters, (Enrique Castillo *et al.*, 2005). Extreme value theory (EVT) plays a very vital role in extremes. The importance of the EVT comes from its role in determining the distribution of extremes (maxima or minima).

There are two approaches for analyzing extreme values under the EVT; the block maxima (BM) and the peaks over threshold (POT) approach. The data at hand and the purpose of analyzing it recommend which approach to use. Each of those approaches depends on a certain limiting distribution. Since in this research we concentrate on the POT approach, the limiting distribution of that approach, the generalized Pareto distribution (GPD), is of great interest. Hence, estimating the parameters of the GPD becomes a very important task in EVT. Various estimation methods for the parameters of the GPD are available such as maximum likelihood estimation (MLE) which was discussed in (Grimshaw, 1993), (Davison & Smith, 1990) and (Smith, 1984), Pickands estimation method which was proposed by Pickands III (1975), method of moments (MOM) and probability weighted moments (PWM) were used by J. R. Hosking and Wallis (1987), method of l-moments was discussed in (J. R. M. Hosking, 1990), method of medians (MED) was proposed by He and Fung (1999), the likelihood moment estimation method (LME) was proposed by Zhang (2007), the nonlinear least square method (NLS) was proposed by Song and Song (2012), then more recently the POT nonlinear least square (POT-NLS) and the weighted nonlinear least square (WNLS) methods were proposed by (Park and Kim (2016)). Most of the above estimators have been investigated for their asymptotic behavior but the three newest estimators; NLS, POT-NLS and WNLS were not investigated for their asymptotic behavior. In this paper we concentrate on describing and comparing the asymptotic behavior of those estimators through a simulation study. Moreover, in this study different distributions which belong to the maximum domain of attraction of the GPD were considered for generating data with different sample sizes and threshold values (u) and quantile levels. Then those estimates were used to calculate the VaR and studying its performance depending on root mean square error (RMSE) and absolute relative bias (ARB).

After comparing the properties of the newest three estimators of the parameters of the GPD, those estimators were applied on one of the famous datasets called the Danish fire loss dataset. This paper is organized as follows. The Peaks over threshold approach of the extreme value theory is summarized in section 2. In section 3, the three new estimators of the GPD parameter that would be compared under the POT were discussed. In section 4, the results of the simulation study of the three estimators were presented. Section 5 is devoted for the analysis of the Danish fire loss dataset. Then conclusion is presented in section 6.

2. Extreme value theory

There are two classical ways to identify extremes in real data. The first approach models the maximum that the variable takes in successive periods. These extreme events are sometimes called block maxima. The block maxima of each block is modeled by the distributions in the classical extreme type theorems. i.e., the limiting distribution of the normalized maximum of the sequence must converge to either a Gumbel, Frechet, or Weibull distribution. This is known as the block maxima (BM) approach.

An alternative approach considers the process when it exceeds a given threshold u . All exceedences of the threshold u constitute extreme events, so u is typically chosen to be large. Then exceedences are fitted by the generalized Pareto distribution. This is known as the peaks over threshold (POT) approach, see (Reiss *et al.*, 2007), (Coles *et al.*, 2001) and (Embrechts *et al.*, 2013) for more details about these two approaches.

2.1 Peaks over threshold (POT)

Let X_1, X_2, \dots be a sequence of independent and identically distributed random variables, having a probability distribution F . Extreme events are the X_i that exceed some high threshold u . A description of the stochastic behavior of extreme events is given by the conditional probability

$$P\{X - u > x | X > u\} = \frac{1 - F(u+x)}{1 - F(u)}, \quad y > 0 \quad (2.1)$$

If the distribution F were known, the distribution of exceedences would also be known. But, in real situation, this is not the case, approximations that are applicable for high values of the threshold are sought. This parallels the use of the GEV as an approximation to the distribution of maxima when the population is unknown.

For a random variable X with cumulative distribution function F , we are interested in estimating the distribution function F_u of excesses above the threshold u . The distribution function F_u is called the conditional excess distribution function and is represented by

$$F_u(x) = P(X - u \leq x | X > u), \quad 0 \leq x \leq x_F - u \quad (2.2)$$

Where $x_F \leq \infty$ is the right endpoint of F : $x_F = \inf\{x: F(x) = 1\}$. Pickands III (1975) showed that for a large class of underlying distribution functions F , the conditional excess distribution function F_u , for large u , is well approximated by

$$F_u(x) = P(X - u \leq x | X > u) \approx F_{k,\beta}(x) \quad (2.3)$$

where $F_{k,\beta}(x)$ is the generalized Pareto distribution with a scale parameter β and a shape parameter k .

2.2 Generalized Pareto distribution (GPD)

Fisher and Tippett (1928) stated the extreme types theorem as follows: Suppose that X_1, \dots, X_n are i.i.d., $M_n = \max\{X_1, \dots, X_n\}$ and a_n, b_n are sequences of constants. If G is a non-degenerate limit distribution of $\frac{M_n - b_n}{a_n}$, i.e.

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = G(x) \quad (2.4)$$

, then G is of one of the following types:

Type I (Gumbel):

$$G(x) = \exp\left[-\exp\left(-\frac{x-\mu}{\beta}\right)\right], \text{ for } x \in R \quad (2.5)$$

where $\beta > 0$. Here, μ is a location parameter and β is a scale parameter.

Type II (Frechet):

$$G(x) = \exp\left[-\left(\frac{x-\mu}{\beta}\right)^{-\alpha}\right] \quad (2.6)$$

Where $x \geq \mu$; $\alpha, \beta > 0$.

Type III (Weibull):

$$G(x) = \exp\left[-\left(-\frac{x-\mu}{\beta}\right)^{-\alpha}\right] \quad (2.7)$$

where $x \geq \mu$; $\alpha, \beta > 0$.

Pickands III (1975) stated that if F satisfies the theorem of Fisher and Tippett (1928), so that for large n

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

where

$$G(x) = \exp\{-[1 + k(\frac{x-\mu}{\beta})]^{-1/k}\} \quad (2.8)$$

for some $\mu, \beta > 0$ and $k \neq 0$, where $G(x)$ is the generalized extreme value (GEV) distribution. Then, for large enough u , the distribution function of $(X-u)$, conditional on $X > u$, is approximately

$$F_{k,\beta}(x) = 1 - \left(1 + \frac{kx}{\tilde{\beta}}\right)^{-\frac{1}{k}} \quad (2.9)$$

where $\tilde{\beta} = \beta + k(u - \mu)$ and $F_{k,\beta}(x)$ is the GPD.

