Estimating the Tail Index for a Truncated Extended Pareto Model

Luka Beverin 04561890

WST795 Research Report

Submitted in partial fulfillment of the degree BSc(Hons) Mathematical Statistics

Supervisor: Dr. G Maribe

Department of Statistics, University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

September 18, 2019

Abstract

In extreme value statistics, the Pareto distribution (PD) is a good fit to relative exceedances above a predetermined threshold. However, the PD typically requires a high threshold, leaving an insufficient amount of data to be modelled. For this reason, Beirlant et al. (2009) introduced an extension of the PD for heavy-tailed distributions, which has been shown to reduce the bias of the extreme value index (EVI) estimates. In practice, data may be truncated due to some study design or natural cause. We therefore purpose a truncated Bayesian estimator of the EVI for the extended Pareto distribution (EPD). We investigate the performance of our estimator in comparison to other estimators through a simulation study. The novelty of this research report is that the fitting of a truncated EPD on truncated extremes has not been considered in literature.

Declaration

I, *Luka Beverin,* declare that this essay, submitted in partial fulfillment of the degree *BSc(Hons) Mathematical Statistics* , at the University of Pretoria, is my own work and has not been previously submitted at this or any other tertiary institution.

____

Luka Beverin

____

Dr Gaonyalelwe Maribe

____

September 18, 2019

Acknowledgements

I would like to extend my sincere gratitude to my supervisor, Dr Gaonyalelwe Maribe, for his guidance, motivation and prodigious knowledge. His positive attitude towards research and problem solving has inspired me to push my own limits. I would also like to express a special thanks and appreciation to my parents for their unwavering love and support throughout my studies.

Table of contents

List of Figures

List of Tables

1 Introduction

When rare events of catastrophic phenomena take place, we are often left wondering about their occurrence and likelihood, and whether anything could have been done to predict and perhaps even prevent the unfortunate outcome. These could include, for instance, the abnormal dry spell in 2015 that left the city of Cape Town at the brink of an emergency water crisis, the destructive North Sea flood of 1953 and the devastating earthquake that crumbled almost the entire Haitian capital city of Port-ou-Prince. For this reason, an engineer may be quite interested in constructing hydraulic structures that are protected against unexpected floods; or an architect in Asia building a high-rise office block that is able to withstand an earthquake of great magnitude. It becomes apparent that the characteristics of interest in all the aforementioned cases are the minimum and maximum values. This gives rise to extreme alue theory (EVT); a section of mathematical statistics pertaining to the behaviour of extreme events within the tails of a probability distribution. EVT's main objective is to predict the statistical probabilities of rare events that are not frequently observed. In practical applications of extreme value analysis, there are two commonly used methods, namely the Block Maxima (BM) approach and the Peaks-over-Threshold (POT) method. In recent years the POT method, developed by hydrologists, has been the most popular approach in modelling extreme events. In the POT approach, one would select a sufficiently high threshold such that any observation above this threshold would be considered an extreme value. Pickands III (1975) showed that the probability distribution of those selected observations is approximately a generalized Pareto distribution (GPD). However, Beirlant et al. (2009) noticed that while the GPD fits well, one would require relatively high thresholds, which would subsequently mean that only a small upper proportion of the data would be modelled. In the paper, Beirlant et al. (2009) proposes an extended version of the (G)PD, called the extended Pareto distribution (EPD), which would ultimately allow a larger portion of data to be modelled, whilst reducing bias and maintaining stability within the parameters of the distribution. Inferences about the tail behaviour of the underlying distribution is better understood through the extreme value index (EVI). The EVI is a parameter that depicts tail heaviness, and as such, estimation of this parameter has been at the focal point of EVT. However, it is often the case when examining the tail characteristics of large data sets that truncation on extremes can arise. This may be due to physical constraints placed on the data or in some instances the measurement process itself caches extreme values of interest, thus, affecting the procedures used in estimation of the EVI. Truncated tails are common in data sets on diamond sizes, earthquake magnitudes and finance (see Beirlant et al. (2017) and Aban et al. (2006)).

In this research report, we propose a Bayesian estimation method for the EVI and high quantiles under a truncated heavy-tailed distribution. The Bayesian method is constructed by treating the unknown parameters as random variables and specifying a non-informative prior. Using Markov Chain Monte Carlo methods we are able to derive a posterior distribution that provides inferences on the upper tail. This research report is concerned with the behaviour of the proposed truncated Bayesian EPD estimator in comparison to other popular estimation methods. The key comparison will be done through analysing the theory of both models and conducting a simulation experiment.

This paper is structured as follows. Chapter 2 discusses the general background theory of extreme value analysis with links to truncation and Bayesian inference. In Chapter 3 we discuss various estimation techniques of the EVI and the second-order parameter ρ. Chapter 4 describes the EPD. In Chapter 5, we observe the results of our simulation study and discuss the performance of the proposed estimator. Chapter 6 concludes the research report.

2 Probabilistic Background of Extreme Value Theory

2.1 Applications of Extreme Value Theory

Since its inception, EVT has played a vital role in various fields and continues to demonstrate its functionality in solving problems in a material setting. In this section, we give some common applications of EVT. We refer to Coles et al. (2001) and De Haan and Ferreira (2007) for more detailed literature dedicated to EVT and its applications.

Financial applications

In the past few decades, investors, risk managers and bankers have become troubled with circumstances that occur under extreme market conditions. Portfolio managers may be interested in knowing the distribution of large losses and therefore they make use of the Value-at-Risk (VaR) quantity. The VaR of a portfolio is a statistic that measures the level of financial risk of an investment. Applying EVT to finance is an effective additional risk measure because it produces more appropriate distributions to fit extreme data when compared to other VaR methods (see Bensalah (2000)).

Structural engineering

In the field of structural engineering, it is known that wind speeds, seismic activity and design loads are important factors that have to be considered for design purposes. Statistical inferences about the maxima and minima of the relevant extremes is crucial for bettering the design procedures in terms of cost and safety. For more in-depth analysis of this problem, the reader is referred to Castillo (1988).

Insurance applications

The (re)insurance industry has recently become one of the most prominent areas where EVT is applicable. This growing interest is due to natural and man-made catastrophes that lead to enormous losses within the industry. Here, EVT is useful in modelling the loss severity in the tails of the distributions. For a relevant discussion, we refer the reader to Corradin and Verbrigghe (2001).

2.2 Quantiles and tail estimation

When conducting basic statistical analysis, the distribution function $F(x) = P(X \leq x)$ is usually studied, however, in EVT quantiles prove more useful in certain conditions. Denote the quantile function and tail quantile function respectively:

$$
\begin{aligned}
Q(p) &:= \inf\{x : F(x) \geq x\}, \qquad 0 < p < 1, \\
U(v) &:= \inf\{x : F(x) \geq 1 - 1/v\}.
\end{aligned}
$$

Note that $U(v) = Q(1-1/v)$ links the quantile function and tail quantile function. The study of extreme quantiles is directly linked to the estimation of the tail of the distribution. The heaviness of the tail is ruled by the EVI, which makes estimation of this parameter indispensable in the context of EVT. Later in the report, we mention various literature dedicated to the estimation of the EVI parameter.

2.3 Truncation

In a practical setting, large data sets can be subjected to truncation effects. We denote W as the underlying parent random variable with distribution function $F_W(w) = P(W \leq w)$ and RTF $\overline{F}_W(w) = 1 - F_W(w)$. The quantile function is $Q_W(p) = \inf\{w : F_W(w) \geq p\}$ and the tail quantile function is $U_W(u) = Q_W(1 - 1/u)$, where $0 < p < 1$ and $u > 1$. In this research report we consider truncation from the right-hand side from which independent and identically distributed data $X_1, X_2, ..., X_n$ are observed from some distribution with trucation value $T > 0$. In essence,

$$X =_d W | W < T.$$

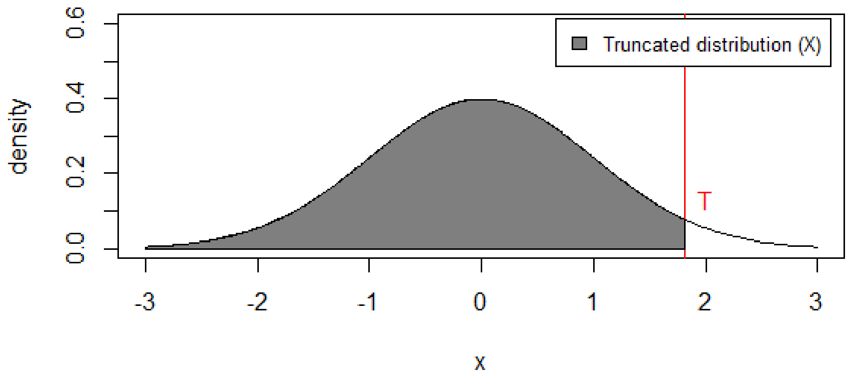


Figure 1: Schematic representation of a truncated standard normal distribution from the right-hand side at truncation level $T = Q(0.965)$.

The corresponding truncated distribution function is denoted $F_T(x) = F(X \leq x)$ with RTF $\overline{F}_T(x) = P(X > x)$. The tail quantile function is given as $U_T(v) = Q_T(1 - 1/v)$. It follows that

$$\begin{aligned}
\overline{F}_T(x) &= 1 - \frac{F_W(x)}{F_W(T)} = \frac{F_W(T) - F_W(x)}{F_W(T)} = \frac{\overline{F}_W(x) - \overline{F}_W(T)}{1 - \overline{F}_W(T)} = (1 + D_T)\overline{F}_W(x) - D_T \\
U_T(v) &= U_X\left(\frac{v}{F_W(T)}\left[1 + vD_T\right]^{-1}\right) \\
&= U_X\left(\frac{v}{F_W(T)}\left[1 + \frac{1}{vD_T}\right]^{-1}\right)
\end{aligned}$$

Where $D_T = \overline{F}_W(T)/F_W(T)$ is equal to the odds ratio of the truncated probability mass to the non-truncated distribution X.

2.4 Bayesian methodology in Extreme Value Theory

One way to draw statistical inferences about the tails of a distributions is to consider a Bayesian framework. At the time of this report, there exists very little literature that links Bayesian techniques to estimation of the EVI. Nevertheless, Coles et al. (2001) and Beirlant et al. (2006) are examples of references that provide a comprehensive approach of Bayesian methodologies in the EVT framework. A Bayesian analysis is set up as the following. Let $\mathbf{x} = (x_1, ..., x_n)$ denote realizations of a random variable X according to a distribution with density function $f(x|\theta)$. It is in our interest to formulate beliefs about the parameter θ. We denote $\pi(\theta)$ as the prior distribution for θ. Specifiying a prior distribution deviates significantly from the inferential framework that is later seen in likelihood-based estimation methods. The parameter θ is observed as a random variable for which the prior distribution $\pi(\theta)$ consists of the parameter distribution that is prior to any additional data. We denote $f(\mathbf{x}|\theta)$ as the likelihood for the parameter θ. In the case of independence we may write the likelihood as $\prod_{i=1}^{n} f(x_i|\theta)$. Bayes' theorem states that

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{\omega} f(\mathbf{x}|\theta)\pi(\theta)d\theta} \propto f(\mathbf{x}|\theta)\pi(\theta), \tag{1}$$

where ω represents the parameter space. Bayes' theorem proves to be an extremely useful tool that allows statisticians to transform a primary set of assumptions about θ, represented by the prior distribution $\pi(\theta)$, into a posterior distribution that is denoted $\pi(\theta|\mathbf{x})$. The posterior distribution is proportional to the product of the likelihood and the prior distribution. This makes calculations numerically challenging, and so we depend on Markov Chain Monte Carlo (MCMC) methods to approximate the posterior. The benefit of the posterior distribution is that it produces additional information that is provided by the data $\mathbf{x}$. Morever, estimation of the parameter θ can be obtianed by calculating the mode and mean of the posterior distribution. Bayesian analysis relies on the specification of a prior $\pi(\theta)$ and as a result many statisticians have opposed its functionality. This viewpoint is understood because when data is scarce, experts are usually unable to construct informative opinions on a prior that benefits from additional information. However, in such a situation, a statistician should make use of a non-informative/objective prior. In the context of EVT, the most popular non-informative priors are Jeffreys' and the maximal data information (MDI) prior.

2.4.1 Jeffreys' prior

We define the Fisher information matrix

$$I_{ij}(\theta) = E\left(-\frac{d^2 \ln f(\mathbf{x}|\theta)}{d\theta_i d\theta_j}\right) \qquad i, j = 1, 2, ..., p$$

where p is the dimension of θ. Then Jeffreys' prior (Jeffreys (1998)) is

$$J(\theta) \propto \det(\pi(\theta))^{\frac{1}{2}} \tag{2}$$

The appealing property of this prior is that it is invariant under reparameterization.

2.4.2 MDI prior

The MDI prior (Zellner (1971)) is a prior that increases in average information. This is only the case if the data is maximized using the likelihood function. The MDI is defined as

$$\pi(\theta) \propto \exp E\{\log f(\mathbf{x}|\theta)\}. \tag{3}$$

2.5 Limiting Behaviour of Block Maxima Approach

The Block Maxima (BM) method is considered as the first approach in modelling extreme observations. This method entails dividing a series of observations into intervals of equal parts such that attention is constricted to the maximum observation within each interval. A more formal definition of the BM method is based on Ferreira and de Haan (2013).

Let $\tilde{X}_1, \tilde{X}_2, \tilde{X}_3, ...$ be independent and identically distributed random variables with continuous and invertible distribution function F. Define for m = 1,2,... and i = 1,2,..., k the block maximum.

$$X_i = \max_{(i-1)m<j\leq im} \tilde{X}_i$$

The sequence is divided into k blocks of size m. Hence, the number of observations in the sequence is written as $n = m$ x k. Figure 1 serves as an illustration.

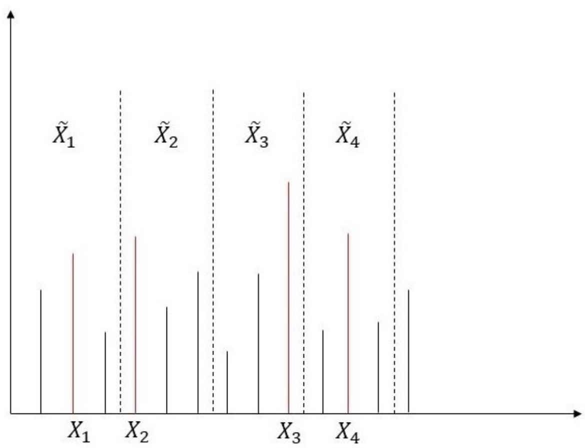


Figure 2: BM method. The X_i's would be considered as extremes values of a given data set.

A main concern in EVT is the exploration of a limiting distribution that models the maxima. More precisely, we are searching for limiting distributions and characterizations of F which ensure that there exists a sequence of real norming constants $a_m > 0$ and b_m such that for all values of x

$$\lim_{m\to 0} P\Big(\frac{X_i - b_m}{a_m} \leq x\Big) = G(x), \qquad i = 1, 2, ..., k, \tag{4}$$

where G is some non-degenerate distribution. If condition (4) holds, then we say that F in the maximal domain of attraction of G and we write $F \in \mathcal{D}_{\mathcal{M}}(G)$. The problem of finding all possible non-degenerate distributions G is better known as the extremal limit problem. It was Tippett's work with Sir Ronald Aylmer Fisher 1928 that lead to the discovery of three asymptotic results, which described the distributions of extreme values. Namely Fisher-Tippett Types I, II and III. These have become known since as the Fréchet, Gumbel and Weibull distributions. In 1936 Richard Von Mises studied EVT, giving in particular the von Mises conditions - simple and sufficient conditions of the limit of a distribution in order to allow a situation in which extreme observations occur. These conditions prompted extremal domains of attractions for some extreme value distribution (EVD), for which Gnedenko (1943) then progressed towards the first meticulous treatment of the fundamental limit of extremes. The search for characterizations of F such that (4) holds is referred to as the first-order conditions and is discussed later in Section 3.2.

