



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

A mixture model approach to extreme value analysis of heavy tailed processes

by
Lizo Sanqela
18082212

Supervisor: Dr Gaonyalelwe Maribe
Co-supervisor(s): Prof Frans Kanfer and Prof Sollie Millard


Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae in Advanced Data Analytics

In the Department of Statistics
In the Faculty of Natural and Agricultural Sciences
University of Pretoria

December 2023

I, *Lizo Sanqela* declare that this mini-dissertation (100 credits), which I hereby submit for the degree Magister Scientiae in Advanced Data Analytics at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature: 

Date: 7 December 2023

Abstract

Extreme value theory (EVT) encompasses statistical tools for modelling extreme events, which are defined in the peaks-over-threshold methodology as excesses over a certain high threshold. The estimation of this threshold is a crucial problem and an ongoing area of research in EVT.

This dissertation investigates extreme value mixture models which bypass threshold selection. In particular, we focus on the Extended Generalised Pareto Distribution (EGPD). This is a model for the full range of data characterised by the presence of extreme values. We consider the non-parametric EGPD based on a Bernstein polynomial approximation. The ability of the EGPD to estimate the extreme value index (EVI) is investigated for distributions in the Fréchet, Gumbel and Weibull domains through a simulation study. Model performance is measured in terms of bias and mean squared error. We also carry out a case study on rainfall data to illustrate how the EGPD fits as a distribution for the full range of data. The case study also includes quantile estimation.

We further propose substituting the Pareto distribution, in place of the GPD, as the tail model of the EGPD in the case of heavy-tailed data. We give the mathematical background of this new model and show that it is a member of the EGPD family and is thus in compliance with EVT. We compare this new model's bias and mean squared error in EVI estimation to the old EGPD through a simulation study. Furthermore, the simulation study is extended to include other estimators for Fréchet-type data. Moreover, a case study is carried out on the Belgian Secura Re data.

Acknowledgements

First and foremost, I am grateful to God for the opportunities He has graced me with.

I wish to express my utmost gratitude and appreciation to my supervisor Dr Gaonyalelwe Maribe for his invaluable insights and the guidance and mentorship he offered throughout my studies. I am also grateful to my co-supervisors Prof SM Millard and Prof FHJ Kanfer for their support.

I greatly appreciate the assistance given to me by Dr Jannie Pretorius from whom I learnt a lot about parallel programming.

A special thanks goes out to Ms Lesego Khumalo for her technical support, without which I would not have completed my simulation studies.

This mini-dissertation is heartily dedicated to my late mother
Ntombekhaya Jeanette Sanqela.

Contents

List of Abbreviations	xii
1 Introduction	1
1.1 Problem statement	1
1.2 Aim and objectives	2
1.2.1 Aim	2
1.2.2 Objectives	2
1.3 Structure of the dissertation	3
2 Extreme Value Theory	4
2.1 Classical theory of extremes	5
2.1.1 Generalised Extreme Value distribution	5
2.1.2 Generalised Pareto distribution	7
2.1.2.1 Pareto distribution	9
2.1.3 Threshold selection methods	9
2.1.3.1 Mean residual life plot	10
2.1.3.2 Pareto quantile plot	10
2.1.3.3 Parameter stability plot	10
2.1.3.4 Other methods	11
2.1.4 Methods of estimating the EVI	11

2.1.4.1	Hill estimator	11
2.1.4.2	Extended Pareto Distribution	11
2.2	Modelling the entire range of data	13
2.2.1	Extreme value mixture models	13
2.3	Areas of application	15
3	Extended Generalised Pareto Distribution	16
3.1	Semi-parametric EGPD models	18
3.1.1	Estimation of the degree of the Bernstein polynomial	21
3.2	Simulation study	24
3.2.1	Design	24
3.2.2	Results	25
3.2.2.1	Distributions in the Fréchet domain	26
3.2.2.2	Distributions in the Gumbel domain	28
3.2.2.3	Distributions in the Weibull domain	30
3.2.3	Discussion of results	32
3.3	Case study	33
3.4	Conclusion	37
4	Semi-parametric Extended Pareto Distribution	39
4.1	Pareto substitution	39
4.2	Simulation study	43
4.2.1	Design	43
4.2.2	Results	43
4.2.2.1	Pareto distributions	44
4.2.2.2	Fréchet distributions	45
4.2.3	Discussion of results	46

4.3	Case study	47
4.4	Conclusion	49
5	Conclusion	51
5.1	Future work	52
Appendices		
Appendix A Results stated without proof		57
Appendix B Derivations		58