Pickands III (1975) theorem implies that, if block maxima have approximating distribution G , then threshold excesses have a corresponding approximate distribution within the generalized Pareto family. Moreover, the parameters of the generalized Pareto distribution of threshold excesses are uniquely determined by those of the associated generalized extreme value (GEV) distribution of block maxima. In particular, the parameter k in $F_{k,\beta}(x)$ is equal to that of the corresponding GEV distribution.

2.3 Tail risk measures

Risk measures are methods used to summarize the level of risk associated with some loss distributions. For example, a risk measure that is used for specifying capital requirements can be thought of as the amount of cash (capital) that must be added to a position to make its risk acceptable to regulators.

2.3.1 The value at risk measure (VaR)

VaR is a widely accepted standard risk measure used in risk analysis. Where it measures the maximum expected loss over a given time period at a given confidence level. It is defined as the q -quantile of the loss distribution, (Reiss *et al.*, 2007). Suppose that a random variable X with continuous distribution function F models losses, then VaR_p at the $100p\%$ quantile is the $100p^{\text{th}}$ quantile of the distribution F .

$$VaR_p(X) = F^{-1}(p) \quad (2.10)$$

For example; if a portfolio of stocks has a 10-day 95% VaR of \$1 million, that means that there is a 0.05 probability that the portfolio will fall in value by more than \$1 million over a 10-day period, $Pr(X > VaR_p) \leq 1 - p$, $Pr(X > 1) \leq 1 - 0.95 \leq 0.05$.

(Park & Kim, 2016) showed that when the POT approach is employed, the underlying loss distribution can be written as a combination of the body and tail parts with the latter modeled by the GPD.

Let $F(X)$ be the distribution function of an arbitrary continuous distribution, then the exceedance distribution of loss events over u is defined by:

$$F_u(y) = P(X - u \leq y | X > u) = \frac{F(u+y) - F(u)}{1 - F(u)} \quad (2.11)$$

which is assumed to be GPD (k, β), then $F(x)$ can be written as:

$$\begin{aligned} F(x) &= P(X \leq x) = (1 - F(u))F_u(x - u) + F(u) \\ &= (1 - F(u))F_{k,\beta}(x - u) + F(u) \end{aligned} \quad (2.12)$$

Replacing F_u by the GPD and $F(u)$ by the estimate $\frac{n - N_u}{n}$ where n is the total number of observations and N_u is the number of observations above the threshold (u)

$$\hat{F}_{k,\beta}(x) = 1 - \frac{N_u}{n} \left(1 + \frac{\hat{k}}{\hat{\beta}}(x - u)\right)^{-1/\hat{k}} \quad (2.13)$$

then the VaR can be obtained by inverting the estimated version of the GPD:

$$\begin{aligned} \widehat{VaR}_p(X) &= F_{\hat{k},\hat{\beta}}^{-1} \left[1 - \frac{1-p}{1-F(u)} \right] \\ &= u + \frac{\hat{\beta}}{\hat{k}} \left[\left(\frac{n}{n_u} (1-p) \right)^{-\hat{k}} - 1 \right] \end{aligned} \quad (2.14)$$

3. GPD estimation

The GPD family depend on the scale and shape parameters. In this section, some methods for the estimation of those parameters were reviewed: nonlinear least square method (NLS), POT-nonlinear least square method (POT-NLS) and weighted nonlinear least squares method (WNLS).

3.1 The nonlinear least square method (NLS)

This method is proposed by (Song & Song, 2012) to estimate the parameters of the GPD. It minimizes the sum of squared deviations between the empirical distribution function (EDF) and the theoretical distribution function of the GPD at observed values. The estimators of k and β can be obtained as follows:

- 1- Pick a threshold value (u) and compute the EDF using observations above u .
- 2- (First –step) Take the log of (1-EDF) and the log of (1-GPD distribution function), which are set as the response and explanatory variables, respectively. Since the GPD distribution function is $F_{k,\beta}(x) = 1 - \left(1 + \frac{kx}{\beta}\right)^{-1/k}$ for $k \neq 0$,

$$\log\left(1 - F_{k,\beta}(x)\right) = \left(-\frac{1}{k}\right) \log\left(1 + \frac{kx}{\beta}\right)$$

where $F_n(x)$ is the empirical distribution function,

$$\log(1 - F_n(x)) = \left(-\frac{1}{k}\right) \log\left(1 + \frac{kx}{\beta}\right) + \varepsilon_i \quad (3.1)$$

which can be solved in the following way:

$$(\hat{k}_1, \hat{\beta}_1) = \arg \min_{(k,\beta)} \sum_{i=1}^{n_u} \left[\log(1 - F_n(x_i)) - \log\left(1 - F_{k,\beta}(x_i)\right) \right]^2 \quad (3.2)$$

where
$$\hat{k}_1 = \frac{\partial \sum_{i=1}^{n_u} [\log(1 - F_n(x_i)) - \log(1 - F_{k,\beta}(x_i))]^2}{\partial k}$$

and
$$\hat{\beta}_1 = \frac{\partial \sum_{i=1}^{n_u} [\log(1 - F_n(x_i)) - \log(1 - F_{k,\beta}(x_i))]^2}{\partial \beta}$$

3- (Second step) fit the NLS regression of the original EDF on the original GPD function with the initial values of $(\hat{k}_1, \hat{\beta}_1)$.

$$(\hat{k}_2, \hat{\beta}_2) = \arg \min_{(k,\beta)} \sum_{i=1}^{n_u} [F_n(x_i) - F_{k,\beta}(x_i)]^2 \quad (3.3)$$

Because the NLS is minimizing the difference between the empirical distribution function which starts from the value of the threshold and the theoretical distribution function which starts from $u=0$, hence the difference between them is not measured in consistent way and the NLS can only be applicable when $u=0$, (Park & Kim, 2016).