Theorem 2.1 (Fisher-Tippett-Gnedenko Theorem)**.** *If condtion (4) holds for some non-degenerate distribution G, then G will belong to one of the following three distribution families:*

$$\text{\textit{Gumbell:}} \quad \Lambda_\alpha(x) = e^{-e^{-x}}, \qquad x \in \mathbb{R} \tag{5}$$

$$\text{\textit{Fréchet:}} \quad \Phi_\alpha(x) = e^{-x^{-\alpha}}, \qquad x > 0, \alpha > 0 \tag{6}$$

$$\text{\textit{Weibull:}} \quad \Psi_\alpha(x) = e^{-(-x)^{-\alpha}}, \qquad x \leq 0, \alpha < 0 \tag{7}$$

The distributions are often referred to as the three domains of attractions. Real case scenarios and examples can easily demonstrate different parent tail characteristics that have a variety of distributions attached to their respective limits. Notably, the maximums do not appear to converge in distribution to a normal distribution. This is due to the intrinsic properties that are observed in the distribution of the maximum values.

2.5.1 The Generalized Extreme Value Distribution

The generalized extreme value (GEV) distribution, introduced by Jenkinson (1955) and Von Mises (1936), unites the Gumbel, Fréchet and Weibull distributions into a single parametric representation

$$G_\xi(x) := \begin{cases} e^{-(1+\xi x)^{-\frac{1}{\xi}}}, & \text{if} \quad \xi \neq 0, \quad 1 + \xi x > 0, \\ e^{-e^{-x}} & \text{if} \quad \xi = 0, \quad x \in \mathbb{R}. \end{cases} \tag{8}$$

This generalization is obtained by introducing a new parameter, $\xi = 1/\alpha$. The quantity ξ is a tail shape parameter and is referred to as the EVI, which is a primary parameter in the context of extreme value analysis. This parameter determines the shape of the survivor function $\overline{F}(x) := 1 - F(x)$ and moments of extremes.

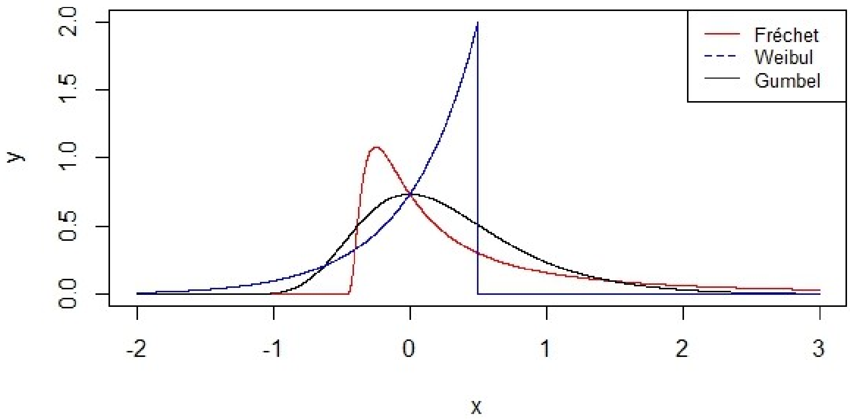


Figure 3: The GEV distribution functions; Fréchet ($\xi > 0$), Weibull ($\xi < 0$) and Gumbel ($\xi = 0$) for $u = 0$ (location) and $\sigma = 0.5$ (scale)

We observe that the tail shape parameter $\xi = 0$ yields the Gumbel domain, which is a class of unbounded thin-tailed distributions that includes distributions such as the exponential and gamma. Furthermore, distributions associated with $\xi < 0$ form part of the Weibull family of distributions. It is noticeable that the Weibull distribution class has finite upper bounds which incorporates the uniform distribution amongst others (see Beirlant et al. (2006)). Finally, in the case where $\xi > 0$, the corresponding distribution is the Fréchet. The Fréchet distribution is a class of heavy-tailed distributions, which is most useful in finance, hydrology and insurance. Note that $\xi > 0$ situates us in the Fréchet domain and it has been shown that for a truncated model in such a domain the EVI exhibits a value of $\xi = -1$, which produces the finite upper bounds visible in the Gumbel domain. See Figure 2.8 in (Beirlant et al. (2006)).

2.6 The Peaks-Over-Threshold Method

The BM approach is favoured in the environmental sciences where cyclic patterns are repeatedly observed, for example, yearly maximum flood levels and yearly maximum temperatures. This approach is appropriate when our data set consists of a set of maxima. However, in statistical analysis, the BM method often leads to a loss of information when other large samples are being considered. Thus, we are not only interested in the behaviour of the maxima, but also in the behaviour of observations that exceed some high threshold t. This methodology is referred to as the Peaks-over-Threshold (POT) and was introduced by Davison and Smith (1990). The method aims to remedy the drawbacks of the BM approach. Here, emphasis is placed on modelling the relevant high observations. Figure 4 serves as an illustration.

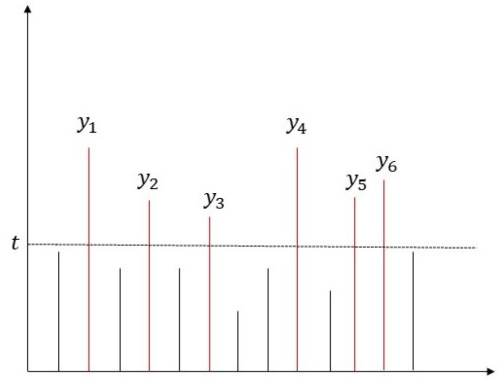


Figure 4: POT method.

It was shown in Pickands III (1975) that the exceeses above the threshold t have an approximate limiting distribution in the form of the generalized Pareto distribution (GPD). The GPD has since then been further developed by, amongst others, Hosking and Wallis (1987), Davison and Smith (1990) and Smith (1984).

Theorem 2.2 (Picklands-Balkema-de Haan). *Let X be a random variable with continous distribution fuction (cdf) F and right endpoint $x_F = sup\{x : F(x) < 1\}$. An exceedance of a predetermined threshold t occurs when $X > t$. The excess over t is defined as $y = X_i - t, \quad i = 1, 2, 3, .., n$. Then, for a high threshold t, the excess distribution function of the random variable X over threshold t is given by,*

$$\begin{aligned} F_t(y) = P(X - t \leq y | X > t) &= \frac{P(X - t \leq y, X > t)}{P(X > t)} \\ &= \frac{F(y + t) - F(t)}{1 - F(t)}. \end{aligned} \tag{9}$$

for $0 \leq y \leq x_F - t$. For sufficiently high threshold t, $F_t(y)$ can be approximated by the generalized Pareto distribution (GPD). In general, the GPD is defined as

$$G_{\xi,\sigma}(x) = \begin{cases} 1 - (1 + \frac{\xi x}{\sigma})^{-\frac{1}{\xi}}, & if \quad \xi \neq 0, \\ 1 - e^{-\frac{x}{\sigma}}, & if \quad \xi = 0 \end{cases} \tag{10}$$

where $\sigma > 0$ is the scale factor, and $x \geq 0$ if $\xi \geq 0$ and $0 \leq x \leq -\sigma/\xi$ if $\xi < 0$.

When $\sigma = 1$, the representation is referred to as the standard GPD. Similar to the GEV distribution, $\xi = 1/\alpha$ is the shape parameter, with α being the tail index. In particular, the duality between the GPD and GEV distribution means that the EVI ξ is dominant in determinig inferences of the tails. The GPD takes the form of various distributions with varying shape parameter. When $\xi = 0$, the GPD takes the form of an exponential distribution with mean σ. In the case of $\xi > 0$, the GPD reduces to the reparametrised Pareto distribution. The Pareto-type tails are typically heavier-tailed than the exponential and appear in various branches of insurance and finance. If $\xi < 0$ then the distribution of excesses has a finite upper limit. The distribution (10) is known if the parent F distribution is known. Since in practical applications we seldom know the parent distribution, there has been a search for approximations that are broadly applicable for high thresholds. To better understand the GPD, we consider an example from Coles et al. (2001), which is of the Fréchet model $F(x) = e^{-x^{-\alpha}}$, for $x > 0$ and $a > 0$. Recall that EVI $1/\alpha = \xi > 0$. Hence,

$$\frac{1 - F(u+y)}{1 - F(u)} = \frac{1 - e^{-(u+y)^{-1}}}{1 - e^{-u^{-1}}} \sim \left(1 + \frac{y}{u}\right)^{-1}$$

as $u \to \infty$ and for all $y > 0$. The limit distribution of threshold exceedances for the Fréchet model corresponds to the GPD with $\xi = 1$ and $\sigma = u$. The same procedure can be repeated with other distributions such as the uniform, exponential and many others. For the remainder of this research report we will pay strict attention to the case of $\xi > 0$.

2.6.1 Threshold seletion

The issue of selecting an appropriate threshold t is ultimately a trade-off between bias and variance. A threshold that is too high will only produce few excesses with which the model can estimated, leading to high variance; too low a threshold is likely to alter the asymptotic bias of the model. A common method to approach the issue is to make use of an explanatory technique that is carried out before estimation of the model. The method is based on the mean of the GPD. If Y has a GPD with location parameter σ and shape parameter ξ then,

$$\mathrm{E}(Y) = \frac{\sigma}{1 - \xi}, \quad \xi < 1. \tag{11}$$

When $\xi \geq 1$ the mean is infinite. It is shown in Coles et al. (2001) that by (11), the expected value of the excesses above the threshold t is $\mathrm{E}(X - t | X > t)$. The graphical interpretation of this method is known as the mean excess plot and is introduced in Section 2.8.2. Another approach to choosing a threshold is to asses the stability of the parameter estimates when the model is fitted across an array of thresholds. For this method we refer the reader to Section 4.3.4 of Coles et al. (2001) for a more detailed discussion.

2.7 First and Second-Order Framework

In the context of extreme values, we are often only concerned about the tails of a distribution and not the entire distribution. As such, we are interested in making inferences about the tail and require more knowledge about the EVI ξ, which depicts the heaviness and upper tail behaviour of the underlying distribution. The study of first and second-order regular variation is crucial in understanding the asymptotic theoretical properties of EVI estimators. The aim of this chapter is to introduce the required background mathematical tools required to approach regular variation and to discuss the first and second-order framework. In Section 3.2 we detail the first-order condition, otherwise known as the domain of attraction problem. The Hall class of distributions is introduced in Section 3.2.3 and lastly, second-order variation is discussed in Section 3.3. We first state the GEV distribution obtained in 8:

$$G_\xi(x) := \begin{cases} e^{-(1+\xi x)^{-\frac{1}{\xi}}}, & \text{if} \quad \xi \neq 0, \quad 1+\xi x > 0, \\ e^{-e^{-x}} & \text{if} \quad \xi = 0, \qquad x \in \mathbb{R}, \end{cases} \tag{12}$$

where F is in the maximal domain of attraction of G and we write $\mathcal{D}_{\mathcal{M}}(G)$.

2.7.1 Functions of regular variation

In this subsection, we define a result from Section 2.9.4 of Beirlant et al. (2006) which is necessary to thoroughly understand the mathematical derivations that follow. Regular variation is a fundamental property from a class of functions which are extensively useful in several mathematical applications.

Definition 1 (Beirlant et al. (2006)). *Let f be an ultimately positive and measurable function on $(0, \infty)$. We say, f is regular varying if and only if there exists $p > 0$ for which*

$$\lim_{x \uparrow \infty} \frac{f(xt)}{f(x)} = t^p \qquad \textit{for all} \quad t > 0 \tag{13}$$

We call p the index of regular variation and write $f \in \mathcal{R}_p$. The function f will be called slowly varying in the case of $p = 0$. A function of slow variation will be denoted by the symbol ℓ. Finally, the class of regular varying functions is denoted by $\mathcal{R}$.

2.7.2 First-order condition

In EVT we pay special attention to the convergence of distribution of the maxima X_i. This leads us to the domain of attraction problem: the search for sufficient and necessary conditions on the distribution of X such that the limit distribution specified in (4) holds. Simply put, we would like to know for what kind of distributions of X are the maximums attracted to some specific extreme value distribution $G_\xi(x)$. Firstly, we require some domain of attraction condition which has to be satisfied by the underlying distribution F. Such a condition is derived in Beirlant et al. (2006) and relies on the well known Helly-Bray theorem in Patrick (1995). For some positive function a and any $u > 0$,

$$\lim_{x \to \infty} \left(\frac{U(xu) - U(x)}{a(x)} \right) =: h(u) \quad \text{exists}, \qquad (C) \tag{14}$$

with the limit function h not being identically equal to zero.

2.7.2.1 General tail ($\xi \in \mathbb{R}$)

The first-order condition stated in 14 is sufficient for a general case, however, extreme value analysis requires a more informative condition which is indexed by the parameter ξ. Let's consider a general tail $\xi \in \mathbb{R}$. Then the following extended regular variation property (de Haan (1984)), denoted ERV_ξ, is a necessary and sufficient condition for $F \in \mathcal{D_M}(G_\xi)$:

$$h_\xi(u) := \lim_{x\to\infty} \frac{U(xu) - U(x)}{a(x)} = \int_1^u s^{\xi-1} ds = \begin{cases} \frac{u^\xi - 1}{\xi} & \text{if} \quad \xi \neq 0 \\ \ln u & \text{if} \quad \xi = 0, \end{cases} \qquad (C_\xi) \tag{15}$$

for all $u > 0$. The above result indicates that under the condition (C_ξ) the possible limits are described by the one-parameter family, indexed by ξ. We can also say that if F satisfies $C_\xi(a)$ with auxilliary function a then $F \in \mathcal{D_M}(G_\xi)$, where G_ξ is some extreme value distribution.

2.7.2.2 The Fréchet-Pareto case ($\xi > 0$)

The most classical case in EVT is the Fr/'echet-Pareto case, which is a heavy-tailed distribution obtained with $\xi > 0$ in (12). We can take $U(x) = x^\xi \ell_U(x)$ with the slowly varying function ℓ_U such that (C_ξ) is also satisfied. Then, for $x \uparrow \infty$,

$$\begin{aligned} \frac{U(xu) - U(x)}{a(x)} &= \frac{(xu)^\xi \ell_U(xu) - x^\xi \ell_U(x)}{a(x)} \\ &= \frac{x^\xi \ell_U(xu)}{a(x)} \left(\frac{\ell_U(xu)}{\ell_U(x)} u\xi - 1 \right), \end{aligned}$$

choosing auxiliary function $a(x) = \xi x^\xi \ell_U(x) = \xi U(x)$ we have that

$$\frac{U(xu) - U(x)}{a(x)} \sim \frac{u^\xi - 1}{\xi}.$$

Note that distributions for which $U(x) = x^\xi \ell_U(x)$ are referred to as Pareto-type distributions. It is shown in Beirlant et al. (2006) that the Pareto-type tail, which is the first-order condition (C_ξ) for $\xi > 0$ is equivalent to:

$$\frac{1 - F(tx)}{1 - F(t)} \to x^{-1/\xi} \quad \text{as} \quad t \to \infty \quad \text{for any} \quad x > 1,$$

which may be interpreted as relative conditional excess distribution:

$$\overline{F}(x) = \frac{1 - F(tx)}{1 - F(t)} = P\left(\frac{X}{t} > x | X > t \right) \to x^{-1/\xi} \quad t \to \infty, \quad x > 1.$$

This is specifically saying that $x^{1/\xi}(1 - F(x))$ is slowly varying. Hence, there exists a slowly varying function $\ell_F(x)$ such that $\overline{F}(x) = x^{-1/\xi}$. Moreover, $F \in \mathcal{D_M}(\Phi_\alpha)$ if and only if there exists a slowly varying function $\ell_F(x)$ for which $\overline{F}(x) = x^{-1/\xi}$. Indeed, it is not always rudimentary to compute the quantile tail function of a distribution. So, it could be recommended to express the relation between the first-order condition (C_ξ) and the underlying distribution F. Beirlant et al. (2006) showed that the Pareto-type distributions can be formulated in terms of the underlying distribution F and as well in terms of the tail quantile function. The link between the two is dependant on the de Bruyn conjugate. This yields full equivalence between the two statements

$$1 - F(x) = x^{-1/\xi}\ell_F(x) \qquad \text{and} \qquad U(x) = x^{\xi}\ell_U(x)$$

where the two slowly varying functions ℓ_F and ℓ_U are linked together via the de Bruyn conjugation.