List of Figures

2.1	Illustration of how extreme values are chosen under the <i>block maxima</i> and <i>peaks over threshold</i> methods, respectively.	8
2.2	An example of the density function of an extreme value mixture model, without a continuity constraint at the threshold.	13
3.1	Bias($\hat{\gamma}_k$) for distributions in the Fréchet domain, i.e. Fréchet($\alpha = 2$), Burr XII($\lambda = 1, \tau = 2, \eta = 1$) and half-T($\nu = 2$).	26
3.2	MSE($\hat{\gamma}_k$) for distributions in the Fréchet domain, i.e. Fréchet($\alpha = 2$), Burr XII($\lambda = 1, \tau = 2, \eta = 1$) and half-T($\nu = 2$).	27
3.3	Bias($\hat{\gamma}_k$) for distributions in the Gumbel domain, i.e. Weibull($\theta = 1, \tau = 3/2$), Gamma($\theta = 2, \kappa = 3/2$) and log-Normal($\mu = 0, \sigma^2 = 1$).	28
3.4	MSE($\hat{\gamma}_k$) for distributions in the Gumbel domain, i.e. Weibull($\theta = 1, \tau = 3/2$), Gamma($\theta = 2, \kappa = 3/2$) and log-Normal($\mu = 0, \sigma^2 = 1$).	29
3.5	Bias($\hat{\gamma}_k$) for distributions in the Weibull domain, i.e. Beta($\alpha = 1, \beta = 2$), Reversed Burr($t_1 = 1, t_2 = 2$) and Generalised Pareto Distribution (GPD)($\gamma = -1/2, \psi_t = 1$).	30
3.6	MSE($\hat{\gamma}_k$) for distributions in the Weibull domain, i.e. Beta($\alpha = 1, \beta = 2$), Reversed Burr($t_1 = 1, t_2 = 2$) and GPD($\gamma = -1/2, \psi_t = 1$).	31
3.7	Gamma Q-Q plot for positive rainfall (mm) at Chartes	33
3.8	Q-Q plots for positive rainfall (mm) at Chartes, under the GPD tail, Gamma-GPD mixture and Extended Generalised Pareto Distribution (EGPD) models.	35
3.9	Histogram of positive rainfall (mm) at Chartes with Fitted GPD, Gamma-GPD and EGPD ($m = 5$) densities overlaid.	36
3.10	Estimates of return levels for rainfall (mm) recorded at Chartes weather station.	37

4.1	Bias($\hat{\gamma}_k$) and MSE($\hat{\gamma}_k$) for Pareto($\alpha = 4$) and Pareto($\alpha = 2$) distributions. . .	44
4.2	Bias($\hat{\gamma}_k$) and MSE($\hat{\gamma}_k$) for Fréchet($\alpha = 4$) and Fréchet($\alpha = 2$) distributions. .	45
4.3	Estimates of the EVI for the Secura Belgian Re data.	48

List of Tables

3.1	Distributions in the Fréchet, Gumbel and Weibull domains of attraction, used in the simulation study.	25
3.2	Estimated shape and scale parameters of the GPD, under the GPD tail fit, Gamma-GPD mixture model and the EGPD.	34
4.1	Estimated net premium, in €, calculated for varying priority levels, R for the Secura Belgian Re data.	48
4.2	Estimated exceedance probabilities, $p = P(Y > R)$ for $R = \text{€}7,000,000$, for the Secura Belgian Re data.	49

List of Algorithms

1	Estimation of the semi-parametric EGPD model, given m	21
2	Estimation of the semi-parametric Extended Pareto Distribution (EPD) model, given m	41

List of Abbreviations

BM	Block Maxima
CDF	Cumulative Distribution Function
ECDF	Empirical Cumulative Distribution Function
EGPD	Extended Generalised Pareto Distribution
EPD	Extended Pareto Distribution
EVI	Extreme Value Index
EVT	Extreme Value Theory
GEV	Generalised Extreme Value
GPD	Generalised Pareto Distribution
IID	Independent and Identically Distributed
LSCV	Least Squares Cross Validation
ML	Maximum Likelihood
MSE	Mean Squared Error
PD	Pareto Distribution
PDF	Probability Density Function
POT	Peaks Over Threshold
PWM	Probability Weighted Moments

Chapter 1

Introduction

Extreme value theory encompasses statistical tools for modelling extreme events ([Embrechts et al., 1997](#)). Examples of extreme events include stock market crashes, floods and tropical cyclones. These kind of events occur with very little frequency but have high severity. Extremal data is thus inherently scarce ([Hu, 2013](#)). The power of extreme value models lies in their ability to fit well to this scarce extremal data and provide reliable estimates of extreme quantiles even beyond the range of the empirical distribution ([Embrechts et al., 1997](#)).

Heavy-tailed data can be observed in various fields: in computer science (e.g. sizes of files in computer networks and systems), genetics (e.g. lengths of protein sequences in an organism's genetic makeup), and insurance (e.g. the sizes of the claims filed at an insurance company). These phenomena are typically seen as anomalies, and not as events to be expected. As such, modelling of these events is not popular in the literature, as most models assume light-tail behaviour ([Nair et al., 2022](#)). Extreme value theory, however, can model the tails of distributions that exhibit heaviness. This dissertation will place a particular focus on heavy-tailed data.

The rest of this chapter details the problem to be addressed in this dissertation as well as the objectives of the study.

1.1 Problem statement

Extreme Value Theory (EVT) is concerned primarily with studying the tails of a distribution ([Frigessi et al., 2002](#)). The *peaks-over-threshold* approach to EVT is the most popular approach in the literature. This method considers extreme observations as those exceeding a certain high threshold ([Smith, 1990](#)). A crucial challenge within this framework is the determination of the threshold. Setting the threshold too low results in ordinary values being considered extreme

observations, thus leading to biased estimates in subsequent inference. In contrast, should this value be set too high, then too few extreme observations are available to carry out a meaningful analysis and the variances of parameter estimates are inflated. Threshold selection is an area of ongoing research in EVT as no single best technique exists ([Papastathopoulos and Tawn, 2013](#)).