3.2 POT nonlinear least squares method (POT-NLS)

This method is proposed by (Park & Kim, 2016) to improve the NLS method. The idea of this method is to minimize the squared distance between the empirical distribution function $F_n(x)$ and the cumulative distribution function $F(x)$, not the distance between $F_n(x)$ and theoretical distribution function $F_{k,\beta}$.

$$\begin{aligned} F(x) &= P(X \leq x) = (1 - F(u))F_u(x - u) + F(u) \\ &= (1 - F(u))F_{k,\beta}(x - u) + F(u) = (1 - F(u))F_{k,u,\beta}(x) + F(u) \end{aligned}$$

So the revised first step in the NLS method becomes:

$$\begin{aligned}
(\hat{k}_1, \hat{\beta}_1) &= \arg \min_{(k, \beta)} \sum_{i=1}^{n_u} [\log(1 - F_n(x_i)) - \log(1 - F_{k, \beta}(x_i))]^2 \\
&= \arg \min_{(k, \beta)} \sum_{i=1}^{n_u} [\log(1 - F_n(x_i)) \\
&\quad - \log(1 - (1 - F_n(u)) F_{k, u, \beta}(x_i) - F_n(u))]^2 \\
&= \arg \min_{(k, \beta)} \sum_{i=1}^{n_u} \left[\log \frac{1 - F_n(x_i)}{1 - F_n(u)} - \log(1 - F_{k, u, \beta}(x_i)) \right]^2 \tag{3.4}
\end{aligned}$$

and the revised second step becomes:

$$\begin{aligned}
(\hat{k}_2, \hat{\beta}_2) &= \arg \min_{(k, \beta)} \sum_{i=1}^{n_u} [F_n(x_i) - F(x_i)]^2 \\
&= \arg \min_{(k, \beta)} \sum_{i=1}^{n_u} [F_n(x_i) - (1 - F_n(u)) F_{k, u, \beta}(x_i) - F_n(u)]^2 \\
&= \arg \min_{(k, \beta)} \sum_{i=1}^{n_u} (1 - F_n(u))^2 \left[\frac{F_n(x_i) - F_n(u)}{1 - F_n(u)} - F_{k, u, \beta}(x_i) \right]^2 \\
&= \arg \min_{(k, \beta)} \sum_{i=1}^{n_u} \left[\frac{F_n(x_i) - F_n(u)}{1 - F_n(u)} - F_{k, u, \beta}(x_i) \right]^2 \tag{3.5}
\end{aligned}$$

Park and Kim (2016) called this revised second step "POT-NLS" which they claim to fit the tail part in a more appropriate way than the NLS, by using the truncated distribution. As Park and Kim (2016) criticized the NLS estimators proposed by Song and Song (2012) because the NLS is minimizing the difference between the empirical distribution function which starts from the value of the threshold and the theoretical distribution function which starts from $u=0$, hence the difference between them is not measured in consistent way and the NLS can only be applicable when $u=0$. In other words, since $F_{k, u, \beta}$ is the conditional distribution of $X|X > u$ whereas the EDF $F_n(x)$ is the distribution of the entire X . And the summation in the second step of NLS is carried out for only the observations exceeding the threshold, the EDF conditional on $X > u$ cannot be created by just restricting the range of the observations but this should be done using truncation.

3.3 The weighted nonlinear least squares method (WNLS)

This method is proposed also by (Park & Kim, 2016) after proposing the POT-NLS to improve the NLS method. This method proposed adding suitably chosen weights for each squared deviance term. To determine the suitable weight for each squared error; the first line of the revised second step is equivalent to applying the least squares method to the following regression setup:

$$F(x_i) = F_n(x_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (3.6)$$

Where ε_i is the error term with $E(\varepsilon_i)=0$, that is from the fact that the EDF is an unbiased estimate of $F(x)$ for all x .

Noting that $F(x)$ is a uniform distributed random variable for any x , Park and Kim (2016) proposed the weight for each response variable $F(x_i)$ to be determined as the reciprocal of its variance. Assuming that $X_1 > \dots > X_n$, the distribution of $F(x_i)$ is that of $U_{n-i+1:n}$, and where the $(n - i + 1)^{\text{th}}$ order statistic of the uniform random variable has a Beta distribution with $B(n - i + 1, i)$ from the standard distribution theory, (Krishnamoorthy, 2006). Where the first two moments of $B(n - i + 1, i)$ are:

$$E(U_{n-i+1:n}) = \frac{n-i+1}{n+1}, \quad (3.7)$$

$$Var(U_{n-i+1:n}) = \frac{i(n-i+1)}{(n+1)^2(n+2)} \quad (3.8)$$

, hence the weight for $F(x_i)$ is $(Var(U_{n-i+1:n}))^{-1}$

Consequently the WNLS becomes:

$$\begin{aligned} (\hat{k}_3, \hat{\beta}_3) &= \arg \min_{(k, \beta)} \sum_{i=1}^{n_u} [Var(U_{n-i+1:n})]^{-1} [F_n(x_i) - F(x_i)]^2 \\ &= \arg \min_{(k, \beta)} \sum_{i=1}^{n_u} \left[\frac{i(n-i+1)}{(n+1)^2(n+2)} \right]^{-1} \left[\frac{F_n(x_i) - F_n(u)}{1 - F_n(u)} - F_{k,u,\beta}(x_i) \right]^2 \end{aligned} \quad (3.9)$$

One advantage of the WNLS over the NLS is that it estimates the extreme quantiles in a more stable manner as larger weights are given for $F(x_i)$ values as x_i moves towards the tail side.

All these estimators were not investigated for asymptotic behavior by their authors. In this paper a simulation study was conducted for studying the asymptotic behavior of the three newest estimators and also permit the study of the behavior of those estimators for small sample sizes and for different distributions of the tail.

4. Simulation study

The performance of the three estimation methods of scale and shape parameters was compared using the Monte Carlo simulation. Simulation has been performed using statistical programming R. VaR as a Tail risk measure was estimated using the different estimators from different heavy-tailed distributions including (GPD, log-gamma, Pareto and Cauchy). The GPD is fitted to the sample exceedances above a high threshold to estimate the VaR. Different VaR levels including VaR 95%, 99%, 99.9% and 99.99% were computed.