2.7.2.3 Hall class of distributions

The Hall class of Distributions Hall and Welsh (1985) is often referred to as the corresponding sub-class of useful heavy-tailed Pareto-Type Distributions and includes distributions such as the Burr, Fréchet and student-t. We suppose that is no loss of generality in assuming that the underlying distribution F satisfies $F(0) = 0$. Then the survival function can be written as:

$$\overline{F}(x) = Cx^{-\alpha}(1 + Dx^{-\beta} + o(x^{-\beta})), \qquad x \to \infty, \tag{16}$$

where $\alpha > 0$, $C > 0$, $\beta > 0$ and $D \in \mathbb{R}$. The flexible requirement of $1 - F(x) = x^{-\alpha}\ell_F(x)$, plays an important role in constructing estimators for EVI $\xi > 0$. The tail quantile function of $\overline{F}(x)$ is then

$$U(x) = Cx^{\xi}(1 + Dx^{-\beta} + o(x^{-\beta})).$$

2.7.3 Second-order condition

2.7.3.1 General tail ($\xi \in \mathbb{R}$)

We are now interested in the second-order condition, specifying the convergence rate for the distribution of the maxima X_i to the extreme value distribution. We start by assuming the existence of a function $A(x) \to 0$ as $x \to \infty$, which is possibly not changing in sign, such that

$$\lim_{x\to\infty} \frac{\frac{U(xu)-U(x)}{a(x)} - \frac{u^{\xi}-1}{\xi}}{A(x)} = H_{\xi,p}(u) := \frac{1}{p}\left(\frac{u^{\xi+p}-1}{\xi+p} - \frac{x^{\xi-1}}{\xi}\right) \tag{17}$$

for all $u > 0$, where $\rho \leq 0$ is a second-order parameter that controls the rate of convergence of maximum values towards the limit law in (12), for a general $\xi \in \mathbb{R}$. The function U is said to be of second-order extended regular variation, and we write $U \in 2ERV_{\xi,\rho}$. Written more definitely by Alves et al. (2007),

$$H_{\xi,\rho}(u) = \begin{cases} \frac{1}{\rho}\left(\frac{u^{\rho}-1}{\rho} - \ln u\right) & \text{if} \quad \xi = 0, \rho = 0 \\ \frac{1}{\xi}\left(u^{\xi}\ln u - \frac{u^{\xi}-1}{\xi}\right) & \text{if} \quad \xi \neq 0, \rho = 0 \\ \frac{\ln^2 x}{2} & \text{if} \quad \xi = \rho = 0. \end{cases} \tag{18}$$

Noting that $|A| \in RV_{\rho}$.

2.7.3.2 Heavy tail ($\xi > 0$)

In order to measure the rate of convergence for heavy-tails, Alves et al. (2007) considered one of the following condtions:

$$\lim_{x\to\infty} \frac{\frac{U(xu)-U(x)}{a(x)} - u^{\xi}}{\tilde{A}(x)} = u^{\xi}\frac{u^{\tilde{\rho}}-1}{\tilde{\rho}} \Longleftrightarrow \lim_{x\to\infty} \frac{\ln U(xu) - \ln U(x) - \xi\ln u}{\tilde{A}(x)} = \frac{u^{\tilde{\rho}}-1}{\tilde{\rho}} \tag{19}$$

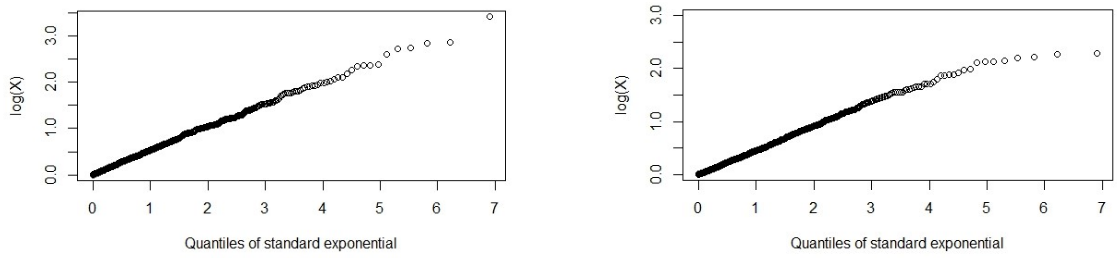
for all $x > 0$, where $\tilde{\rho} \leq 0$ is again as before the rate of convergence of maximum values towards the limit law in (12). The function U is also of second-order extended regular variation, and we write $U \in 2ERV_{\xi,\tilde{\rho}}$. We also remark that $|\tilde{A}(x)| \in RV_{\tilde{\rho}}$.

2.8 Graphical tools

When observing extreme values, a statistician may want to use graphics that are able to explain the data in a clear and efficient way. Graphical tools are essential in providing statistical information about distributions and their tails. In EVT, the most common graphical tools used are quantile-quantile (QQ) plots and mean excess (or mean residual life) plots.

2.8.1 Q-Q plots

A QQ plot is a probability plot created by plotting two sets of quantiles against each other. This is done so that we are able to see whether two probability distributions are related to each other. If both sets of quantiles come from the same distribution, we should expect to see the points forming a straight line. The linearity is easily checked by eye and furthermore, can be quantified by means of a correlation coefficient. Moreover, this graphical tool makes it possible to assess how well the selected distribution fits the tail of the empirical distribution.



(a) QQ-plot of a simulated Pareto distribution

(b) QQ-plot of a simulated truncated Pareto distribution

Figure 5: QQ-plots of truncated and non-truncated Pareto distributions

It is evident that the left image (a) of Figure 5 exhibits a straight line, which is indicative of an underlying Pareto distribution. Notice how the right image (b) of Figure 5 curves downwards at its largest values. This curvature represents a truncated distribution. Apart from informing us about which distribution the data is likely to be realized from, QQ-plots visual simplicity is also able to aid statisticians in determing whether a distribution is subjected to truncation effects.

2.8.2 Mean excess plots

In practice, the main use of mean excess plots is to aid in selecting a sufficiently high threshold t, for which the exceedances above t are approximated by the GPD. Let X be a random variable given threshold t, then the mean excesses (e) function is of X is defined as:

$$e(t) = \mathrm{E}(X - t|X > t)$$

assuming that $\mathrm{E}(x) < \infty$. The estimates of the mean excesses for a representative sample $x_1, x_2, ..., x_n$ are given by

$$e_{k,n} := \hat{e}_n(x_{n-k,n}) = \frac{1}{k}\sum_{i=1}^{k} x_{n-i+1,n} - x_{n-k,n} \tag{20}$$

where the emperical function $\hat{e}_n$ is plotted at different threshold values $t = x_{n-k,n}$, $k = 1, ..., n-1$, the $(k+1)$-largest observation. The emperical function $\hat{e}_n$ is obtained by averaging data that is larger than t and then subtracting t.

3 Estimation procedures

The aim of this chapter is to consider the estimation of the EVI ξ, which is the tail parameter that depicts the heaviness of the underlying distribution. The most well-known and natural estimate is the maximum likelihood (ML). However, ML estimates for ξ and σ in the GPD are known to be difficult to manage (see Grimshaw (1993)), and as such, several authors have proposed different estimators to the ML. In this chapter we discuss the ML approach to estimation and the extremely popular Hill estimator (1975); an estimator that is applicable to heavy-tailed distributions. Afterwards we define the method of moments (MOM) estimator and in addition we provide an algorithm that is used to estimate the second-order parameter ρ that was discussed in Section 3.3. We also discuss estimation through Bayesian inference, which is the method used in our proposed estimator for the truncated EPD. Some estimators that are also worth mentioning are the probability weighted moments (PWM) and the elemental percentile method (EPM). Hosking and Wallis (1987) introduced the PWM estimators; popular in its approach for small samples but lacking in ability to estimate strong heavy-tails. The EPM, suggested by Castillo and Hadi (1997), overcomes some of the complications associated with the ML and PWM estimators by combining the robustness of the PWM estimator and the flexibility of the ML estimator.

3.1 Maximum Likelihood

The method of maximum likelihood (ML) is seen as the natural estimate for the parameters of extremes. The ML estimates are obtained by maximizing the log-likelihoods and solving the equations. The absence of an explicit formula entails that underlying asymptotic properties have to be shown for the different parameters. In the case of $\xi > -0.5$, Smith (1985) proved that the usual properties of consistency, asymptotic efficiency and asymptotic normality hold. Furthermore, Zhou (2009) proves that that the first-order condition implies asymptotic normality for $\xi > 1$. Retaining our focus to the POT method, we are interested in estimating the parameter ξ of the GPD. Suppose that the values $y_1, ..., y_k$ are the k excesses of a threshold t. The log-likelihood for such a sample for the GPD is given by:

$$\log L(\sigma, \xi|y) = \begin{cases} -k\log\sigma - \left(\frac{1}{\xi}+1\right)\sum_{i=1}^{k}\log\left(1+\frac{\xi y_i}{\sigma}\right) & \text{if} \quad \xi \neq 0 \\ -k\log\sigma - \frac{1}{\sigma}\sum_{i=1}^{k} y_i & \text{if} \quad \xi = 0. \end{cases} \tag{21}$$

Thereafter the usual confidence intervals and standard errors are acquired from standard likelihood theory.

3.2 Hill estimator ($\xi > 0$)

The popular Hill 1975 estimator is an estimator for the Pareto-type tail 1-F, that is $\overline{F} \in \mathcal{R}_{-1/\xi}$ with $\xi > 0$. Beirlant et al. (2006) considers various ways to derive the Hill estimator, however, we will approach the estimator with a probabilistic view. Let $\xi > 0$, such that we are in the Fréchet domain of attraction. From Section 2.7.2.2 we know that

$$\frac{1-F(tx)}{1-F(t)} = P\left(\frac{X}{t} > x | X > t\right) \to x^{-1/\xi} \quad t \to \infty, \quad x > 1. \tag{22}$$

As before, we interpret (22) as the relative conditional excess distribution approximating to the strict Pareto distribution. We order the observations $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$ and denote $Y_j = X_i/t$ where i is the index of the j-th exceedance in the original sample and $j = 1, ..., N_t$. Then the log-likelihood conditionally on N_t is

$$\log L(Y_1, ..., Y_{N_t}) = -N_t \log\xi - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^{N_t} \log Y_j.$$

Deriving the log-likelihood with respect to ξ:

$$\frac{d\log L}{d\xi} = -\frac{N_t}{\xi} + \frac{1}{\xi^2} \sum_{j=1}^{N_t} \log Y_j,$$

leads to

$$\hat{\xi} = \frac{1}{N_t} \sum_{j=1}^{N_t} \log Y_j.$$

We choose a threshold $t = X_{n-k,n}$ that is an upper order statistic such that $N_t = k$. The Hill estimator based on $k+1$ upper order statistics is

$$H_{k,n} := \frac{1}{k} \sum_{j=1}^{k} \log \frac{X_{(i)}}{X_{(k+1)}} = \frac{1}{k} \sum_{j=1}^{k} \log X_{n-j+1,n} - \log X_{n-k,n} \quad (23)$$

for $k = 1, ..., n-1$. Simply put, the Hill estimator is written as the average of scaled *log*-spacings. It was shown by Mason (1982) that the Hill estimator $H_{k,n}$ is a consistent estimator for ξ. This is to say that if $(k_n)_{n\in\mathbb{N}}$ is an intermediate sequence then $k_n \to \infty$ and $k_n/n \to 0$. The performace of $H_{k,n}$ depends heavily on the choice of k, since for every choice of k we obtain a different estimator for ξ. This makes the appropriate choice of k a difficult one. As k increases, the bias will increase since the tail satisfies less and less of the convergence criterion, while if less data is used, the variance increases. The complication gives rise to the Hill plot: $\{(k, H_{k,n}) : 1 \leq k \leq n-1\}$. The plotting of estimates $H_{k,n}$ against k such that the parameter ξ is inferred from a stable region in the graph.

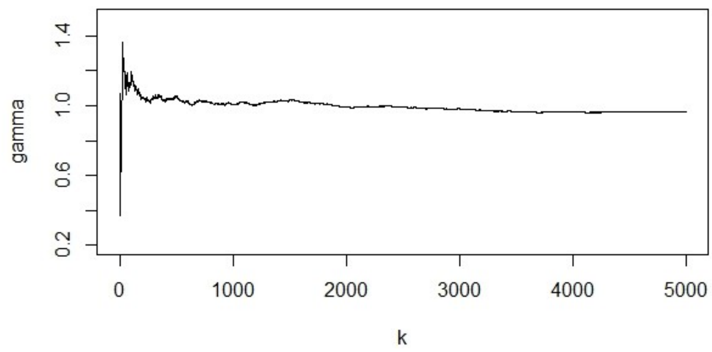


Figure 6: Hill plot of 5000 observations of Pareto Distribution, $\xi = 1$.

The plot in Figure 6 is a Hill plot of 5000 independent and identically distributed observations from a Pareto distribution with $\xi = 1/\alpha = 1$. It is clear that the right-hand side of the plot is stable about the value ξ=1. This is to be expected since the Hill estimator is the ML estimator in the Pareto distribution. It becomes coherent that the Hill estimator is most effective when the underlying distribution is Pareto or similarly heavy-tailed. As such, there are instances when the Hill estimator is not as informative, for example, the Burr distribution. Resnick and Stărică (1997) suggested to plot $\{(\log k, H_{k,n}) : 1 \leq k \leq n-1\}$. This method, named alt (alternative) plotting, centers the graphics on the appropriate stable area. In the case of truncation, Beirlant et al. (2016) obtains an estimator $\hat{\xi}^{\mathrm{T}}_{k,n}$ by proposing Newton-Raphson iterations to solve the equation

$$H_{k,n} = \frac{1}{\hat{\xi}^{\mathrm{T}}_{k,n}} + \frac{\mathrm{R}_{k,n}^{\hat{\xi}^{\mathrm{T}}_{k,n}} \log \mathrm{R}_{k,n}}{1 - \mathrm{R}_{k,n}^{\hat{\xi}^{\mathrm{T}}_{k,n}}} \tag{24}$$

where $\mathrm{R}_{k,n} = X_{n-k,n}/X_{n,n}$.

3.3 Method of moments

The MOM estimator is a consistent estimator for ξ, i.e not restricted to ξ >0. The estimator was introduced in Dekkers et al. (1989) Dekkers et al (1989) and is defined by

$$\hat{\xi}^{\mathrm{MOM}}_{k,n} = \mathrm{M}^{(1)}_{k,n} + 1 - \frac{1}{2}\left(1 - \frac{(\mathrm{M}^{(1)}_{k,n})^2}{\mathrm{M}^{(2)}_{k,n}}\right)^{-1} \tag{25}$$

with

$$\mathrm{M}^{(l)}_{k,n} = \frac{1}{k}\sum_{i \leq k} \log^l \left(\frac{X_{n-i,n}}{X_{n-k,n}}\right), \qquad l = 1, 2.$$

Furthermore, the MOM estimator can be applied in estimating quantiles at the far right-end of a distribution. The MOM estimator for high quantiles is defined as

$$\hat{q}_p^{\mathrm{MOM}} = X_{n-k,n} + X_{n-k,n}\mathrm{M}^{(1)}_{n,k}\left(1 - 1 + \frac{1}{2}\left(1 - \frac{(\mathrm{M}^{(1)}_{n,k})^2}{\mathrm{M}^{(2)}_{n,k}}\right)^{-1}\right)\frac{\left(\frac{k}{np}\right)^{\hat{\xi}^{\mathrm{MOM}}_{n,k}} - 1}{\hat{\xi}^{\mathrm{MOM}}_{n,k}} \tag{26}$$

3.4 Bayesian estimation

In this section we consider the estimation of a parameter through Bayesian inference. For simple models with only few parameters, the posterior distribution can sometimes be derived directly. However, complicated models require a more intensive approach. A method called Markov Chain Monte Carlo (MCMC) was introduced as a class of algorits that allow us to approximate the posterior distribution of models with many parameters. The MCMC algorthim takes a prior distribution and likelihood function to generate random samples from a posterior distribution. The initial MCMC algorithm was the Metropolis algorithm (Hastings (1970)), and it is usually the algorithm of choice for many experts. Another algorithm is the Gibbs sampler, which has had several software applications compiled for its use. Software applications for the Gibbs sampler have built in distributions which allow statisticians to specify their desired prior distribution. This becomes an inconvenience when one wants to add a custom distribution of their own, such as the case for the EPD. For this and many other reasons a recently developed algorithm has emerged, called Stan (Team (2018)). In this research report we implement Stan to propose an estimator for a truncated EPD. For a more detailed discussion on how to implement a Bayesian model via Stan we refer the reader to Annis et al. (2017). The advantages of a Bayesian approach to parameter estimation is that additional knowledge can be taken into account and the methods do not rely on regulatory assumptions.