The *peaks-over-threshold* methodology is applied when the full data, that is both non-extreme and extreme observations, are available. While tail estimation and extrapolation of risk quantities falling outside the range of the observed data is the main focus in EVT, there is also practical significance in considering the distribution of the bulk data below the threshold ([Naveau et al., 2016](#)).

There is, therefore, a need to develop models that can either efficiently estimate the threshold or bypass it and its associated uncertainties. Furthermore, such models have to simultaneously be able to model the bulk of the data and the tail well.

1.2 Aim and objectives

1.2.1 Aim

The aim of this study is to investigate and develop flexible extreme value mixture models that can model well the bulk of the distribution along with its tail when the data is characterised by the presence of extreme values.

1.2.2 Objectives

The primary objectives of this study are as follows:

- i. Conduct an extensive simulation study to investigate the performance of the semi-parametric EGPD model in estimating the extreme value index of data arising from distributions in the Gumbel, Fréchet and Weibull domains of attraction (see [Theorem 2.1](#) for an explanation of the different domains).
- ii. Investigate the suitability of the semi-parametric Extended Generalised Pareto Distribution (EGPD) as both a bulk model and a tail model, through a comparative study.
- iii. Implement the Pareto distribution as a tail model in the EGPD and study its performance on data arising from distributions in the Fréchet domain.

1.3 Structure of the dissertation

This dissertation is structured as follows:

[Chapter 2](#) provides an overview of EVT. [Section 2.1](#) details the block maxima and peak over threshold methods of choosing extreme values, the Generalised Extreme Value and Generalised Pareto distributions as well as threshold selection methods. [Section 2.2](#) discusses methods of modelling the entire range of data in the EVT context. [Section 2.3](#) details some case studies in different areas of application of EVT.

[Chapter 3](#) introduces the Extended Generalised Pareto Distribution (EGPD) for modelling the entire range of data. [Section 3.1](#) details the semi-parametric EGPD model based on Bernstein polynomial approximation. [Section 3.2](#) is a detailed simulation study investigating how the semi-parametric EGPD model fares in estimating the extreme value index of data arising from distributions in the Gumbel, Fréchet and Weibull domains of attraction. [Section 3.3](#) is a comparative study to investigate the suitability of the semi-parametric EGPD as both a bulk model and a tail model.

[Chapter 4](#) discusses the Pareto tail model substitution in the EGPD for the case of distributions in the Fréchet domain. [Section 4.1](#) details the mathematical background of the EGPD (Pareto) model and shows that it is a member of the EGPD family. A comprehensive simulation study is carried out in [Section 4.2](#). [Section 4.3](#) is a case study illustrating the practical usefulness of the EGPD (Pareto) model.

Finally, [Chapter 5](#) concludes this dissertation by summarising the main objectives and findings of the research and suggesting further research possibilities.

Chapter 2

Extreme Value Theory

Extreme Value Theory (EVT) is a branch in the field of statistics that studies the distribution of rare or unusual (from here on "extreme") events. EVT is primarily concerned with the limit distribution of the maximum (minimum) variable amongst a sequence of Independent and Identically Distributed (IID) random variables.

The earliest work in this field goes back to [Bortkiewicz \(1922\)](#) who is credited as having been the first to explore the distribution for the largest value in an IID sequence of variables. This distribution was further explored by [von Mises \(1923\)](#) who derived its expected value, and [Dodd \(1923\)](#) who focused on the median value.

Furthermore, [Fréchet \(1927\)](#) discovered one possible limit distribution for the largest order statistic among an IID sequence of variables. [Fisher and Tippett \(1928\)](#) established that limit distributions for extreme order statistics can only be one of three (3) types. [Jenkinson \(1955\)](#) combined these three distributions and introduced the Generalised Extreme Value (GEV) distribution. The work of [Gnedenko \(1943\)](#) proved pivotal as it established a rigorous mathematical foundation for the field of EVT, and developed the necessary and sufficient conditions for weak convergence of extreme order statistics.

[Juncosa \(1949\)](#) provided an extension of the work done by [Gnedenko \(1943\)](#), which investigates the asymptotic distribution of the minimum in the case when the sequence of variables is not IID. Although this work is well-founded theoretically, it has but marginal practical applicability.

2.1 Classical theory of extremes

Let Y_1, Y_2, \dots, Y_N be a sequence of IID random variables. The main concern in classical EVT is to find a limit distribution for the sample maximum of this sequence, $Y_{N:N} = \max\{Y_1, Y_2, \dots, Y_N\}$. Let F be the underlying Cumulative Distribution Function (CDF) of $\{Y_i\}$, with right endpoint at y_F , that is $y_F = \sup\{y : F(y) < 1\}$. The sample maximum, $Y_{N:N}$, converges in probability to y_F . Furthermore, the limit distribution of $Y_{N:N}$ is degenerate, which means that it is centred at a single point, as illustrated below (Smith, 1990):

$$\begin{aligned} P(\max\{Y_1, Y_2, \dots, Y_N\} \leq y) &= P(Y_1 \leq y, Y_2 \leq y, \dots, Y_N \leq y) \\ &= P(Y_1 \leq y) \times P(Y_2 \leq y) \times \dots \times P(Y_N \leq y) \\ &= F^N(y) \\ &\xrightarrow[N \rightarrow \infty]{} \begin{cases} 0 & \text{if } y < y_F \\ 1 & \text{if } y \geq y_F \end{cases} \end{aligned}$$