The simulated samples were generated from different heavy tailed distributions: GPD, Pareto, Cauchy and Log-gamma. It should be noted that if X is a random variable that has a Pareto distribution with a shape parameter $\alpha > 0$ and a scale parameter $\theta > 0$, then its cdf is given by:

$$F(x) = 1 - \left(\frac{x}{\theta}\right)^{-\alpha}, \quad x \geq \theta \quad (4.1)$$

, if X is a Cauchy random variable with parameter (α, θ) , then X has a cdf:

$$F(x) = \frac{1}{\pi} \operatorname{artcan}\left(\frac{x - \alpha}{\theta}\right) + \frac{1}{2}, \quad -\infty < x < \infty \quad (4.2)$$

and if X is a gamma random variable with parameter (α, θ) , then $Y = \exp(x)$ is a log-gamma with pdf:

$$f(y) = \frac{(\log y)^{\alpha-1}}{y\theta^\alpha \Gamma(\alpha)} \exp\left(-\log \frac{y}{\theta}\right), \quad y > 1, \theta > 0, \alpha > 0 \quad (4.3)$$

A random sample of size 10000 is generated from the given distribution, then a threshold u (the 100 p th sample quantile) is picked, then the GPD is fitted with the observations above u using different estimators to estimate k and β , and these estimates are substituted for in the formula of VaR in (2.14). To compare the estimating techniques the performances are evaluated by the absolute relative bias (ARB). In addition to the ARB, in the GPD model the root mean square error (RMSE) is also used as a method for comparing the estimating methods. The RMSE and ARB are defined as

$$ARB = \frac{|\theta_{est} - \theta_{true}|}{\theta_{true}} \quad (4.4)$$

$$RMSE = \sqrt{(\theta_{est} - \theta_{true})^2} \quad (4.5)$$

where $\theta_{est}, \theta_{true}$ are the estimated and true values of the quantiles of the GPD, respectively.

4.1 GPD (k, β)

Trials have been made with $n = 1000$, $k = -0.5, -0.4, \dots, 0, \dots, 0.4, 0.5$, $\beta = 1, 10$ and 100 and $VaR = 0.95, 0.99, 0.999$ and 0.9999 .

Table 1

Estimating VaR 95, 99, 99.9 and 99.99% of the GPD for $n = 1000$ and $\beta = 1$: RMSE and ARB.

shape			-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5
VaR 95%	NLS	RMSE	0.01	0.01	0.02	0.02	0.03	0.04	0.05	0.07	0.09	0.13	0.16
		ARB	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02
	POT-NLS	RMSE	0.01	0.01	0.02	0.02	0.03	0.04	0.06	0.07	0.1	0.14	0.17
		ARB	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02
	WNLS	RMSE	0.01	0.01	0.02	0.02	0.03	0.04	0.06	0.07	0.1	0.14	0.17
		ARB	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02
VaR 99%	NLS	RMSE	0.01	0.02	0.2	0.04	0.07	0.1	0.16	0.25	0.38	0.62	0.94
		ARB	0	0.01	0.08	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.04
	POT-NLS	RMSE	0.01	0.01	0.02	0.04	0.06	0.1	0.16	0.26	0.4	0.66	1.04
		ARB	0	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.05
	WNLS	RMSE	0.01	0.01	0.02	0.03	0.05	0.09	0.14	0.21	0.33	0.56	0.86
		ARB	0	0	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.04
VaR 99.9%	NLS	RMSE	0.02	0.03	0.05	0.09	0.17	0.3	0.56	1.03	1.84	3.65	6.82
		ARB	0.01	0.01	0.01	0.02	0.03	0.03	0.04	0.06	0.06	0.08	0.09
	POT-NLS	RMSE	0.02	0.03	0.07	0.14	0.25	0.47	0.86	1.64	3.07	5.8	11.23
		ARB	0.01	0.01	0.02	0.03	0.04	0.05	0.07	0.09	0.1	0.12	0.14
	WNLS	RMSE	0.01	0.02	0.04	0.08	0.15	0.31	0.56	1.14	2.15	4.19	7.98
		ARB	0	0.01	0.01	0.02	0.02	0.04	0.05	0.06	0.07	0.09	0.1
VaR 99.99%	NLS	RMSE	0.02	0.03	0.07	0.14	0.29	0.6	1.32	2.96	6.37	15.48	35.43
		ARB	0.01	0.01	0.02	0.03	0.04	0.05	0.07	0.09	0.1	0.13	0.14
	POT-NLS	RMSE	0.03	0.06	0.12	0.27	0.56	1.19	2.59	5.81	13.38	30.41	71.95
		ARB	0.01	0.02	0.03	0.05	0.07	0.1	0.13	0.17	0.2	0.23	0.27
	WNLS	RMSE	0.01	0.03	0.07	0.16	0.33	0.74	1.62	3.88	8.83	20.54	47.93
		ARB	0.01	0.01	0.02	0.03	0.04	0.06	0.09	0.12	0.14	0.16	0.19

Table 2

Estimating VaR 95, 99, 99.9 and 99.99% of the GPD for $n = 1000$ and $\beta = 10$: RMSE and ARB.