3.5 Estimation of second-order parameter ρ for heavy tails

In this section we provide the algorithm of Alves et al. (2007) used to estimate ρ.

1. Given a sample $(X_1, X_2, ..., X_n)$, plot the estimates of

$$\hat{\rho}_{k,r} := \min\{0, \frac{3(T_{k,n}^{(\tau)} - 1)}{T_{k,n}^{(\tau)} - 3}\}, \tag{27}$$

for $\tau = 0$, where

$$T_{k,n} := \frac{\left(M_{k,n}^{(1)}\right)^{\tau} - \left(\frac{M_{k,n}^{(2)}}{2}\right)^{\tau/2}}{\left(\frac{M_{k,n}^{(2)}}{2}\right)^{\tau/2} - \left(\frac{M_{k,n}^{(3)}}{6}\right)^{\tau/3}} \tag{28}$$

and as before

$$M_{k,n}^{(l)} := \frac{1}{k}\sum_{i\leq k}(\log X_{n-i+1:n} - \log X_{n-k:n})^{l}, \quad l \geq 1. \tag{29}$$

since $\tau = 0$, the notation $a^{b\tau} = b \ln a$ is used and therefore (27) is written as

$$T_{k,n} := \frac{\log(M_{k,n}^{(1)}) - \frac{1}{2}\log\left(\frac{M_{k,n}^{(2)}}{2}\right)}{\frac{1}{2}\log\left(\frac{M_{k,n}^{(2)}}{2}\right) - \frac{1}{3}\log\left(\frac{M_{k,n}^{(3)}}{6}\right)} \tag{30}$$

2. Consider $\{\hat{\rho}_{k,n}\}_{k\in\mathcal{K}}$ as given in (20) for large k, say $k \in \mathcal{K} = ([n^{0.990}], [n^{0.999}])$, compute their median, denoted ρ_τ and work with $\hat{\rho} \equiv \hat{\rho}_{\tau_0} := \hat{\rho}_{k_1}$, with

$$k_1 = [n^{0.995}] \tag{31}$$

The choice of the level k_1 is not very important, we can consider any reasonably large value of k of order $n^{1-\epsilon}$ from some intermediate sequence of integers $k = k(n)$ satisfying $\lim_{n\to\infty} \sqrt{k}A(n/k) = \infty$. This estimator uses more upper order statistics than the first-order parameter ξ.

4 The Extended Pareto Distribution

The POT approach consists of modelling the exceedances above some sufficiently high threshold, for which we have shown that the exceeses are defined by the GPD. In order to demonstrate the soundness of the model, we must examine whether or not the estimates of the GPD shape parameter ξ are steady when the model is applied to across an array of thresholds.

We are interested in how to proceed with the possibility of the threshold stability property not being visible. When fitting the GPD to the exceedances above the threshold, the lack of the stability property can be clarified by a slow rate of convergence in theorem 2.2. The same issue arises when fitting the Pareto distribution (PD) to relative excesses in the case of heavy-tailed distributions. Beirlant et al. (2009) proposed an extended version of the PD which aims to reduce asymptotic bias of the estimators and ultimately perform better for a larger range of thresholds. For heavy-tailed distributions, the derivation of the EPD was obtained by the reparametrizion of the Hall-class of distributions and more generally via regular variation of second-order conditions.

When fitting the EPD under the Fréchet domain of attraction (common heavy-tailed distributions), it is more appropriate to make use of the relative excesses X/t rather than absolute excesses $X - t$ since the limiting distribution of $P(X/t > x|X > t) \approx x^{-1/\xi}$ for t large and $x > 1$ is the PD.

The rest of this chapter is structured as follows. We provide the definition of the E(G)PD and show that the EPD produces a more accurate approximation to the limiting distribution of relative excesses for a broad case of heavy-tailed distributions. Estimation procedures for the parameters of the EPD are also discussed along with supplementary comments. Some final remarks are made comparing the GPD and the EPD.

4.1 The Extended Pareto distribution model

Definition 2 (Beirlant et al. (2009)). *The Extended Pareto Distribution is defined by its distribution function*

$$G_{\xi,\delta,\tau}(y) = \begin{cases} 1 - \{y(1+\delta-\delta y^{\tau})\}^{-\frac{1}{\xi}}, & \textit{if} \quad y > 1, \\ 0, & \textit{if} \quad y \leq 1, \end{cases} \tag{32}$$

where $\delta > max(-1, 1/\tau)$ *and* $\tau < 0 < \xi$.

The extended generalized Pareto distribution (EGPD) is then given as

$$H_{\xi,\delta\ ,\tau}\ (x) = G_{\xi,\delta\ ,\tau}\ (1+x), \qquad x \in \mathbb{R}. \tag{33}$$

Note that $H_{\xi,\delta\ ,\tau}\ (x)$ is the reparamaterization of the GPD with $\tau = -1$. The PD and GPD are both members of the EPD family. The E(G)PD is used to model the tails of heavy-tailed distributions which are satisfied by the following second-order condition.

Definition 3 (Beirlant et al. (2009)). *Let $\xi > 0$ and $\tau < 0$ be constants. A distribution function F is said to belong to the class $\mathcal{F}(\xi,\tau)$ if $x^{1/\xi}\overline{F}(x) \to C \in (0,\infty)$ as $x \to \infty$ and if the function δ defined via*

$$\overline{F}(x) = Cx^{-1/\xi}\{1+\xi^{-1}\delta(x)\} \tag{34}$$

is eventually non-zero and constant sign and such that $|\delta| \in \mathcal{R}_\tau$.

Definition 3 essentially describes the class of distributions in the Fréchet maximal domain of attraction with shape parameter $1/\xi$. The function $\delta(x) \sim Dx^\tau$ as $x \to \infty$ for some nonzero constant D, which comes from the Hall-class of distributions discussed in Subsection 2..7.2.3. In the following table we depict a few examples of the Hall class and include the second-order constant $p = \xi\tau < 0$; the parameter that commands the rate of convergence within the second-order framework.

Distribution	Distribution function	Parameters	ξ	τ	$\rho = \xi\tau$
Burr(ξ,ρ,β)	1-(1+$x^{-\rho/\xi}/\beta)^{1/\rho}$	$[\xi > 0, \rho < 0, \beta > 0]$	ξ	ρ/ξ	ρ
Fréchet(α)	$\exp(-x^{-\alpha})$	$[\alpha > 0]$	$1/\alpha$	$-\alpha$	-1
GPD(ξ,σ)	$1-(1+\xi x/\sigma)^{-1/\xi}$	$[\xi > 0, \sigma > 0]$	ξ	-1	-ξ
Student-t_u	$C(u)\int_{-\infty}^{x}(1+\frac{y^2}{u})^{-(u+2)/2}$	$[u > 0]$	$1/u$	-2	-2/u

Table 1: Selected heavy-tailed distributions with their respective extreme value indexes ξ and second-order constants τ and $\rho = \xi\tau$

Beirlant et al. (2009) also showed that for the distribution F belonging to the class $\mathcal{F}(\xi,\tau)$, the EPD improves the accuracy of the PD approximations to its respective excess distribution with some order of magnitude.

Proposition 1 (Beirlant et al. (2009)). *If $F \in \mathcal{F}(\xi,\tau)$, then as $t \to \infty$,*

$$\sup_{y\geq 1}\left|\frac{\overline{F}(ty)}{\overline{F}(t)} - \overline{G}_{\xi,\delta(t)\ ,\tau}\ (y)\right| = o\{|\delta(t)|\} \tag{35}$$

If in equation (35) the EPD tail function $\overline{G}_{\xi,\delta(t)\ ,\tau}\ (y)$ was replaced by the PD tail function $(y^{-1/\xi})$ for some $\sigma = \sigma(t)$ and $\tau \neq -1$, then the rate of convergence to the limit distribution would be $O\{|\delta(t)|\}$ only. In essence, when fitting the PD to its respective conditional relative excess distributions, the bias of the approximation is of the same order as $|\delta(t)|$. With the proposed EPD, we see that the bias of the approximation is ultimately smaller than $|\delta(t)|$. This is an important result since it implies that the EPD is able to model a larger portion of the data and that the EPD significantly improves the approximation $\overline{F}(ty)/\overline{F}(t) \approx y^{-1/\xi}$ for t large.

4.2 Parameter estimation

4.2.1 Maximum Likelihood

Our aim in this section is to obtain statistical inferences on the distribution function F that is to the right of some sufficiently high threshold. We let $X_1, X_2, ..., X_n$ be a random sample from F and denote the order statistics

$$X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$$

With regards to (35), estimation of the EPD parameters will be based on the relative excesses X_i/t. We choose a data-adaptive threshold $t = t_n = X_{n-k,n}$ where k is an intermediate sequence of integers and $i \in \{1, ..., n\}$ such that $X_i > t$.

The estimates of ξ and δ_n are obtained by maximizing an approximation to the EPD likelihood and solving the equations. The density function of the EPD is obtained by deriving (32) with respect to x. It is given by

$$\frac{d}{dx} = \frac{d}{dx}[G_{\xi,\delta,\tau}(x)] = \frac{1}{\xi}x^{-1/\xi-1}\{1+\delta(1-x^{\tau})\}^{-1/\xi-1}[1+\delta\{1-(1+\tau)x^{\tau}\}]. \tag{36}$$

The score functions given by Beirlant et al. (2009) admit the following expansions in $\delta \to 0$:

$$\frac{d}{d\xi}\log g_{\xi,\delta,\tau}(x) = \frac{-1}{\xi} + \frac{1}{\xi^2}\log x + \frac{\delta}{\xi^2}(1-x^{\tau}) + O(\delta^2) \tag{37}$$

$$\frac{d}{d\delta}\log g_{\xi,\delta,\tau}(x) = \frac{1}{\xi}\{(1-\xi\tau)x^{\tau}-1\} + \{1-2(1-\xi\tau)x^{\tau} + (1-2\xi\tau-\xi\tau^2)x^{2\tau}\}\frac{\delta}{\xi} + O(\delta^2) \tag{38}$$

For the moment we assume that τ is known. After substitution and simplification, Beirlant et al. (2009) gives the simplified estimators as

$$\hat{\delta}_{k,n} = \mathrm{H}_{k,n}(1-2\hat{p}_n)(1-\hat{p}_n)^3\frac{1}{\hat{p}_n^4}\left(\mathrm{E}_{k,n}(\hat{p}_n/\mathrm{H}_{k,n}) - \frac{1}{1-\hat{p}_n}\right) \tag{39}$$

$$\hat{\xi}_{k,n} = \mathrm{H}_{k,n} - \hat{\xi}_{k,n}\frac{\hat{p}_n}{1-\hat{p}_n} \tag{40}$$

where we define

$$H_{k,n} := \frac{1}{k}\sum_{j=1}^{k} \log \frac{X_{n-k+i:n}}{X_{n-k:n}},$$

$$\mathrm{E}_{k,n}(s) = \frac{1}{k}\sum_{j=1}^{k} \log\left(\frac{X_{n-k+i:n}}{X_{n-k:n}}\right)^{s}, \quad s \leq 0.$$

Note that $\hat{\delta}_{k,n}$ can be expected to be of order $O_p(k^{-1/2})$ as $n \to \infty$. In order to estimate $\tau = \hat{\tau}_{k,n} = \hat{p}_n/\mathrm{H}_{k,n}$, the unknown second-order parameter p is replaced by $\hat{p}_n$; a weakly consistent estimator sequence of $p = \xi\tau$. We usually let $p = -1$, however, the parameter can be estimated externally; see Section 3.5. This substitution will not affect the asymptotic distributions of the other estimators.
Theorem 3.1 of Beirlant et al. (2009) shows that the asymptotic bias of the EPD EVI estimator ($\hat{\xi}_{k,n}^{\mathrm{EPD}}$) has been reduced when compared to that of the GPD ($\hat{\xi}_{k,n}^{\mathrm{GPD}}$) and Hill 1975 estimators. The following limit distributions are expressed explicity under the conditions of Theorem 3.1:

$$\sqrt{k}(\hat{\xi}_{k,n}^{\mathrm{EPD}} - \xi) \sim N(0, \xi^2\frac{(1-p)^2}{p^2}), \qquad n \to \infty \tag{41}$$

$$\sqrt{k}(\mathrm{H}_{k,n} - \xi) \sim N(\lambda\frac{p}{1-p}, \xi^2), \qquad n \to \infty \tag{42}$$

$$\sqrt{k}(\hat{\xi}_{k,n}^{\mathrm{GPD}} - \xi) \sim N(\lambda b(\xi,p), (1+\xi)^2), \qquad n \to \infty \tag{43}$$

where $b(\xi, p) = \frac{p(1+\xi)(\xi+p)}{\xi(1-p)(1+\xi-p)}$. As can be seen from the equations above, the Hill 1975 estimator has the smallest asymptotic variance

4.2.2 Bayesian inference

Maribe (2016) proposed a Bayesian EPD estimator using a non-informative prior. One of the main aims of the paper was to assess the Bayesian estimators performance in comparison to the ML approach. An MDI prior is assigned to the parameter ξ and a vague truncated Normal prior is assigned to δ. Beirlant et al. (2006) showed that for a Pareto distribution, the MDI prior is given by:

$$\pi(\xi) \propto \frac{e^{-\xi}}{\xi} \tag{44}$$

It seems natural to select the Pareto MDI prior for the EPD parameter ξ since the EPD is a second-order generalization of the simple Pareto distribution. Thereafter MCMC methods are performed to obtained random samples from a posterior distribution.

4.3 Remarks

In this Chapter, we discussed the EPD as a limiting distribution along with its properties. Through asymptotic theory and simulations, Beirlant et al. (2009) concluded that the EPD improves approximations to the excess distribution with some order of magnitude. The Fréchet, student-t, Pareto mixture and log gamma distributions were simulated to observe the different estimators. For each distribution, the EPD ML estimator performs relatively well when compared to the Hill and GPD ML estimators in terms of bias, asymptotic variance and mean squared error. Overall the EPD is a better POT model to consider than the GPD. Furthermore, Maribe (2016) showed that a Bayesian approach to estimating EVI for heavy-tailed distributions reduces the bias and improves stability when compared to the popular ML estimation method.

5 A simulation study

In this section we assess the performance of our proposed Bayesian EPD estimator by performing a simulation study on truncated and non-truncated distributions, taking 120 repetitions each of sample size $n = 200$. The distributions under study are the heavy-tailed distributions: Pareto, Burr and Fréchet.

Non-truncated distributions

1) Pareto(α), $\alpha = 2, \xi = 0.5$

$$F_W(x) = 1 - x^{-\alpha} \qquad x > 1, \alpha > 0,$$

2) Burr(α, ρ), $\alpha = 2, \rho = -2$

$$F_W(x) = 1 - (1 + x^{-\rho\alpha})^{1/\rho}, \qquad x > 0, \rho < 0, \alpha > 0,$$

3) Fréchet(α), $\alpha = 4$

$$F_W(x) = e^{-x^{-\alpha}}, \qquad x > 0, \alpha > 0$$

.

Truncated distributions

1) Pareto(α, T), $\alpha = 2$

$$F_T(x) = \frac{1 - x^{-\alpha}}{1 - T^{-\alpha}} \qquad x > 1, \alpha > 0,$$

2) Burr(α, ρ, T), $\alpha = 2, \rho = -2$

$$F_T(x) = \frac{1 - (1 + x^{-\rho\alpha})^{1/\rho}}{1 - (1 + T^{-\rho\alpha})^{1/\rho}}, \qquad x > 0, \alpha > 0$$

3)Fréchet(α, T), $\alpha = 4$

$$F_T(x) = \frac{e^{-x^{-\alpha}}}{e^{-T^{-\alpha}}}, \qquad x > 0, \alpha > 0$$

The random variable W describes a distribution before truncation. We refer the reader back to Section 2.3. In practice, one does not know if data comes from a truncated or non-truncated distribution, and as such, we need to analyse the performance of our estimator in both cases. We consider no truncation $(Q_W(1))$, light truncation $(Q_W(0.99))$ and rough truncation $(Q_W(0.90))$.