This degenerate distribution for the sample maximum has no inferential value. A non-degenerate distribution is thus necessary for inference regarding $Y_{N:N}$, and addressing the following pertinent problems in EVT (Smith, 1990):

- Estimation of the tail of the survival function, $\bar{F} = 1 - F$. That is, calculating $p = \bar{F}(y_*)$ when $y_* > Y_{N:N}$.
- Estimation of extreme quantities: $y_* = \bar{F}^{-1}(p)$ for small values of p .

The possibility of obtaining a non-degenerate limit distribution is explored by normalising $Y_{N:N}$. Suppose there exist sequences $(\alpha_N > 0)$ and (β_N) such that:

$$P\left(\frac{Y_{N:N} - \beta_N}{\alpha_N} \leq y\right) \xrightarrow[N \rightarrow \infty]{} H(y) \quad (2.1)$$

where H is a real-valued function which maps to the unit interval. When H is a non-degenerate CDF, it is said that F is in the maximum domain of attraction of H and is denoted $F \in MDA(H)$. The challenge is then to find the appropriate sequences (α_N) and (β_N) such that H is non-degenerate (de Haan and Ferreira, 2006).

2.1.1 Generalised Extreme Value distribution

Fisher and Tippett (1928) and Gnedenko (1943) solved the limit problem in (2.1), and in so doing established a class of non-degenerate limit distributions called *extreme values distributions*.

Theorem 2.1 (Fisher-Tippett-Gnedenko)

Let Y_1, Y_2, \dots, Y_N be a sequence of IID random variables from a population with CDF F .

If for (2.1), $F \in MDA(H)$, then H necessarily belongs to one of the following classes of distributions:

- I. *Gumbel*: $H(y) = \exp(-\exp(-y))$ for all $y, \gamma = 0$
- II. *Fréchet*: $H(y) = \exp(-y^{-1/\gamma})$ for $y > 0, \gamma > 0$
- III. *Weibull*: $H(y) = \exp(-|y|^{-1/\gamma})$ for $y < 0, \gamma < 0$



The parameter γ in [Theorem 2.1](#) is called the *Extreme Value Index (EVI)*. This parameter is of great importance in the classification of the three classes of extreme value distributions, and in measuring the tail heaviness of the underlying distributions:

- Distributions in the *Gumbel* domain of attraction ($\gamma = 0$) have an infinite right endpoint, an exponentially decaying tail, and all their moments exist.
- The *Fréchet* class ($\gamma > 0$) is also the limit case for distributions with an infinite right endpoint. However, these distributions are heavy-tailed, and their moments of orders greater than $1/\gamma$ do not exist.
- When the underlying distribution has a finite right endpoint and thus a short tail, then the limiting distribution is in the *Weibull* class ($\gamma < 0$).

The three classes of distributions in [Theorem 2.1](#) have been unified into the standard Generalised Extreme Value (GEV) distribution ([Jenkinson, 1955](#)):

$$H_\gamma(y) = \begin{cases} \exp \left[- (1 + \gamma y)^{-1/\gamma} \right] & \gamma \neq 0 \text{ and } 1 + \gamma y > 0 \\ \exp [-\exp(-y)] & \gamma = 0 \end{cases}$$

Defining $H_{\gamma,\mu,\psi}(y) = H_\gamma\left(\frac{y-\mu}{\psi}\right)$ leads to the three-parameter, non-standard GEV distribution. The parameters μ and ψ in $H_{\gamma,\mu,\psi}$ are the location and scale, respectively.

The GEV distribution is typically applied in the Block Maxima (BM) approach to EVT, illustrated in [Figure 2.1](#). For a sample of size N , this method entails subdividing the N observations into M non-overlapping blocks of size n . The blocks can sometimes appear naturally, for example, a time series of daily returns of a stock can be blocked quarterly, semi-annually or annually. In some applications, however, the number of blocks and block sizes are up to the practitioner's discretion. The set of extreme values under this approach is chosen as the maximum (minimum) observation in each block. For n large enough, the distribution of the block maxima (minima) is approximated by the GEV distribution ([Smith, 1990](#)). This distribution is commonly fitted using Maximum Likelihood (ML) estimation, or alternatively, the method of Probability Weighted Moments (PWM) estimation can be used.

Application of the BM method and fitting the GEV distribution allows the practitioner to, amongst other things, (1) estimate extreme quantiles, (2) compute the *return level*, which is the magnitude of a stress event (that is, the extreme event being modelled) and (3) compute the *return period*, that is, the frequency of occurrence of a stress event. These quantities are computed through the inverse of the CDF of the GEV distribution:

$$y_p = \begin{cases} \mu + \frac{\psi}{\gamma} \{[-\log(1-p)]^{-\gamma} - 1\} & \text{if } \gamma \neq 0 \\ \mu - \psi \log[-\log(1-p)] & \text{if } \gamma = 0 \end{cases}$$

with $p = P(Y > y)$. From the above, y_p can be interpreted as the level expected to be exceeded on average in $1/p$ periods (McNeil et al., 2015).