shape			-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5
VaR 95%	NLS	RMSE	0.08	0.12	0.15	0.2	0.29	0.37	0.53	0.69	0.95	1.27	1.73
		ARB	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02
	POT-NLS	RMSE	0.09	0.12	0.17	0.21	0.31	0.39	0.57	0.74	1.02	1.4	1.85
		ARB	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02
	WNLS	RMSE	0.09	0.12	0.16	0.21	0.31	0.39	0.56	0.74	1.01	1.38	1.84
		ARB	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02
VaR 99%	NLS	RMSE	0.09	0.16	0.26	0.41	0.66	0.99	1.62	2.45	3.84	6.15	9.21
		ARB	0	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.04
	POT-NLS	RMSE	0.09	0.15	0.23	0.39	0.64	0.99	1.61	2.54	4.04	6.28	10.06
		ARB	0	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.04
	WNLS	RMSE	0.08	0.13	0.2	0.32	0.53	0.81	1.32	2.07	3.36	5.29	8.41
		ARB	0	0	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.04
VaR 99.9%	NLS	RMSE	0.15	0.29	0.51	0.92	1.65	2.91	5.51	9.9	18.61	36.78	64.21
		ARB	0.01	0.01	0.01	0.02	0.03	0.03	0.04	0.05	0.06	0.08	0.08
	POT-NLS	RMSE	0.2	0.37	0.71	1.34	2.52	4.66	8.49	16.34	30.97	58.5	109.13
		ARB	0.01	0.01	0.02	0.03	0.04	0.05	0.07	0.09	0.1	0.12	0.14
	WNLS	RMSE	0.1	0.19	0.39	0.79	1.51	2.94	5.51	11.09	21.33	39.92	77.82
		ARB	0	0.01	0.01	0.02	0.02	0.03	0.04	0.06	0.07	0.09	0.1
VaR 99.99%	NLS	RMSE	0.18	0.37	0.72	1.43	2.9	5.94	13.11	28.08	64	158.97	333.42
		ARB	0.01	0.01	0.02	0.03	0.04	0.05	0.07	0.08	0.1	0.13	0.13
	POT-NLS	RMSE	0.3	0.59	1.27	2.61	5.65	12.33	25.52	59.29	135.42	317.7	703.76
		ARB	0.01	0.02	0.03	0.05	0.07	0.1	0.13	0.17	0.2	0.24	0.27
	WNLS	RMSE	0.13	0.3	0.67	1.48	3.21	7.29	15.84	38.25	88.21	200.08	468.18
		ARB	0.01	0.01	0.02	0.03	0.04	0.06	0.08	0.11	0.14	0.16	0.19

Table 3

Estimating VaR 95, 99, 99.9 and 99.99% of the GPD for $n = 1000$ and $\beta = 100$: RMSE and ARB.

shape			-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5
VaR 95%	NLS	RMSE	0.86	1.17	1.56	2.1	2.83	3.89	5.15	6.8	9.3	13.46	16.67
		ARB	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02
	POT-NLS	RMSE	0.91	1.2	1.66	2.31	2.92	4.15	5.74	7.88	9.9	13.9	17.75
		ARB	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02
	WNLS	RMSE	0.89	1.16	1.63	2.29	2.87	4.09	5.68	7.82	9.83	13.71	17.66
		ARB	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02
VaR 99%	NLS	RMSE	0.97	1.66	2.63	5.33	6.62	10.44	16	25.38	38.67	59.48	93.62
		ARB	0	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.04
	POT-NLS	RMSE	0.93	1.4	2.43	4.04	6.43	10.25	16.15	25.55	40.94	64.41	101.23
		ARB	0	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.04
	WNLS	RMSE	0.81	1.23	2.05	3.38	5.26	8.51	13.49	21.12	33.77	55.24	84.25
		ARB	0	0	0.01	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.04
VaR 99.9%	NLS	RMSE	1.51	2.89	5.19	14.28	16.58	30.68	55.16	101.27	188.41	355.99	6053.5
		ARB	0.01	0.01	0.01	0.02	0.03	0.04	0.04	0.05	0.06	0.08	1
	POT-NLS	RMSE	1.95	3.5	6.94	13.22	25.8	46.64	84.64	160.38	305.76	558.09	1103.4
		ARB	0.01	0.01	0.02	0.03	0.04	0.05	0.07	0.08	0.1	0.12	0.14
	WNLS	RMSE	0.93	1.99	3.91	8.07	15.53	29.27	56.36	106.91	208.12	408.55	786.05
		ARB	0	0.01	0.01	0.02	0.03	0.03	0.04	0.06	0.07	0.09	0.1
VaR 99.99%	NLS	RMSE	1.83	3.73	7.3	26.9	28.97	62.24	131.6	286.29	657.05	1501	3533
		ARB	0.01	0.01	0.02	0.03	0.04	0.05	0.07	0.09	0.1	0.12	0.14
	POT-NLS	RMSE	2.88	5.71	12.21	25.7	57.87	120.32	254.39	579.73	1323.9	2856.2	7227.3
		ARB	0.01	0.02	0.03	0.05	0.07	0.1	0.13	0.17	0.2	0.23	0.26
	WNLS	RMSE	1.3	3.09	6.56	15.39	32.99	72.1	162.74	362.63	836.99	1983.53	4743.23
		ARB	0.01	0.01	0.02	0.03	0.04	0.06	0.08	0.11	0.14	0.16	0.19

From tables 1, 2 and 3 it is found that as the level of VaR increases, both the RMSE and ARB of the three estimators increase. Also as the value of the shape parameter (k) increases, both the RMSE and ARB of the three estimators increase. Whereas, when the value of the scale parameter (β) increases, only the RMSE of the three estimators increases while the ARB remains almost the same. In general, POT-NLS have generally the highest RMSE and ARB. While for $u = 0.99$, WNLS has the best performance. For $u = 0.999$, WNLS is better for $-0.5 < k < 0$ and NLS is better for $0 < k < 0.5$. While for $u = 0.9999$, WNLS is better for $-0.5 < k < -0.30$ and NLS is better for $-0.2 < k < -0.5$.

4.2 Log-gamma (2, 3)

Trials have been made with $n = 100, 200$, shape = 2, scale = 3 and VaR = 0.99, 0.999 and 0.9999.

Table 4

Estimating VaR 99, 99.9 and 99.99% of log-gamma (2, 3) for $n = 100$ and 200: ARB.

n	u	VaR 99%			VaR 99.9%			VaR 99.99%		
		NLS	POT-NLS	WNLS	NLS	POT-NLS	WNLS	NLS	POT-NLS	WNLS
		ARB	ARB	ARB	ARB	ARB	ARB	ARB	ARB	ARB
100	0.99				1.58	1	1	6.62	1	1
200	0.98	0.33	0.87	0.87	1.52	1.04	0.98	5.32	2.05	1.24

From table 4 it is clear that WNLS has the best performance in general. Also, POT has a good performance in the case of VaR99% and $u = 0.98$, VaR 99.9% and $u = 0.98$, VaR99.9% and $u = 0.99$ and VaR99.99% and $u = 0.99$. While, NLS has the worst performance for $u = 0.98$ and $u = 0.99$.

4.3 Pareto (2, 0.5)

Trials have been made with $n = 100, 200, 300$, shape = 2, scale = 0.5 and VaR = 0.99, 0.999 and 0.9999.