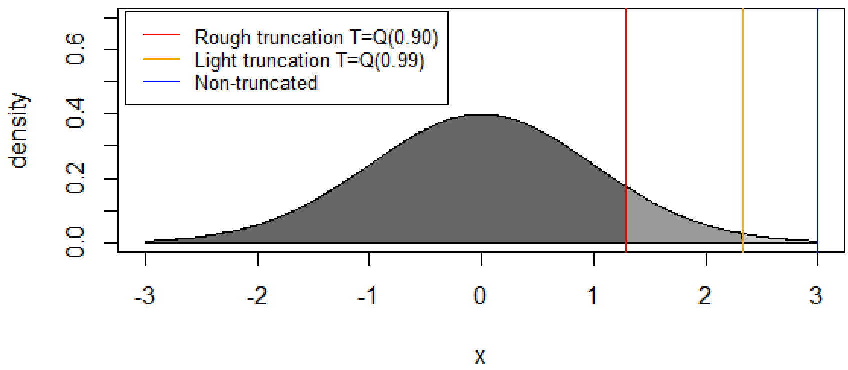


Figure 7: Schematic representation of a standard normal distribution truncated at different levels T.

Figure 8 illustrates the approach of our simulation. We fit the EPD to a truncated heavy-tailed distribution with the aim of estimating the EVI using Bayesian techniques. The skyblue shaded region represents the truncated EPD at an arbitrary truncation level T.

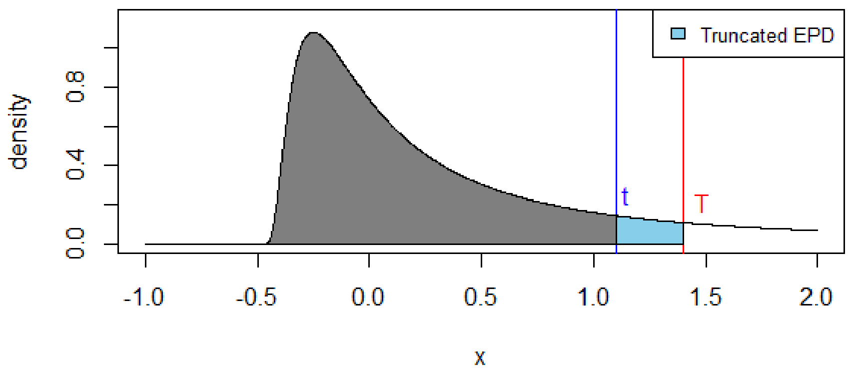


Figure 8: Schematic representation of a truncated EPD distribution from the right-hand side with threshold value t and truncation value T.

The behaviour of our proposed EPD Bayes estimator $\hat{\xi}^{T}_{\mathrm{EPD}}$ has been obtained through the use of MCMC methods, which are executed via Stan in the R studio statistical software. Inferences on the EVI ξ are then achieved by taking the mode of the posterior samples. Other estimators such as the truncated Hill 1975 are obtained using the ReIns package located in R, whilst the MOM and ρ estimators are coded explicitly.

```
data
{
  int  <lower=1> N;     // k in the global function
  real <lower=0> y[N]; // relative excesses
  real t;
  real <upper=0> rho;   //second order p[ar]
  real <lower=0> Xnn;   //Endpoint in this case I took the max
  real U;               //I set an upper limit and lower limit for \gamma
  real L;
  real sigma;
  real Truncated;
}
parameters
{
  real <lower=L,upper=U>gama;
  real <lower=fmax(-1,(gama/rho))> delta;
}
transformed parameters
{
  real<upper=0>taou;
  real T_gam;
  T_gam=pow((Xnn/t),(-1/gama));
  taou=rho/gama;
}
```

Figure 9: Stan code of data, parameters and transformed parameters code blocks

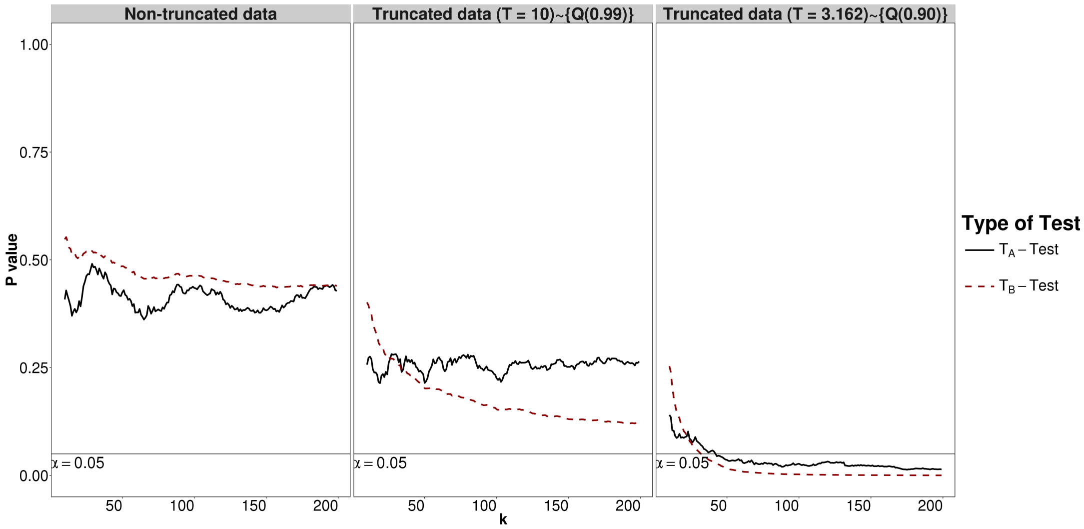
The first code block of Stan is the data block, which contains variables that are to be used in the model. The second code block is the parameters block. The parameters declared in our code block are the parameters from the EPD function in Definition (2) that Stan will estimate. The third code block contains transformed parameters because it is often suitable to work with the transformation of certain parameters.

```
model
{
    delta ~ normal(0,sigma);
    target += -log(gama)-gama;

    for (i in 1:N){
        y[i] ~ pdfEPD(gama,delta,taou);
        if (Truncated==1){
          target += -log(1-pow(((Xnn/t)*(1+delta-delta*pow((Xnn/t),taou))),(-1/gama)));
        }
    }
}
```

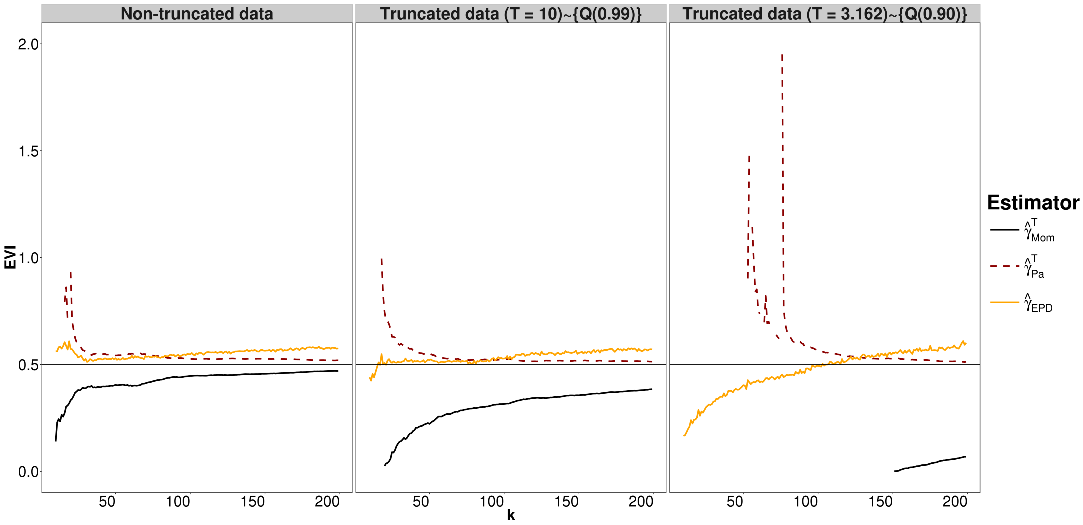
Figure 10: Stan code of model code block.

In Figure 10 we have the statistical model code block which contains the prior distribution and the likelihood. We let δ of Definition (2) take on a normal distribution with mean 0 and variance σ^2. The EVI parameter of the truncated EPD takes on the MDI prior (44) (derived for Pareto-type tails). The *for loop* in the code is where we assign the likelihood for the model. In the case of no truncation, the likelihood is the simple EPD function, and in the case of truncated data, we use the truncated EPD. The syntax target $+=$ is programming notation used to increment the log probability density function.



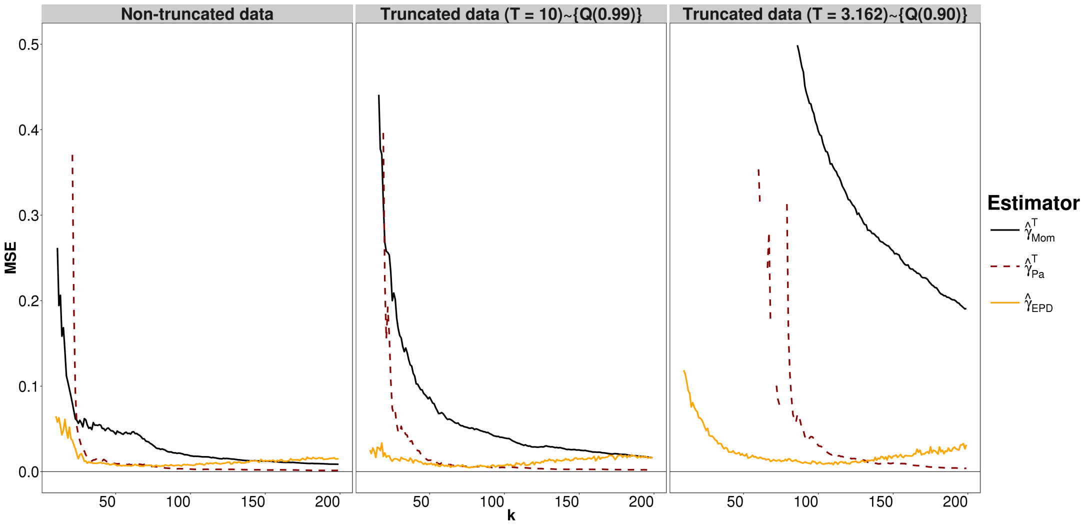
(a) Truncation test at a 5% level of significance using mean p-values.

As discussed earlier, in practical applications it can be difficult to detect whether a dataset is subjected to truncation effects. For this reason, we have to test whether a dataset is truncated to some level. There exists two different types of truncation tests. The first truncation test in Figure 11(a) (black line) is used to test the null hypothesis $H_0 : T = \infty$, and the second test (red line), is the test light versus rough truncation, where $H_0 : T =$ light truncation. For the tests, we have selected a 5% level of significance. If on average the plot is below the straight line drawn at 0.05, we reject the null hypothesis. Under non-truncation (left column) and light truncation (middle column), it is clear that we do not reject the null hypothesis. In the case of rough truncation (right column), we notice that the plots are on average below the specified p-value and thus we are to reject the null hypothesis for both tests.



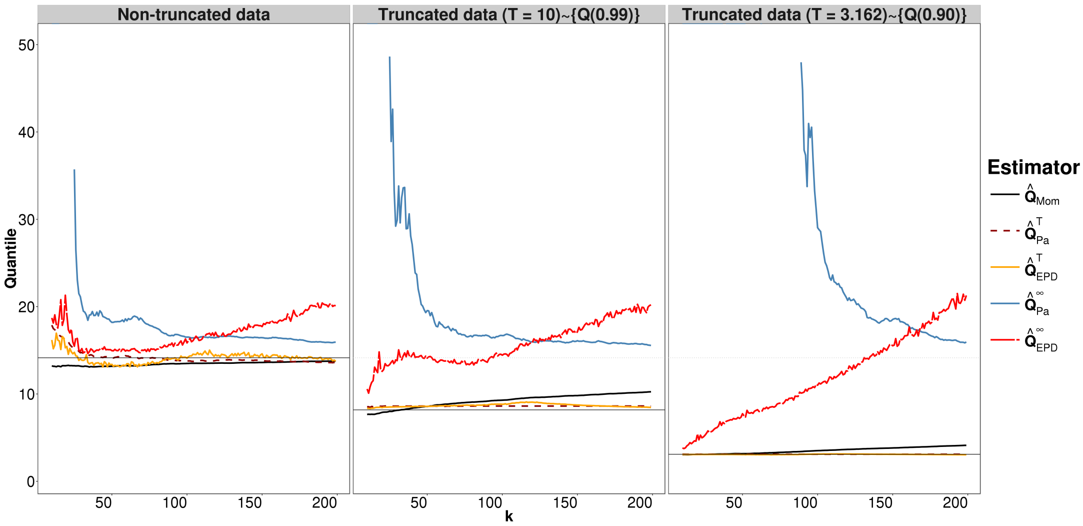
(b) EVI estimates of Pareto $\gamma = 0.5$ distributed data

The primary aim of this simulation study is to observe the performance of our estimator for the EVI parameter. To draw comparisons, we assess the performance of three different estimators: MOM estimator (black line), truncated Hill/Pareto estimator (red dotted line) and our own proposed truncated Bayesian EPD estimator (orange line). The MOM estimator was discussed in Section 3.3 and the truncated Hill/Pareto estimator is detailed at the end of Section 3.2. A well-performing estimator should vary close to the horizontal line, which is drawn at the true value of the EVI. In Figure 11(b) we have that the EVI is equal to 0.5. Under no truncation (left column) and light truncation (middle column), we notice that the truncated Hill/Pareto estimator and our Bayes estimator do relatively well in both cases, whilst the MOM estimator underestimates in the case of light truncation (middle column). However, under light truncation (middle column), our estimator performs significantly better for small values of k (exceedances above the threshold). Observing the plot under rough truncation (right column) for small values of k , we notice that the truncated Hill/Pareto overestimates whilst our estimator underestimates. In the same plot the MOM estimator is almost non-existent. The truncated Hill/Pareto and Bayesian EPD estimates tend to perform better as k increases.



(c) MSE estimates of Pareto γ= 0.5 distributed data

In Figure 11(c) we plot the Mean Squared Error (MSE) for the three different estimators. The MSE is written as $(\hat{\xi} - \xi)^2$, and is a useful statistical tool that helps us compare the estimators. Our Bayesian estimator has the samllest MSE compared to the other estimators under a Pareto distribution that is subjected to rough truncation (right column).

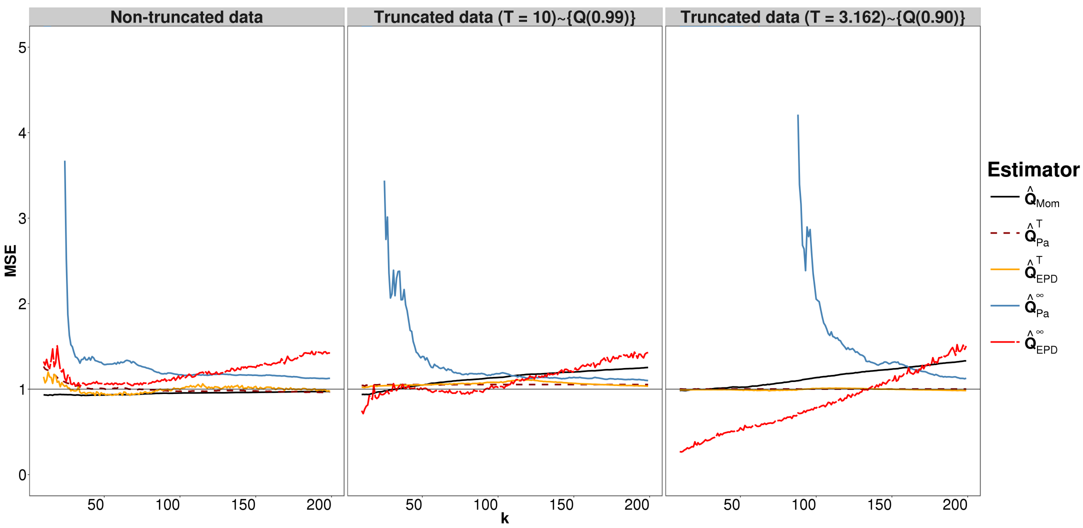


(d) Quantile estimators for Pareto $\gamma = 0.5$ distributed data.