The BM approach has the weakness of being wasteful of data when there are multiple extremes in the same block. Furthermore, in the case when there are blocks which do not contain extremes, the GEV distribution is fitted to moderate observations, which skews inferences (Hu, 2013).

2.1.2 Generalised Pareto distribution

The Peaks Over Threshold (POT) approach to EVT offers a workaround for the issues experienced under the BM approach (Smith, 1990). This method defines extremes as those observations exceeding a high threshold, t . The main objective is then to approximate the distribution of the absolute excesses, $X = Y - t|Y > t$:

$$P(Y - t \leq x|Y > t) = \frac{F(x+t) - F(t)}{1 - F(t)} \quad (2.2)$$

where F is the CDF of Y .

The works of Balkema and De Haan (1974) and Pickands III (1975) proposed the Generalised Pareto Distribution (GPD) as an approximate distribution for the excesses above a high threshold.

Theorem 2.2 (Pickands-Balkema-de Haan)

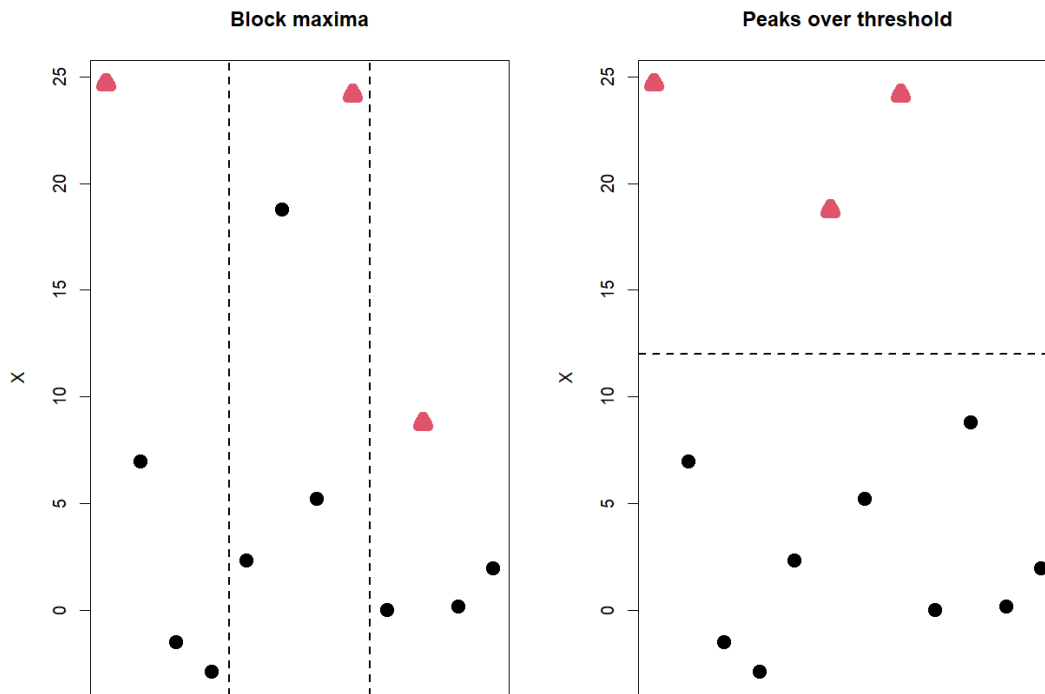
Let Y be a random variable with CDF F which has a right endpoint at y_F . For a large enough threshold, t , the conditional distribution of the excesses above the threshold, that is, the distribution of $X = Y - t|Y > t$, can be approximated by the GPD with distribution function:

$$H_{\gamma, \psi_t}(x) = \begin{cases} 1 - \left(1 + \frac{\gamma x}{\psi_t}\right)^{-1/\gamma} & \gamma \neq 0 \\ 1 - \exp\left(-\frac{x}{\psi_t}\right) & \gamma = 0 \end{cases}$$

where $\psi_t > 0$, $x \geq 0$ when $\gamma \geq 0$ and $0 \leq x \leq -\psi_t/\gamma$ when $\gamma < 0$. ▲

The parameter ψ_t is the threshold-dependent scale parameter, and γ is the shape parameter. As with the GEV distribution, γ is the EVI. A necessary and sufficient condition for the approximation in [Theorem 2.2](#) is that $F \in MDA(H_\gamma)$.

The POT method is illustrated using sample data in [Figure 2.1](#) below, alongside the BM method.



Note: ▲ indicates an extreme value.

Figure 2.1: Illustration of how extreme values are chosen under the *block maxima* and *peaks over threshold* methods, respectively.

The GPD is most commonly fitted using PWM or ML estimation to the extreme values that have been determined by the POT method ([Leadbetter, 1991](#)). Similar to using the GEV distribution, the GPD can be used to calculate return periods and return levels using its quantile function:

$$y_p = \begin{cases} \frac{\psi_t}{\gamma} (p^{-\gamma} - 1) & \text{if } \gamma \neq 0 \\ -\psi_t \log p & \text{if } \gamma = 0 \end{cases}$$

where y_p is the return level and $p = P(Y > y)$.