Table 5

Estimating VaR 99, 99.9 and 99.99% of Pareto (2, 0.5) for $n = 100, 200$ and 300: ARB.

n	u	VaR 99%			VaR 99.9%			VaR 99.99%		
		NLS	POT-NLS	WNLS	NLS	POT-NLS	WNLS	NLS	POT-NLS	WNLS
		ARB	ARB	ARB	ARB	ARB	ARB	ARB	ARB	ARB
100	0.99				0.28	0.52	0.42	0.43	2.08	1.11
200	0.98	0.14	0.15	0.15	0.27	0.51	0.42	0.48	1.54	1.01
300	0.97	0.15	0.14	0.14	0.29	0.49	0.4	0.47	1.07	0.78

From table 5, it is found that NLS has the best performance, while, POT has the worst performance. It is also noticed that for high levels of VaR (99.99%), as the sample size (n) increases, the ARB decreases except for the NLS estimator.

4.4 Cauchy (0, 1)

Trials have been made with $n = 100, 200, 300$, shape = 0, scale = 1 and VaR = 0.99, 0.999 and 0.9999.

Table 6

Estimating VaR 99, 99.9 and 99.99% of Cauchy (0,1) for $n = 100, 200, 300$ and 1000: ARB.

n	u	VaR 99%			VaR 99.9%			VaR 99.99%		
		NLS	POT-NLS	WNLS	NLS	POT-NLS	WNLS	NLS	POT-NLS	WNLS
		ARB	ARB	ARB	ARB	ARB	ARB	ARB	ARB	ARB
100	0.99				0.22	0.26	0.22	0.44	0.87	0.55
200	0.98	0.07	0.08	0.08	0.2	0.26	0.21	0.35	0.35	0.45
300	0.97	0.07	0.07	0.07	0.15	0.26	0.21	0.23	0.31	0.42

From table 6, it is concluded that NLS has the best performance except for VaR99% and $u = 0.9$ and for VaR99.9% and $u = 0.99$, where WNLS is the best. While POT-NLS has the worst performance except for VaR 99% and $u = 0.9$ and VaR 99% and $u = 0.98$. Also, for VaR99.9% and $u = 0.99$, WNLS has the worst performance. Also, it is noticed that as the sample size (n) increases, the ARB decreases especially in the case of high VaR levels (VaR 99.99%).

5. Application

Insurance is one of the fields where EVT is extensively used. Insurance companies can use the POT approach to determine the price of the insurance layer and also to determine the limits of each layer.

5.1 Description of the dataset

The full Danish data comprise 2492 losses and can be considered as being essentially all Danish fire losses over one million Danish Krone (DKK) from 1980 to 1990 plus a number of smaller losses below one million DKK. The data represent historical data on insurance losses which exceed a certain amount known as a displacement, it is practically impossible to collect data on all losses and data on small losses are of less importance.

So, Insurance is generally provided against significant losses and insured parties deal with small losses themselves and may not report them. Thus, the data should be thought of as being realizations of random variables truncated at a displacement δ where δ is much less than r ($\delta \ll r$). We restrict our attention to the 2156 losses exceeding one million so that the effective displacement δ is one. The loss figure is a total loss figure for the event concerned and includes damage to building, damage to furniture, and personal property as well as loss of profit. For details, one may look at (McNeil, 1997).

Suppose we are interested in a high excess loss layer with lower and upper attachment points r and R , respectively, where r is large and $R > r$.

This means the payout Y_i on a loss X_i is given by:

$$Y_i = \begin{cases} 0 & \text{if } 0 < X_i < r \\ X_i - r & \text{if } r \leq X_i < R \\ R - r & \text{if } R \leq X_i < \infty \end{cases} \quad (5.1)$$

The following two questions need to be answered:

1. The pricing problem given r and R . what should this insurance layer cost a customer?

For a general layer (r, R) , price is given by

$$price = \int_r^R (x - r) f_{X^\delta}(x) dx + (R - r) (1 - F_{X^\delta}(R)) \quad (5.2)$$

Where $f_{X^\delta} = \frac{dF_{X^\delta}(x)}{dx}$ denotes the density function for the losses truncated at δ .

Picking a high threshold u ($< r$) and fitting a GPD model to the excesses, $F_{X^\delta}(x)$ for $x > u$ can be estimated using the tail estimation procedure

$$\hat{F}_{X^\delta}(x) = (1 - F_n(u)) F_{\hat{k}, u, \hat{\beta}}(x) + F_n(u) \quad (5.3)$$

where \hat{k} and $\hat{\beta}$ are parameter estimates and $F_n(u)$ is an estimate of $P\{X^\delta \leq u\}$ based on the empirical distribution function of the data.

2. The optimal attachment point (r) problem. If we want payouts greater than a specified amount to occur with at most a specified frequency. How low can we set r ?

Fire Danish loss data is used to answer two questions; first: the price of high excess layer, second: how to determine high excess layer. It is clear that we are interested in extreme losses and specifically a certain high loss value (threshold)

that it is expected to exceed with certain low probability. The available data is on daily basis. Also, data are independent of time as there is no relation between the occurrence of a loss and the time of occurrence. All these reasons suggest the use of the POT approach as both time series and block maxima approaches will not answer the required questions.

5.2 Statistical analysis

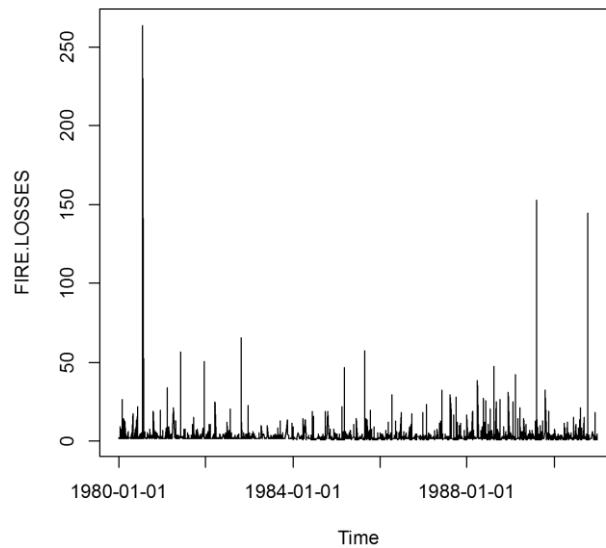


Figure 5.1.1 The time series plot of fire loss data

It can be seen from figure 5.5.1 that there is no trend or cycles in the data. Also the data takes a random look which means that the data are independent from time.