In Section 2.2 we discussed quantiles and tail estimation. The estimation of extreme quantiles provides additional statistical inferences about the upper-end of the underlying distribution and the truncated distribution. Here we compare the performance of five extreme quantile estimators. We consider the MOM (detailed in Section 3.3),$\hat{Q}^{\infty}_{\mathrm{Pa}}$, $\hat{Q}^{T}_{\mathrm{Pa}}$, $\hat{Q}^{\infty}_{\mathrm{EPD}}$ and $\hat{Q}^{T}_{\mathrm{EPD}}$ quantile estimators. The ∞ notation implies that the estimator is the true estimator under no truncation.

The non-truncated $\hat{Q}^{\infty}_{\mathrm{Pa}}$ estimator is essentially an adaption to the Weissman estimator 1978 and is given as Equation (24) in Beirlant et al. (2016). In the same paper, the truncated Pareto-type quantile estimator is defined as Equation (23). Both estimators are located in the ReIns package of R Studio. Furthermore, the $\hat{Q}^{\infty}_{\mathrm{EPD}}$ estimator is derived in Goegebeur et al. (2014). In order to obtain the truncated version, we replace the Weissman estimator with equation with Equation (23) of Beirlant et al. (2016). These estimators are coded explicitly in the R code of the appendix.

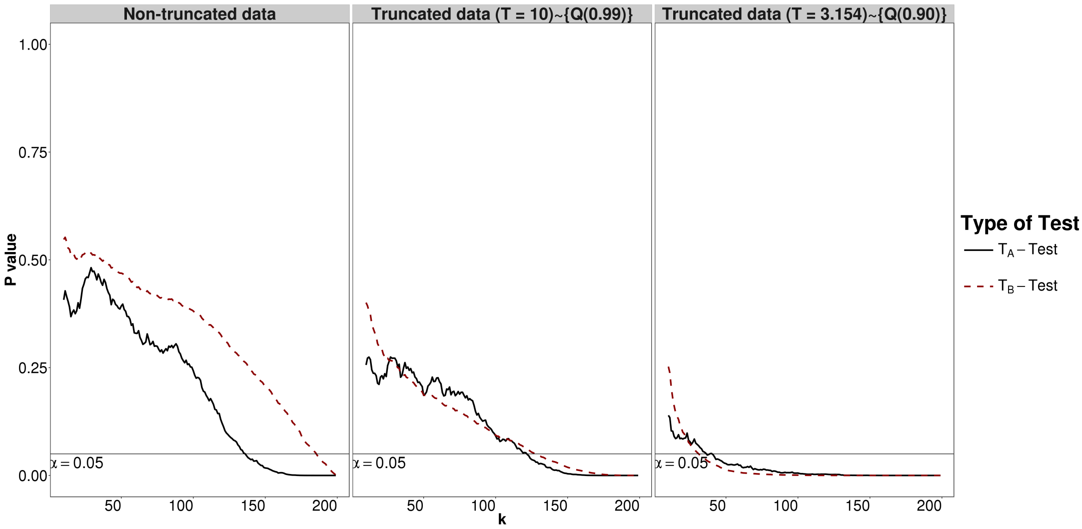
The left column is the underlying distribution W with a horizontal line drawn at the extreme quantile $Q(1-1/n)$ (tail quantile function) where $n = 200$. The same process is applied under light and rough truncation. For example, in the middle column, the horizontal line for the extreme quantile is drawn just below the truncation value of 10. Apart from the $\hat{Q}^{\infty}_{\mathrm{Pa}}$ and $\hat{Q}^{\infty}_{\mathrm{EPD}}$ estimators, the quantile estimators under non-truncated data (left column) seem to perform relatively well. Again, under light truncation (middle column), the truncated quantile estimators perform well, apart from the MOM estimator, which tends to overestimate as k increases. Noticeably, the $\hat{Q}^{\infty}_{\mathrm{Pa}}$ and $\hat{Q}^{\infty}_{\mathrm{EPD}}$ estimators under light truncation (middle column) vary close to the horizontal line drawn at the extreme quantile of the non-truncated distribution (left column), with our estimator $\hat{Q}^{\infty}_{\mathrm{EPD}}$ being slightly more stable than $\hat{Q}^{\infty}_{\mathrm{Pa}}$. The behaviour of the non-truncated (left column) extreme quantiles estimators is interesting considering that they have no information about the previous extreme quantile. In the case of rough truncation (right column) the three estimators $\hat{Q}^{T}_{\mathrm{EPD}}$, $\hat{Q}^{T}_{\mathrm{Pa}}$ and $\hat{Q}_{\mathrm{MOM}}$ achieve plots near the extreme quantile whilst $\hat{Q}^{\infty}_{\mathrm{Pa}}$ overestimates and $\hat{Q}^{\infty}_{\mathrm{EPD}}$ underestimates. The poor performance of the latter two may be due to the higher truncation level, which results in less information being provided to the estimators.



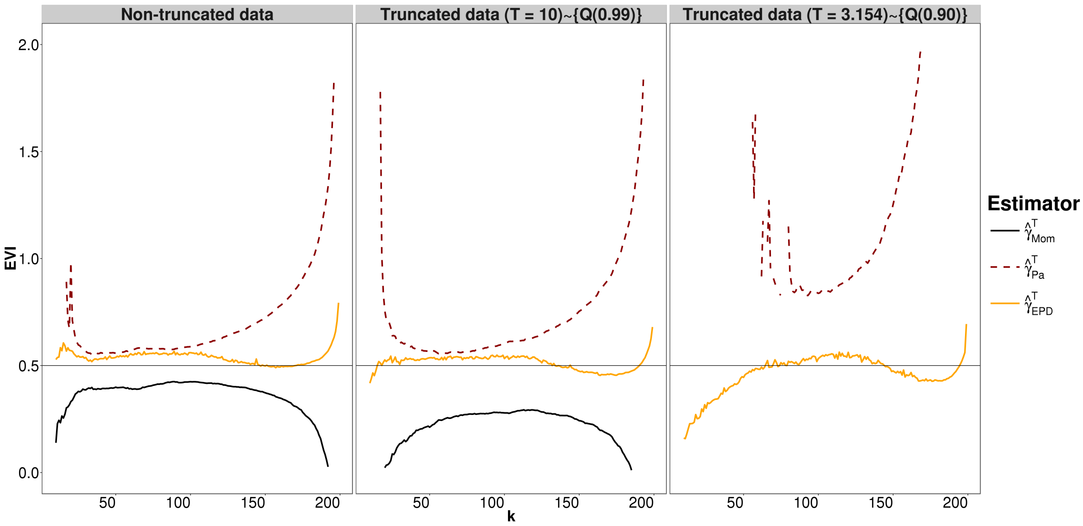
(e) MSE of Quantile estimators for Pareto $\gamma = 0.5$ distributed data.

Figure 11: Pareto(α=2).Left column: no truncation; Middle column: light truncation; Right column: rough truncation.

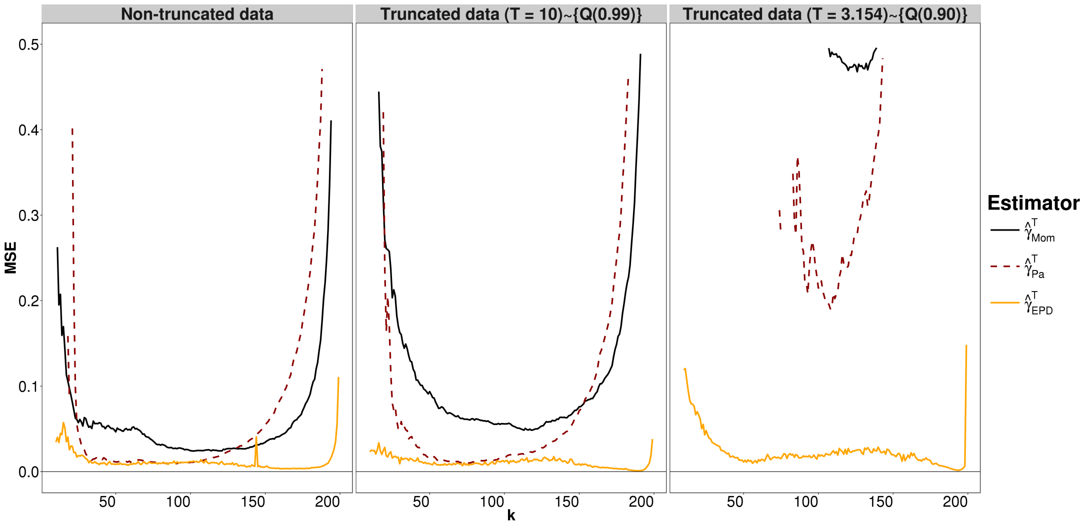
Similar to the MSE of the EVI parameter, the MSE of the quantile estimators is denoted as $(\hat{Q} - Q)$. Under no truncation (left column) and light truncation (middle column) all the estimators vary close to the desired value. In the case of rough truncation (right column) the $\hat{Q}^{T}_{\mathrm{EPD}}$ and $\hat{Q}^{T}_{\mathrm{Pa}}$ estimators perform the best in terms of truncation, whilst the non-truncated quantile estimators support the results from the previous figure.



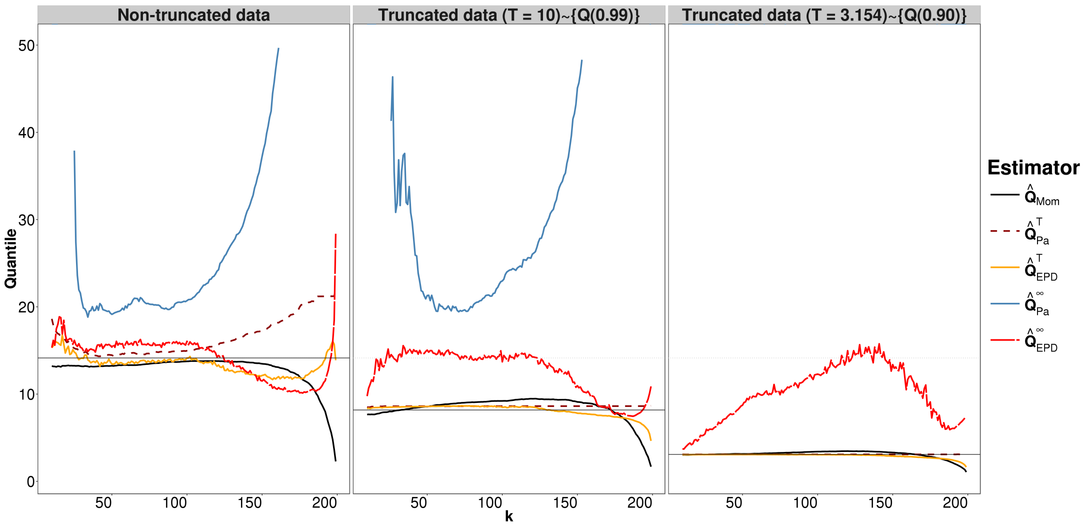
(a) Truncation test at a 5% level of significance using mean p-values.



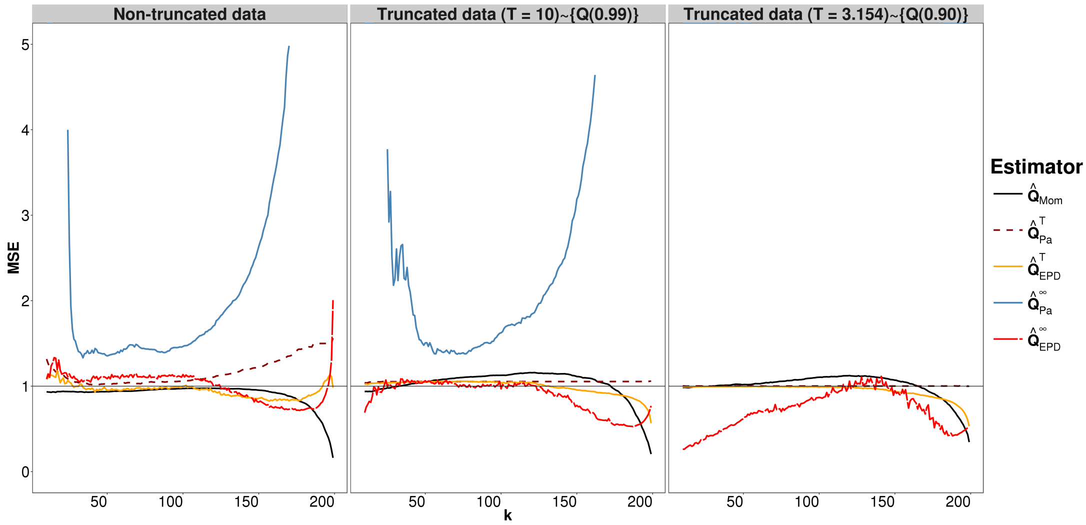
(b) EVI estimates of Burr $\gamma = 0.5$ distributed data



(c) MSE of estimators for Burr γ=0.5 distributed data.



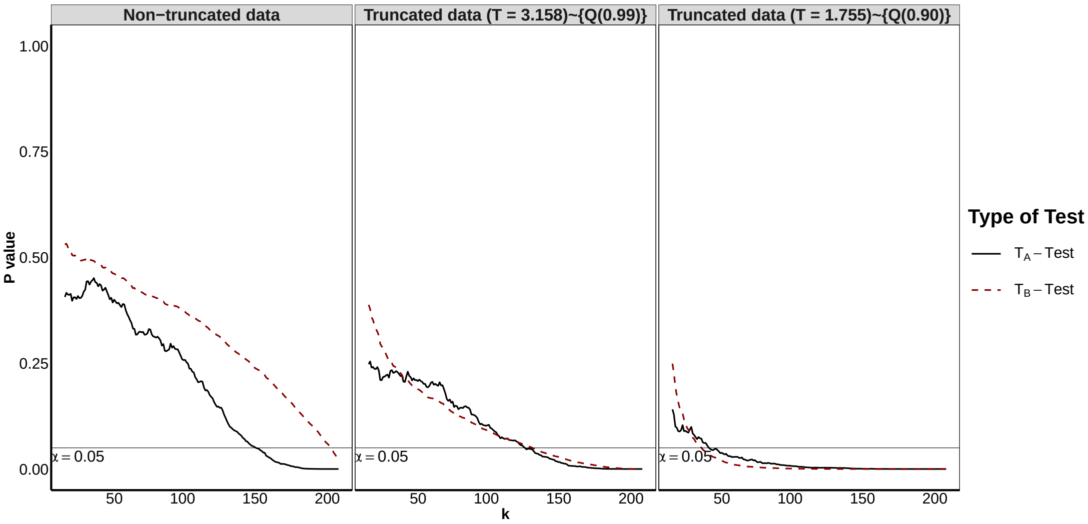
(d) Quantile estimators for Burr γ = 0.5 distributed data.



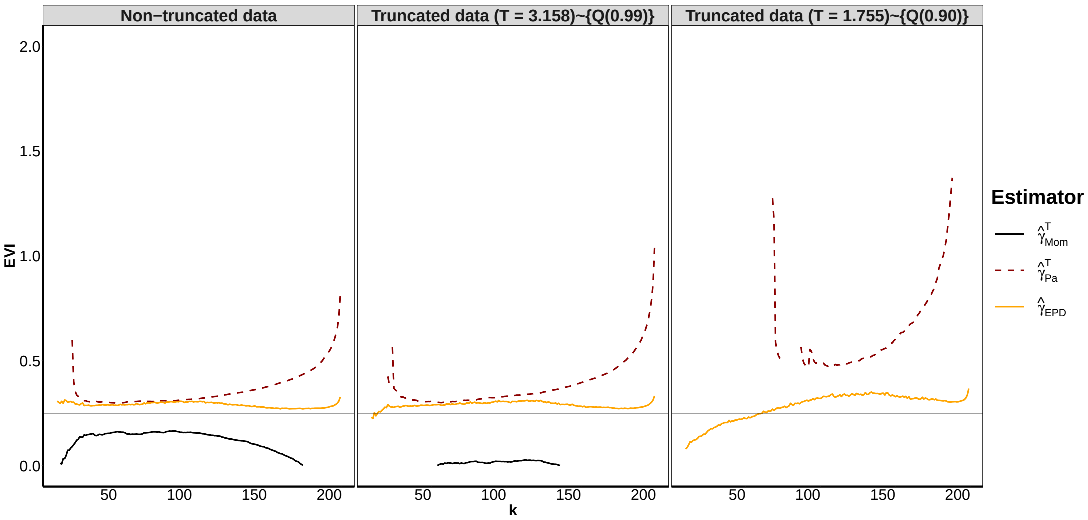
(e) MSE of Quantile estimators for Burr $\gamma = 0.5$ distributed data.