2.1.2.1 Pareto distribution

The POT methodology, in the case of heavy-tailed data, is usually applied to the relative excesses above the threshold (i.e. Y/t) instead of the absolute excesses. The limit distribution for the excesses is obtained from the condition of regular variation, as defined below in [Definition 2.1](#) (for an expansive definition, please see [Bingham et al. \(1989\)](#)). This condition is useful in studying the rate of convergence of distributions.

Definition 2.1 (Regular variation)

A positive, Lebesgue measurable function ([Baker, 1991](#)), a , is regularly varying with index α if

$$\lim_{t \rightarrow \infty} \frac{a(ty)}{a(t)} = y^\alpha \text{ for } y > 0 \text{ and } \alpha \in \mathbb{R}.$$

▲

The probabilistic interpretation of [Definition 2.1](#) is that a distribution function F , defined on \mathbb{R} , has positive tail index γ if its survival function \bar{F} is regularly varying with index $-1/\gamma$ ([Charpentier and Flachaire, 2021](#)):

$$P\left(\frac{Y}{t} > z | Y > t\right) \xrightarrow{P} z^{-1/\gamma} \text{ for } t \rightarrow \infty \text{ and } 1 \leq z < \infty$$

The above expression is the survival function of a Pareto Distribution (PD) with shape parameter $\alpha = 1/\gamma$ and minimum $y_m = 1$. Therefore, the distribution of the relative excesses above a sufficiently high threshold is approximated by the PD, with CDF:

$$H_{\gamma,1}(y) = \begin{cases} 0 & \text{for } y < 1 \\ 1 - y^{-1/\gamma} & \text{for } y \geq 1 \end{cases}$$

2.1.3 Threshold selection methods

Successful application of the POT method relies greatly on the threshold ([Smith, 1990](#)). As per [Theorem 2.2](#), t has to be large enough for the GPD (and PD, if applicable) approximation to be valid. Selecting a threshold, t is a trade-off between variance and bias. Should t be set too low, then the asymptotic justification of the GPD is not met, and bias is created in the parameter estimation. Conversely, choosing t too high means too few observations lie above the threshold for any meaningful inference to occur, and thus the variances of the parameters are inflated. The topic of threshold selection is an area of ongoing research in EVT as no single best technique exists ([Beirlant et al., 2022](#)). Below are a few of the most popular threshold selection methods in the literature.

2.1.3.1 Mean residual life plot

The mean residual life plot is a popular tool for finding a suitable threshold beyond which the distribution of the excesses can be approximated by the GPD (Hu, 2013), as per Theorem 2.2. The *mean excess function* is defined as

$$e(t) = E(Y - t | Y > t)$$

assuming that $E(Y) < \infty$. For a given sample (y_1, y_2, \dots, y_N) , $e(t)$ is estimated by $\hat{e}_N(t)$, defined as:

$$\hat{e}_N(t) = \frac{\sum_{k=1}^N (y_k - t) I_{(t, \infty)}(y_k)}{\sum_{k=1}^N I_{(t, \infty)}(y_k)}$$

where $I_{(t, \infty)}(z) = 1$ if $z > t$ and 0 otherwise (Beirlant et al., 2004).

In practical applications, $\hat{e}_N(t)$ is plotted at $t = y_{N-j:N}$, $j = 1, 2, \dots, N-1$ which is the $(j+1)^{th}$ largest sample observation. The plot is obtained by using the coordinates $(y_{N-j:N}, \hat{e}_{j,N})$ in increasing values of $y_{N-j:N}$, or equivalently $(j, \hat{e}_{j,N})$ in decreasing order of j . The optimal threshold is chosen as the order statistic at the point where the plot begins to show linearity (McNeil, 1999).

2.1.3.2 Pareto quantile plot

Quantile-Quantile (Q-Q) plots assess the goodness-of-fit of a proposed distribution to data by comparing their quantiles. The distributions being compared are accepted as being the same if the plot shows a linear relation, with the points on the graph lying along the 45° line. In EVT, a Pareto Q-Q plot is used with coordinates

$$\{(-\log(1 - p_k), \log y_k)\}_{k=1}^N$$

where y_1, y_2, \dots, y_N is an observed sample and $p_i = \frac{k}{N+1} \in (0, 1)$. The log observations, for a suitable fit, will lie linearly in the plot (Beirlant et al., 2004).

2.1.3.3 Parameter stability plot

The GPD shows a property of threshold stability, as is demonstrated in Lemma 2.1 below. As such, this method chooses the threshold as the point beyond which, for higher thresholds, the estimated shape parameter remains constant, while the scale parameter varies linearly.

Lemma 2.1

Given a threshold, t for which the GPD approximation is met, any higher threshold, say $u \geq t$,

will result in a GPD approximation for the excesses with the same shape parameter γ , but a modified scale parameter $\psi_* = \psi_t + \gamma(u - t)$. ▲

2.1.3.4 Other methods

The graphical methods described above are difficult to use when the sample size is large. Therefore, arbitrary thresholds are set at the discretion of the practitioner. A common choice is to use a high quantile, like the 95th quantile, as the threshold.