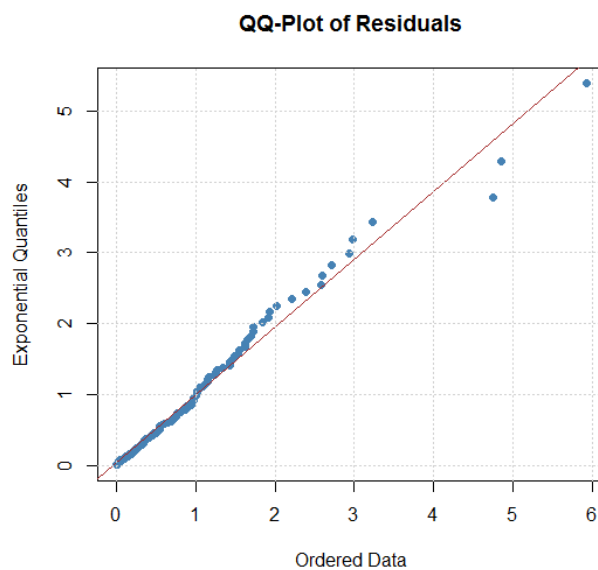


Figure 5.1.2 Q-Q plot of fire loss data.

The QQ plot in figure 5.1.2 examines visually the hypothesis that the losses come from an exponential distribution with a medium sized tail. If the points lie approximately along a straight line, the data are an i.i.d sample from an exponential distribution. A concave departure from ideal shape, as in this example, indicates a heavier tail. A convexity indicates a shorter tailed distribution.

5.2.1 Threshold selection

To select an appropriate threshold, three tools were used including the sample mean excess plot, the parameter plot and the goodness of fit test.

a) Sample mean excess function plot

Assuming that the data could be fitted by the GPD, then it is known that the theoretical mean excess function of the GPD which is known to be linear is given by:

$$e(u) = E(X - u | X > u) = \frac{\beta + ku}{1 - k}, \quad \beta + ku > 0 \tag{5.4}$$



Figure 5.1.3 mean excess plot of fire loss data.

The empirical plot seems to follow a reasonably straight line with positive gradient above a certain value u , then this is an indication that the data follow a GPD with positive shape parameter in the tail area above u . There is evidence of a straightening out of the plot above a threshold of 10 or perhaps again above a threshold of 20.

b) *The parameter plot*

Using the MLE to estimate the parameters of the GPD, it can be seen that there is stability in the estimates of the shape and scale parameters before $u=4$ and again before $u=10$, figure 5.1.4.

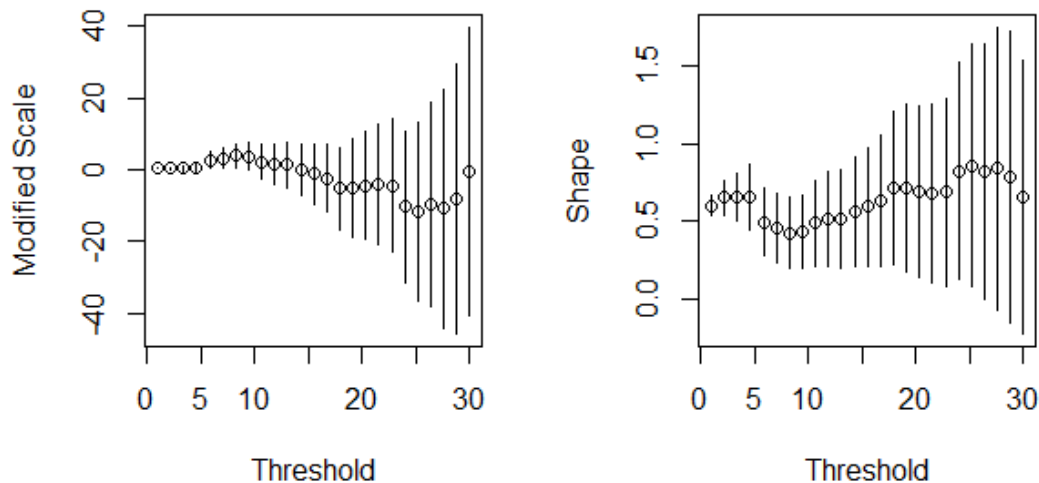


Figure 5.1.4 the parameter plot of the loss data.

c) *Goodness of fit test*

Also goodness of fit test can be used to select the appropriate threshold. The method of choosing the threshold by using as many exceedances as possible subject to passing a test for GPD can be investigated for power as follows: The data analyst wishes to fit a GPD to the exceedances far enough in the tail. Therefore start with as many as possible exceedances, say 100, and tests; if the test fails, omit the smallest values one by one and test again until the test yields acceptance for the GPD, See (Choulakian & Stephens, 2001). Using the Anderson Darling test recommends the use of $u=10$ as seen in table 5.1.1.

Table 5.1.1 Anderson-Darling test for goodness of fit.

Threshold	p.value	Est.shape	Est.scale
3	0.21	0.67	2.18
4	0.04	0.72	2.62
5	0.01	0.63	3.81
10	0.51	0.5	6.95
15	0.32	0.58	8.09
20	0.26	0.69	9.5

Thus the choice of threshold depend on the question we want to answer, If we are trying to determine the optimal attachment point or to price a high layer, we may choose to fit the GPD model with the threshold of 4 rather than 10, because this will be more conservative leading to higher estimates of r and price. Although the GPD model may not fit the data quite well above this lower threshold as it does above the high threshold of 10 but it might be safer to use the low threshold to make calculations. On the other hand, there may be business reasons for trying to keep the attachment point or premium low, there may be a competition to sell high excess policies and this may mean that basing calculations only on the highest observed losses is favored since this will lead to more attractive products.

5.2.2 Fitting the GPD to Danish loss data

After the graphical investigation of the data, it is clear that the data is coming from a heavy tailed distribution, and that the GPD could be appropriate for modelling the data. Then, the scale and shape parameters have been determined for the three estimation method; NLS, POT-NLS and WNLS.