Figure 12: Burr(α=2,ρ=-2).Left column: no truncation; Middle column: light truncation; Right column: rough truncation.

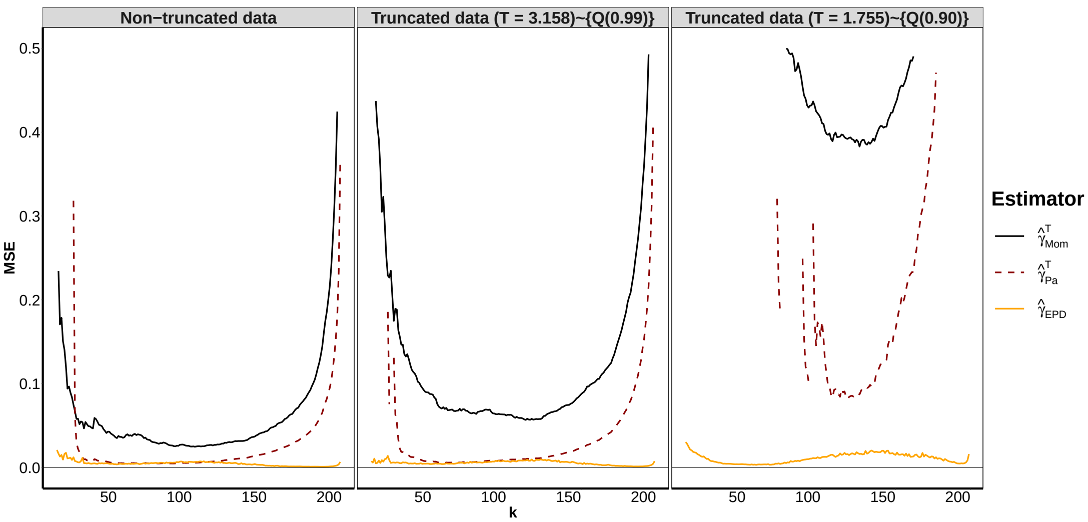
In the case of the Burr distribution, we reject the null hypothesis under rough truncation (right column) for both tests. In the case of light truncation (middle column), the tests seem to have the first half of the mean values above the p-value, which leaves rejection of the tests to interpretation and expert opinion. However, as k increases we approach a leniency towards rejecting both tests. The behaviour of the EVI estimators of Figure 12(b) indicate that our estimator (orange line) outperforms the other estimators for all levels of truncations. This is especially visible as k increases. Furthermore, the results of the MSE graphs show that our estimator has the smallest MSE. The extreme quantile estimators for the Burr distribution again show that our estimator $\hat{Q}^{\infty}_{\mathrm{EPD}}$ performs well under light truncation (middle column). With regards to the truncated extreme quantile estimator, there is not a significant difference in their performances for all levels of truncation.



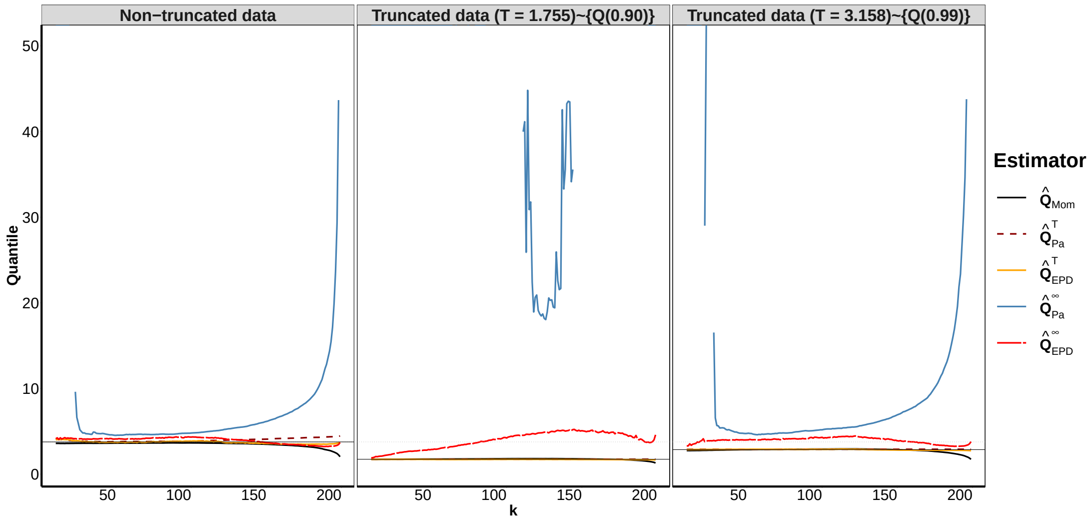
(a) Truncation test at a 5% level of significance using mean p-values.



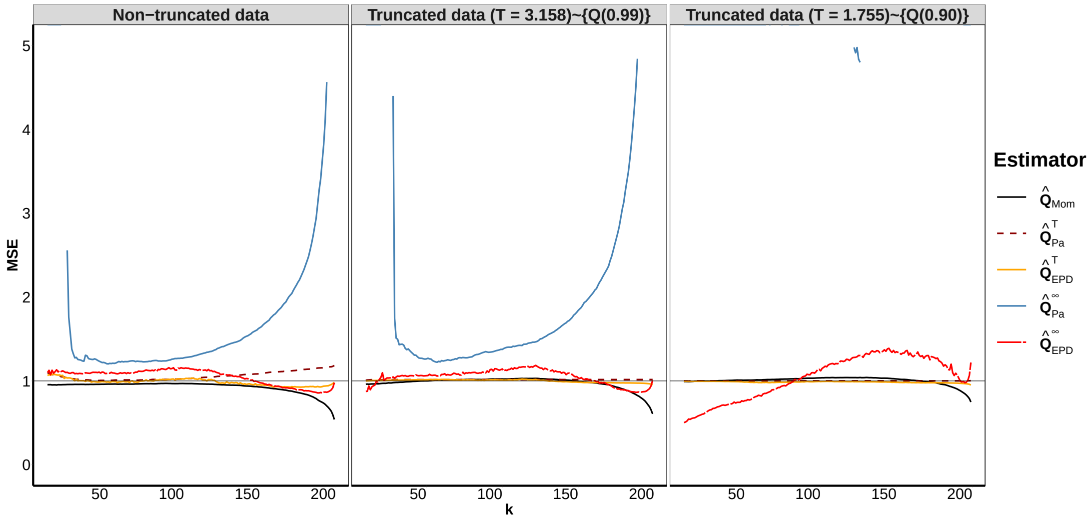
(b) EVI estimates of Fréchet γ= 0.5 distributed data



(c) MSE of estimators for Fréchet γ=0.25 distributed data.



(d) Quantile estimators for Fréchet γ =0.25 distributed data.



(e) MSE of Quantile estimators for Fréchet γ = 0.25 distributed data.

Figure 13: Fréchet(α=2).Left column: no truncation; Middle column: light truncation; Right column: rough truncation.

The results of Figure 13 portray a similar behaviour to that of the Burr distribution, with our estimator clearly outperforming the other estimators in terms of EVI and extreme quantile estimation. Noticably our estimator has the smallest MSE for the EVI estimations under rough truncation (right column).

6 Conclusions and recommendations

The main aim of this research report was to investigate the truncated EPD model and analyse various methods of parameter estimation. We introduced a truncated Bayesian EPD parameter estimator, which was obtained through an exhaustive MCMC algorithm. From the simulation study, we were able to observe the behaviour of our estimator in comparison to the popular ML estimation method and the

MOM estimator. The simulation study showed that our Bayesian approach to estimating the EVI ξ of truncated and non-truncated distributions reduces the bias and mean squared error when compared to the ML estimators. This may be linked to the EPD being a better fit than the PD in terms of bias and variance. Our estimator is especially better performing under rough truncation.

While the results are useful and informative, there still exists further research possibilities. The Bayesian approach seems to underestimate the EVI for small values of k (number of exceedances above the threshold) in the case of rough truncation. One could further investigate on how to improve the estimator when there is insufficient data above a large threshold. In this research report, we assumed that the MDI prior of the EVI ξ was that of a Pareto-Type tail. Since we are applying a Bayesian framework, an informative prior based on expert opinion and insight might improve the estimator. Moreover, one could derive a more accurate MDI prior. There is also the possibility of deriving the ML for a truncated EPD along with necessary asymptotic theory.

An interesting result that has emerged in the simulation study is the behaviour of our non-truncated Bayes EPD quantile estimator under light truncation. For a truncated distribution, to some extent, the estimates vary close to the previous extreme quantile without any knowledge about its predecessor. In essence, this captivating result implies that fitting an EPD model to truncated heavy-tailed distribution may provide information about the extreme quantiles of the underlying distribution i.e if the dataset had not been subjected to truncation effects.