2.1.4 Methods of estimating the EVI

The extreme value index (EVI) describes the tail heaviness of a distribution. This parameter is thus crucial in estimating tail probabilities as well as extreme quantiles. Several methods for estimating the EVI are discussed below.

The GPD is fitted to the excesses over a high threshold using ML or PWM estimation. The strength of the GPD is in its ability to estimate the EVI on \mathbb{R} . That is, the GPD can provide an estimate for the EVI regardless of whether the tail of the underlying distribution is short, light or heavy. It is important to note, however, that fitting the GPD using classical PWM estimation is recommended for small values of the EVI, usually $\gamma \in [-1/2, 1/2]$ (Naveau et al., 2016).

2.1.4.1 Hill estimator

The Hill estimator, developed by Hill (1975), is a popular estimator of the EVI for heavy-tailed data ($\gamma > 0$). The estimator, which is based on the top $k + 1$ order statistics is defined as

$$H_{k,N} = \frac{1}{k} \sum_{j=1}^k \log \frac{Y_{N-k+j:N}}{Y_{N-k:N}} \text{ for } k = 1, 2, \dots, N - 1$$

where $Y_{1:N} \leq Y_{2:N} \leq \dots \leq Y_{N-1:N} \leq Y_{N:N}$ are the order statistics. One challenge in using the Hill estimator is determining the appropriate value of k . A visual tool that can be used to choose this value is the Hill plot with coordinates $(k, H_{k,N})$ in increasing values of k . The appropriate value of k is chosen where the estimate shows stability (McNeil, 1999).

2.1.4.2 Extended Pareto Distribution

Beirlant et al. (2009) developed the EPD for modelling the excesses above a threshold in the case of heavy-tailed data. The EPD has CDF:

$$H_{\gamma,\delta,\tau}(y) = \begin{cases} 0 & \text{for } y \leq 1 \\ 1 - [y(1 + \delta - \delta y^\tau)]^{-1/\gamma} & \text{for } y > 1 \end{cases}$$

for $\tau < 0$ and $\delta > \max\{-1, 1/\tau\}$. This distribution is known to reduce the bias of estimates of the EVI. Furthermore, the EPD can be fitted to bigger portions of data, as it allows for lower thresholds to be used in the POT methodology.

2.2 Modelling the entire range of data

The POT method is applied when the full data (that is, both moderate and extreme data) are available. Modelling excesses over a threshold and discarding the rest of the data may be a wasteful exercise. Therefore, there is a great interest in literature to model the bulk and tail in unison [Hu \(2013\)](#).

This section provides a brief overview of some of the existing methodologies for modelling the entire range of data.

2.2.1 Extreme value mixture models

Extreme value mixture models emerged as an objective approach to choosing a suitable threshold for the POT modelling. The main concept behind these models is to construct a two-component model; wherein the first component captures the bulk (non-extreme or moderate) data, and the other is a model for the excesses over the threshold (i.e a tail model). The model fitted to the bulk component may be (1) parametric, (2) semi-parametric or (3) non-parametric, while the tail is typically modelled with a GPD. These models typically treat the threshold as a parameter to be estimated, thereby allowing for quantifying the uncertainties associated with threshold selection. [Figure 2.2](#) below illustrates the density function of an extreme value mixture model.

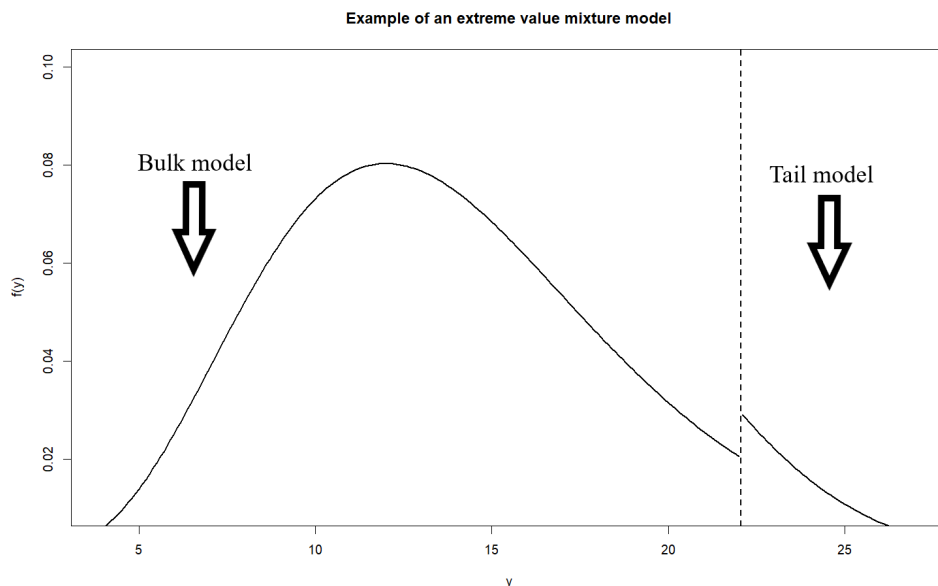


Figure 2.2: An example of the density function of an extreme value mixture model, without a continuity constraint at the threshold.

An overview of the mixture model(s) which will be used in this dissertation is given below. A comprehensive discussion of extreme value mixture models can be found in [Hu \(2013\)](#).