The following table 5.1.2 includes the estimates of the scale and shape parameters of the GPD at a threshold =10 using different estimation methods.

Table 5.1.2: Estimates of the scale and shape parameters of the GPD

Estimation method	Estimate of scale	Estimate of shape
NLS	0.9919578	0.5105
POT-NLS	7.9008673	0.2853565
WNLS	7.7000671	0.3238651

Considering the values of the estimated shape and scale parameters in the Anderson Darling goodness of fit in table 5.1.1, and the results of the simulation studies, we could find that WNLS gives the more accurate estimates.

5.3 VaR estimation

Table 5.1.3 shows the different levels of VaR estimates of the insurance dataset. If we take VaR 99.9% for example, it means that there is a 0.001 probability that the losses exceed \$70.68 million and if this happened the expected value of losses will be \$111.19 million. In other words if we set the lower attachment point $r = 10$, then there is a probability 0.001 that the loss exceeds \$70.68 million.

Table 5.1.3: Estimates of VaR of the insurance dataset.

Estimation method	VaR 0.95	VaR 0.99	VaR 0.999	VaR 0.9999	Price
WNLS	9.94	26.23	70.68	164.38	0.1490

5.4 Model checking

Diagnostic plots in figure 5.1.5 show that that the GPD fit the Danish fire loss data in a reasonable way. As the probability and quantile plots are approximately linear and also the density plot shows how the pdf of the GPD fit the empirical data well. The return level plot is concave recommending a shape parameter k to be greater than 0 as estimates of extreme quantile $F^{-1}(p)$ is known as the return level associated with the return period $1/p$, defining $x_p = (1 - p)$, then

$$F_{k,\beta}^{-1}(p) = \begin{cases} \frac{\beta}{k} [x_p^{-k} - 1], & k \neq 0 \\ \beta[-\ln(x_p)], & k = 0 \end{cases} \quad (5.5)$$

So, if $F_{k,\beta}^{-1}(p)$ is plotted against x_p on a logarithmic scale, the plot is linear if $k = 0$, convex if $k < 0$ and concave with no finite bound if $k > 0$. This graph is a return level plot.

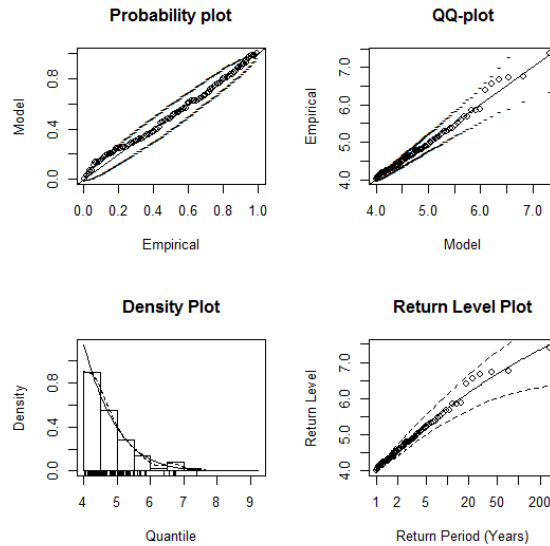


Figure 5.1.5 Diagnostic plots of Danish fire loss data.

6. Summary and Concluding remarks

Throughout this paper, the POT approach was focused on. Three estimation methods for the parameters of the GPD and extreme quantiles were compared: NLS, POT-NLS and WNLS. Different levels of VaR as a tail risk measure were estimated. The aim of this paper is to find out which estimation method performs well on which condition through a simulation study. Then an application of heavy-tailed data was investigated using the POT approach.

The simulation results showed that for estimating the VaR, POT-NLS has almost a bad performance except for small samples generated from the Log-gamma distribution where NLS is the worst while, WNLS and NLS has a better performance in most of the cases.

Reference

- Castillo, E., Hadi, A. S., Balakrishnan, N., & Sarabia, J.-M. (2005). Extreme value and related models with applications in engineering and science.
- Castillo, E., & Sarabia, J. (1994). Extreme value analysis of wave heights. *JOURNAL OF RESEARCH-NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY*, 99, 445-445.
- Choulakian, V., & Stephens, M. A. (2001). Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics*, 43(4), 478-484.
- Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). *An introduction to statistical modeling of extreme values* (Vol. 208): Springer.
- Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 393-442.
- Embrechts, P., Klüppelberg, C., & Mikosch, T. (2013). *Modelling extremal events: for insurance and finance* (Vol. 33): Springer Science & Business Media.
- Fisher, R. A., & Tippett, L. H. C. (1928). *Limiting forms of the frequency distribution of the largest or smallest member of a sample*. Paper presented at the Mathematical Proceedings of the Cambridge Philosophical Society.
- Grimshaw, S. D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics*, 35(2), 185-191.
- He, X., & Fung, W. K. (1999). Method of medians for lifetime data with Weibull models. *Statistics in medicine*, 18(15), 1993-2009.
- Hosking, J. R., & Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29(3), 339-349.
- Hosking, J. R. M. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(1), 105-124.
- Krishnamoorthy, K. (2006). *Handbook of statistical distributions with applications*: Chapman and Hall/CRC.
- McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin: The Journal of the IAA*, 27(1), 117-137.
- Park, M. H., & Kim, J. H. T. (2016). Estimating extreme tail risk measures with generalized Pareto distribution. *Computational Statistics & Data Analysis*, 98, 91-104. doi: 10.1016/j.csda.2015.12.008
- Pickands III, J. (1975). Statistical inference using extreme order statistics. *the Annals of Statistics*, 119-131.
- Reiss, R.-D., Thomas, M., & Reiss, R. (2007). *Statistical analysis of extreme values* (Vol. 2): Springer.

- Smith, R. L. (1984). Threshold methods for sample extremes *Statistical extremes and applications* (pp. 621-638): Springer.
- Song, J., & Song, S. (2012). A quantile estimation for massive data with generalized Pareto distribution. *Computational Statistics & Data Analysis*, 56(1), 143-150.
- Zhang, J. (2007). Likelihood moment estimation for the generalized Pareto distribution. *Australian & New Zealand Journal of Statistics*, 49(1), 69-77.