Appendix

R code for simulation study

```r
# Estimation of extreme value index of truncated Pareto model
rm(list=ls())
load("CompiledEPDModel.RData",envir=.GlobalEnv)

Packages<-c("ReIns","ggplot2","parallel","R.utils","grid","coda","gridExtra")
for(i in 1:length(Packages)){
if(all(installed.packages()[,1]!=Packages[i])){
install.packages(Packages[i])}
library(Packages[i],character.only=TRUE)
}
library(R.utils)

BTest <- function(data, r = 1) {
X <- as.numeric(sort(data))
n <- length(X)
gamma <- numeric(n)
H1 <- numeric(n)
R <- numeric(n);  K <- r:(n-1)
p<-numeric(n)

K1 <- 1:(n-1)
H1[K] <- (cumsum(log(X[n-K1+1]))/K1 - log(X[n-K1]))[K]
R[K] <- X[n-K]/X[n-r+1]
p[K] <- exp(-K*R[K]^(1/H1[K]))
return(p)
}

#Truncated EPD
library(rstan)
library(coda)
library(parallel)
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())
#STan Model
```

```
#--------------------------------------------------
#estimator of p: the second order convergence parameter
rho.fraga.alves <-function(x){
n<-length(x); k1=min((n-1),round(n^0.995))
xk <- x[1:k1]/x[k1+1]
T <- (log(mean(log(xk)))-0.5*log(mean(log(xk)^2)/2))/
(0.5*log(mean(log(xk)^2)/2)-(1/3)*log(mean(log(xk)^3)/6))
r.tau.tilde=-abs((3*(T-1))/(T-3))
return(r.tau.tilde) }

Moment <- function(data, p=pp) {

X <- sort(data)
n <- length(X)
M1 <- numeric(n)
M2 <- numeric(n)
Mom <- Q <- numeric(n)
K <- 1:(n-1)

for (k in (n-1):1) {
Z.values <- log(X[n:(n-k+1)]/X[n-k])
M1[k] <- (1/k)*sum(Z.values)
M2[k] <- (1/k)*sum(Z.values^2)
}
#method of moments estimator
Mom <- M1 + 1 - (1-(M1)^2/M2)^(-1)/2

QMom<-X[n-K]+X[n-K]*M1[K]*((1-(M1[K])^2/M2[K])^(-1)/2)*(((K/(n*p))^Mom[K]-1)/Mom[K])
TM<-X[n-K]-(X[n-K]*M1[K]*((1-(M1[K])^2/M2[K])^(-1)/2))/Mom[K]; TMom=pmax(TM,X[n])
list(k=K, gamma=Mom[K],Q=QMom[K],TMom=TMom[K])
}

QBayes<-function(data,n,k, gamma, delta,rho,pp) {
A <- -1/gamma
R <- data[n]/data[n-k]
G <- (R*(1+delta*(1-R^(-rho*A))))^A
```

```
DT <-  pmax((k+1)/(n+1) * (G-1/k) / (1-G), 0.001)
QT<-data[n-k]*((DT+(k)/(n))/(DT+pp))^(gamma)*exp(-delta*(n/k)^rho)
QW<-data[n-k]*((k/n)/pp)^(gamma)*exp(-delta*(((k/n)/pp))^rho)
return(list(DT=DT,QT=QT,QW=QW))
}

#STan Model
#-----------------------------------------------------
GIBBS <- function(data,k,rho,t,Xnn,sigma.k,Truncated,L,U){
dat <- list(N = k, y = data,rho=rho,t=t,Xnn=Xnn,sigma=sigma.k,
Truncated=Truncated,L=L,U=U)
EPDfit<-tryCatch({withTimeout({
stan(fit=fit, data = dat,iter = 1000, chains = 1,verbose=F);
}, timeout=30, onTimeout="silent")},error=function(e){return(NULL)});
if(class(EPDfit)!="NULL"){
stan2coda<-mcmc.list(lapply(1:ncol(EPDfit), function(x)
 mcmc(as.array(EPDfit)[,x,])))
samps<-as.data.frame(as.matrix(stan2coda))
gamas <- samps$gama  ;deltas<-samps$delta;
return(list(gamas=gamas,deltas=deltas))
}
}
SimulateData<-function(distribution,n,alf,quan,Trunc){
if(distribution =="Pareto"){
Tn<-ifelse(Trunc==1,qpareto(quan, shape=alf),1e99)
simDat<-rtpareto(n, shape=alf, endpoint = Tn)
sampl<-sort(simDat)
Q<-qtpareto(1-1/n, shape=alf, endpoint = Tn)
QW<-qtpareto(1-1/n, shape=alf, endpoint = Inf)
}else if(distribution =="Frechet"){
Tn<-ifelse(Trunc==1,qfrechet(quan, shape=alf),1e99)
simDat<-rtfrechet(n, shape=alf, endpoint = Tn)
sampl<-sort(simDat)
Q<-qtfrechet(1-1/n, shape=alf, endpoint = Tn)
QW<-qtfrechet(1-1/n, shape=alf, endpoint = Inf)
}else if(distribution =="Burr"){
Tn<-ifelse(Trunc==1,qburr(quan,alpha=alf, rho=-2),Inf)
```

```
simDat<-rtburr(n, alpha=alf, rho=-2, endpoint=Tn)
sampl<-sort(simDat)
Q<-qtburr(1-1/n,alpha=alf, rho=-2, endpoint = Tn)
QW<-qtburr(1-1/n,alpha=alf, rho=-2, endpoint = Inf)
}else if(distribution =="GPD"){
Tn<-ifelse(Trunc==1,qgpd(quan,gamma=1/alf, sigma = 1),Inf)
simDat<-rtgpd(n, gamma=1/alf, sigma = 1, endpoint = Tn)
sampl<-sort(simDat)
Q<-qtgpd(1-1/n,gamma=1/alf, sigma = 1, endpoint = Tn)
QW<-qtgpd(1-1/n,gamma=1/alf, sigma = 1, endpoint = Inf)
}else if(distribution =="LogNormal"){

}else{
stop("distribution not in list")
}
return(list(sampl=sampl,Tn=Tn,Q=Q,QW=QW))
}
#--------------------------------------------------
SimFun<-function(sim){
set.seed(123+sim)
data1<-SimulateData(dist,n,alf,quan=TruncLevel[1],Trunc=0)$sampl
data2<-SimulateData(dist,n,alf,quan=TruncLevel[2],Trunc=1)$sampl
data3<-SimulateData(dist,n,alf,quan=TruncLevel[3],Trunc=1)$sampl

rho3 <- -1#rho.fraga.alves(rev(data3))
rho2 <- -1#rho.fraga.alves(rev(data2))
rho1 <- -1#rho.fraga.alves(rev(data1))

MOM1<-MOM2<-MOM3<-TPA1<-TPA2<-TPA3<-TEPD1<-TEPD2<-TEPD3<-vector()
BDT1<-BDT2<-BDT3<-DT1<-DT2<-DT3<-Delta1<-Delta2<-Delta3<-vector()
QMOM1<-QMOM2<-QMOM3<-QTPA1<-QTPA2<-QTPA3<-QTEPD1<-QTEPD2<-QTEPD3<-vector()
QWTPA1<-QWTPA2<-QWTPA3<-QWTEPD1<-QWTEPD2<-QWTEPD3<-vector()
TA1<-TA2<-TA3<-TB1<-TB2<-TB3<-vector()

TestB1<-trTest(data1, plot = F)$Pval[bb:(n-1)]
TestB2<-trTest(data2, plot = F)$Pval[bb:(n-1)]
TestB3<-trTest(data3, plot = F)$Pval[bb:(n-1)]
```

```
TestA1<-BTest(data1)[bb:(n-1)]
TestA2<-BTest(data2)[bb:(n-1)]
TestA3<-BTest(data3)[bb:(n-1)]
#Non-adapted Estimates************************************
M1<-Moment(data1)
M2<-Moment(data2)
M3<-Moment(data3)

MOM1<-M1$gamma[bb:(n-1)];QMOM1<-M1$Q[bb:(n-1)]
MOM2<-M2$gamma[bb:(n-1)];QMOM2<-M2$Q[bb:(n-1)]
MOM3<-M3$gamma[bb:(n-1)];QMOM3<-M3$Q[bb:(n-1)]

TrHill1 <-trHill(data1)
TrHill2<-trHill(data2)
TrHill3<-trHill(data3)

TPA1<-TrHill1$gamma[bb:(n-1)]
TPA2<-TrHill2$gamma[bb:(n-1)]
TPA3<-TrHill3$gamma[bb:(n-1)]

DT1<-trDT(data1,r=1,gamma=TrHill1$gamma)$DT
DT2<-trDT(data2,r=1,gamma=TrHill2$gamma)$DT
DT3<-trDT(data3,r=1,gamma=TrHill3$gamma)$DT

QTPA1<-trQuant(data1, r = 1, rough = TRUE, gamma=TrHill1$gamma, DT=DT1, p=pp)
$Q[bb:(n-1)]
QTPA2<-trQuant(data2, r = 1, rough = TRUE, gamma=TrHill2$gamma, DT=DT2, p=pp)
$Q[bb:(n-1)]
QTPA3<-trQuant(data3, r = 1, rough = TRUE, gamma=TrHill3$gamma, DT=DT3, p=pp)
$Q[bb:(n-1)]

QWTPA1<-trQuantW(data1, r = 1, rough = FALSE, gamma=TrHill1$gamma, DT=DT1, p=pp)
$Q[bb:(n-1)]
QWTPA2<-trQuantW(data2, r = 1, rough = FALSE, gamma=TrHill2$gamma, DT=DT2, p=pp)
$Q[bb:(n-1)]
QWTPA3<-trQuantW(data3, r = 1, rough = FALSE, gamma=TrHill3$gamma, DT=DT3, p=pp)
```

```r
$Q[bb:(n-1)]
#Bayesian Estimates****************************************
for (k in bb:(n-1)){
sigma.k<-(k/n)^2
xT1<-data1[(n-k+1):n]/data1[n-k]
xT2<-data2[(n-k+1):n]/data2[n-k]
xT3<-data3[(n-k+1):n]/data3[n-k]

#Non-adapted Estimates************************************
TBayes1 <- tryCatch({GIBBS(data=xT1,k=k,rho=rho1,t=data1[n-k],Xnn=data1[n],
sigma.k=sigma.k,Truncated=1,L=0,U=2)},
error=function(e) {return(c(NA,NA))})
TBayes2 <- tryCatch({GIBBS(data=xT2,k=k,rho=rho2,t=data2[n-k],Xnn=data2[n],
sigma.k=sigma.k,Truncated=1,L=0,U=2)},
error=function(e) {return(c(NA,NA))})
TBayes3 <- tryCatch({GIBBS(data=xT3,k=k,rho=rho3,t=data3[n-k],Xnn=data3[n],
sigma.k=sigma.k,Truncated=1,L=0,U=2)},
error=function(e) {return(c(NA,NA))})

gamas1  <- TBayes1$gamas; deltas1 <- TBayes1$deltas
gamas2  <- TBayes2$gamas; deltas2 <- TBayes2$deltas
gamas3  <- TBayes3$gamas; deltas3 <- TBayes3$deltas

Quants1<-QBayes(data1,n=n,k=k, gamma=gamas1, delta=deltas1,rho=rho1,pp=pp)
Quants2<-QBayes(data2,n=n,k=k, gamma=gamas2, delta=deltas2,rho=rho2,pp=pp)
Quants3<-QBayes(data3,n=n,k=k, gamma=gamas3, delta=deltas3,rho=rho3,pp=pp)

gamadens1<-density(gamas1,na.rm=T); deltadens1<-density(deltas1,na.rm=T);
qtdens1 <- density(Quants1$QT,na.rm=T); qwdens1 <- density(Quants1$QW,na.rm=T)
gamadens2<-density(gamas2,na.rm=T); deltadens2<-density(deltas2,na.rm=T);
qtdens2 <- density(Quants2$QT,na.rm=T); qwdens2 <- density(Quants2$QW,na.rm=T)
gamadens3<-density(gamas3,na.rm=T); deltadens3<-density(deltas3,na.rm=T);
qtdens3 <- density(Quants3$QT,na.rm=T); qwdens3 <- density(Quants3$QW,na.rm=T)

TEPD1[k-(bb-1)]<-gamadens1$x[which.max(gamadens1$y)];
TEPD2[k-(bb-1)]<-gamadens2$x[which.max(gamadens2$y)];
TEPD3[k-(bb-1)]<-gamadens3$x[which.max(gamadens3$y)];
```

```
QTEPD1[k-(bb-1)]<-qtdens1$x[which.max(qtdens1$y)];
QWTEPD1[k-(bb-1)]<-qwdens1$x[which.max(qwdens1$y)];
QTEPD2[k-(bb-1)]<-qtdens2$x[which.max(qtdens2$y)];
QWTEPD2[k-(bb-1)]<-qwdens2$x[which.max(qwdens2$y)];
QTEPD3[k-(bb-1)]<-qtdens3$x[which.max(qtdens3$y)];
QWTEPD3[k-(bb-1)]<-qwdens3$x[which.max(qwdens3$y)];
}
return(c(TestB1,TestB2,TestB3,TestA1,TestA2,TestA3,MOM1,MOM2,MOM3,TPA1,TPA2,TPA3,
TEPD1,TEPD2,
TEPD3,QMOM1,QTPA1,QTEPD1,QMOM2,QTPA2,QTEPD2,QMOM3,QTPA3,QTEPD3,QWTPA1,
QWTEPD1,QWTPA2,QWTEPD2,QWTPA3,QWTEPD3))
}

bb<-10 #Only consider K from 30 observations above the threshold
n <-200
nsim<-120        #2 or moreeee
alf<-4
TruncLevel<-c(1,0.99,0.9)
dist<-"Frechet"
pp<-1/n

###########################################################
Ext1<-SimulateData(dist,n,alf,quan=TruncLevel[1],Trunc=1)
Ext2<-SimulateData(dist,n,alf,quan=TruncLevel[2],Trunc=1)
Ext3<-SimulateData(dist,n,alf,quan=TruncLevel[3],Trunc=1)
Tn1<-Ext1$Tn; Tn2<-Ext2$Tn;Tn3<-Ext3$Tn
QT1<-Ext1$Q; QT2<-Ext2$Q; QT3<-Ext3$Q
QW<-Ext1$QW
##########################################################3

#Initiate Variables and matrices to be split
#----------------------------------------------
out<-matrix(0,30*(n-bb),nsim)
SlvGrid<- matrix(seq(1:nsim),1,nsim)#seq(1:nsim)

#Run the Simulate in Parallel
```

```r
#-----------------------------------------------
library(parallel)
nsim<-ifelse(nsim<2,2,nsim)
#************Checks what machine you're running on
if(Sys.info()[['nodename']]=="node0407"||Sys.info()[['nodename']]=="node0408")
{#Remote Desktop
nodelist<-ifelse(nsim<64,nsim,64)
}else if(Sys.info()[['sysname']]=="Windows"){#Normal Computer (leave 1 CPU free)
nodelist <-ifelse(nsim<24,nsim,24)
} else{#Then you must be running on a Linux Cluster (MAC OS not allowed, sorry)
nodelist <- unlist(c(read.table('nodelist.txt',sep="+")))}
cl <- makeCluster(nodelist, homogeneous=FALSE)
clusterEvalQ(cl,{c(library(ReIns),library(R.utils),library(rstan),
library(parallel),library(coda))})
clusterExport(cl, ls())
clusterExport(cl, load("CompiledEPDModel.RData",envir=.GlobalEnv))
t<-system.time({out <- tryCatch({parSapplyLB(cl,SlvGrid,SimFun)},
warning=function(w){print("Surpressed")})})
stopCluster(cl)

#Organize the returned variables coloumnwise
#-------------------------------------------------
K<-bb:(n-1)
outs<-out
Means<-rowMeans(out,na.rm=T)
mseEVI<-matrix(rowMeans((out[(6*(n-bb)+1):(15*(n-bb)),]-(1/alf))^2,na.rm=T),(n-bb))
relQerror1<-matrix(rowMeans((out[(15*(n-bb)+1):(18*(n-bb)),]/QT1),na.rm=T),(n-bb))
relQerror2<-matrix(rowMeans((out[(18*(n-bb)+1):(21*(n-bb)),]/QT2),na.rm=T),(n-bb))
relQerror3<-matrix(rowMeans((out[(21*(n-bb)+1):(24*(n-bb)),]/QT3),na.rm=T),(n-bb))
relQerror4<-matrix(rowMeans((out[(24*(n-bb)+1):(nrow(out)),]/QW),na.rm=T),(n-bb))
#Variances<-matrix(apply(out,1,var,na.rm=T),nrow=(n-bb))
#varMeans<-Variances[,7:14]
all.Estimates<-as.data.frame(cbind(K,matrix(Means,(n-bb)),mseEVI,relQerror1,
relQerror2,relQerror3,relQerror4))
names(all.Estimates)<-c("K","TestB1","TestB2","TestB3","TestA1","TestA2","TestA3","MOM1",
"MOM2","MOM3","TPA1","TPA2","TPA3","TEPD1","TEPD2","TEPD3","QMOM1","QTPA1","QTEPD1",
"QMOM2","QTPA2","QTEPD2","QMOM3","QTPA3","QTEPD3","QWTPA1","QWTEPD1","QWTPA2","QWTEPD2",
```

```
"QWTPA3","QWTEPD3","mseMOM1","mseMOM2","mseMOM3","mseTPA1",
"mseTPA2","mseTPA3","mseTEPD1", "mseTEPD2","mseTEPD3","mseQMOM1",
"mseQTPA1","mseQTEPD1","mseQMOM2","mseQTPA2","mseQTEPD2",
"mseQMOM3","mseQTPA3","mseQTEPD3","mseQWTPA1","mseQWTEPD1",
"mseQWTPA2","mseQWTEPD2","mseQWTPA3","mseQWTEPD3")
```

References

Inmaculada B Aban, Mark M Meerschaert, and Anna K Panorska. Parameter estimation for the truncated pareto distribution. *Journal of the American Statistical Association*, 101(473):270–277, 2006.

MI Fraga Alves, M Ivette Gomes, and Laurens de Haan. A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica*, 60(2):193–214, 2003.

MI Fraga Alves, M Ivette Gomes, Laurens de Haan, and Cláudia Neves. A note on second order conditions in extreme value theory: linking general and heavy tail conditions. *REVSTAT Statistical Journal*, 5 (3):285–304, 2007.

Jeffrey Annis, Brent J Miller, and Thomas J Palmeri. Bayesian inference with stan: A tutorial on adding custom distributions. *Behavior Research Methods*, 49(3):863–886, 2017.

Jan Beirlant, Yuri Goegebeur, Johan Segers, and Jozef L Teugels. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, 2006.

Jan Beirlant, Elisabeth Joossens, and Johan Segers. Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference*, 139(8):2800–2815, 2009.

Jan Beirlant, Isabel Fraga Alves, and Ivette Gomes. Tail fitting for truncated and non-truncated pareto-type distributions. *Extremes*, 19(3):429–462, 2016.

Jan Beirlant, Isabel Fraga Alves, and Tom Reynkens. Fitting tails affected by truncation. *Electronic Journal of Statistics*, 11(1):2026–2065, 2017.

Younes Bensalah. *Steps in Applying Extreme Value Theory to Finance: A Review.* Citeseer, 2000.

Enrique Castillo. Extreme value theory in engineering, statistical modeling and decision science. 1988.

Enrique Castillo and Ali S Hadi. Fitting the generalized pareto distribution to data. *Journal of the American Statistical Association*, 92(440):1609–1620, 1997.

Stuart Coles, Joanna Bawa, Lesley Trenner, and Pat Dorazio. *An Introduction to Statistical Modeling of Extreme Values*, volume 208. Springer, 2001.

Stefano Corradin and Benoit Verbrigghe. Economic risk capital and reinsurance: an application to fire claims of an insurance company. *RAS, Pianificazione Redditivita di Gruppo (December)*, 2001.

Anthony C Davison and Richard L Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):393–425, 1990.

Laurens de Haan. Slow variation and characterization of domains of attraction. In *Statistical Extremes and Applications*, pages 31–48. Springer, 1984.

Laurens De Haan and Ana Ferreira. *Extreme Value Theory: An Introduction*. Springer Science & Business Media, 2007.

Arnold LM Dekkers, John HJ Einmahl, and Laurens De Haan. A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, 17(4):1833–1855, 1989.

Ana Ferreira and Laurens de Haan. On the block maxima method in extreme value theory. *arXiv preprint arXiv:1310.3222*, 2013.

Ronald Aylmer Fisher and Leonard Henry Caleb Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge University Press, 1928.

Boris Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, pages 423–453, 1943.

Yuri Goegebeur, Armelle Guillou, and Andréhette Verster. Robust and asymptotically unbiased estimation of extreme quantiles for heavy tailed distributions. *Statistics & Probability Letters*, 87:108–114, 2014.

Scott D Grimshaw. Computing maximum likelihood estimates for the generalized pareto distribution. *Technometrics*, 35(2):185–191, 1993.

Peter Hall and AH Welsh. Adaptive estimates of parameters of regular variation. *The Annals of Statistics*, 13(1):331–341, 1985.

W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

Bruce M Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, pages 1163–1174, 1975.

Jonathan RM Hosking and James R Wallis. Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, 29(3):339–349, 1987.

Harold Jeffreys. *The Theory of Probability*. OUP Oxford, 1998.

Arthur F Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348):158–171, 1955.

Gaonyalelwe Maribe. Second-order estimation procedures for complete and incomplete heavy-tailed data. 2016.

David M Mason. Laws of large numbers for sums of extreme values. *The Annals of Probability*, 10(3): 754–764, 1982.

Jean Nuyts. Inference about the tail of a distribution: Improvementon the hill estimator. *International Journal of Mathematics and Mathematical Sciences*, 2010, 2010.

Billingsley Patrick. Probability and measure. *A Wiley-Interscience Publication, John Wiley*, 1995.

James Pickands III. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1): 119–131, 1975.

Sidney Resnick and Cătălin Stărică. Smoothing the hill estimator. *Advances in Applied Probability*, 29 (1):271–293, 1997.

Richard L Smith. Threshold methods for sample extremes. In *Statistical extremes and applications*, pages 621–638. Springer, 1984.

Richard L Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90, 1985.

Stan Development Team. Rstan: the r interface to stan. r package version 2.17. 3, 2018.

Richard Von Mises. La distribution de la plus grande de n valuers. *Rev. math. Union interbalcanique*, 1: 141–160, 1936.

Ishay Weissman. Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364):812–815, 1978.

Arnold Zellner. *An Introduction to Bayesian Inference in Econometrics*, volume 156. Wiley New York, 1971.

Chen Zhou. Existence and consistency of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, 100(4):794–815, 2009.