[Behrens et al. \(2004\)](#) presented the most basic form of a mixture model. They proposed a model to fit the full data (characterised by extreme values) where the threshold is directly estimated. Let $\{Y_i\}_{i=1}^N$ be an IID sequence of variables, and t a threshold, then this model fits a GPD to $\{Y_i - t | Y_i \geq t\}$ and takes $\{Y_i | Y_i < t\} \sim F(y|\beta)$, where $F(y|\beta)$ can be a Gamma, Weibull or Gaussian CDF, with β being the parameter vector for the chosen distribution. This model has CDF:

$$G(y|\beta, t, \gamma, \psi_t) = \begin{cases} F(y|\beta) & \text{for } y < t \\ F(t|\beta) + [1 - F(t|\beta)] H_{\gamma, \psi_t}(y - t) & \text{for } y \geq t \end{cases}$$

An advantage of this model is that it is straightforward and flexible. Furthermore, Bayesian inference is used throughout and thus expert prior information is drawn from to compensate for the sparsity of extremal data. However, Bayesian inference treats the threshold and ψ_t as independent parameters, thereby ignoring the relationship which clearly exists between them. Moreover, this particular formulation of the model ignores the tail fraction, $\phi_t = P(Y > t)$. This quantity can be included by alternatively defining the model as:

$$G(y|\beta, t, \gamma, \psi_t, \phi_t) = \begin{cases} (1 - \phi_t) \cdot \frac{F(y|\beta)}{F(t|\beta)} & \text{for } y < t \\ (1 - \phi_t) + \phi_t \cdot H_{\gamma, \psi_t}(y - t) & \text{for } y \geq t \end{cases} \quad (2.3)$$

ML estimation is used to fit this form of the mixture model. This can be performed with the `evmix` package in R. It is worth noting that both formulations of this model have Probability Density Function (PDF) that are discontinuous at the threshold. A continuity constraint can be added, but it has been found in the literature to have a negligible effect on the model fit ([Hu, 2013](#)).

2.3 Areas of application

The different methodologies of EVT have been applied in several contexts including finance and insurance. [McNeil \(1999\)](#) demonstrated the quantification of market risk using Value-at-risk and expected shortfall as risk measures, which are calculated using the GPD fitted to excesses over a threshold. [Aviv \(2018\)](#) employed EVT techniques to estimate the loss distributions in the reinsurance context. The GEV distribution is used by [Vyskocil and Koudelka \(2021\)](#) to estimate an insurance company's losses, emanating from operational risk.

The fields of hydrology and meteorology have also made great use of the EVT framework. [Nadarajah and Shiao \(2005\)](#) utilised the GEV distribution on the Taiwanese Pachang River's daily stream-flow data and their inference included estimating the recurrence intervals for flood volume and peak. A study of flood frequencies in eleven southern African countries by [Mkhandi et al. \(2000\)](#) demonstrated that the Generalised Pareto and GEV distributions are suitable for use in such analyses. [Tabari \(2021\)](#) compared the fundamental approaches in EVT, namely BM and POT, in analysing the impact of future climate change on extreme precipitation and global floods. EVT was used by [Maposa et al. \(2021\)](#) in studying extreme temperatures in the Limpopo province of South Africa, which has become prone to heat waves and a shortage of rainfall.

Further applications of EVT include the field of engineering as can be seen in [Chryssolouris et al. \(1994\)](#). [Zheng et al. \(2014\)](#), as well as [Zheng and Sayed \(2019\)](#), applied EVT methodology in road safety analysis. EVT methods have also proven useful in modelling data from a sports context, as [Vicente \(2012\)](#) demonstrated.

Chapter 3

Extended Generalised Pareto Distribution

EVT is frequently applied to sample maxima or high extremes, and thus the upper tail of a distribution, with little focus on the lower tail. Consider a sequence of values, say $\{y_k\}$, then $\min\{y_k\} = -\max\{-y_k\}$. Thus, a sign change allows for applying the same EVT techniques to the lower tail, as would be applied to the upper tail. Under the POT methodology, the threshold (t) would naturally have to be *low* in this case.

Define the random variable $X = -Y$, where Y is non-negative. The distribution of the excesses of this new variable beyond a threshold, $-t$, is approximately GPD, according to [Theorem 2.2](#), with a negative shape parameter, say $\gamma = -1/\kappa$ for some $\kappa > 0$. That is,

$$P(Z = X + t > z | X > -t) \rightarrow \bar{H}_{-1/\kappa, \psi_t}(z) \text{ as } t \rightarrow 0 \text{ and } 0 < z < t$$

The shape parameter is necessarily negative as the underlying distribution is bounded above, and thus short-tailed. Moreover, the threshold has to be chosen such that $\bar{H}_{-1/\kappa, \psi_t}(t) = 0$. This constraint further leads to the condition that low values of Y should be described adequately by a power law ([Naveau et al., 2016](#)):

$$P(Y \leq y) \propto y^\kappa \text{ for small } y \geq 0 \text{ and } \kappa > 0$$

Moving away from exclusively modelling the tail; the Extended Generalised Pareto Distribution (EGPD) is introduced with the aim of modelling the full range of data with a density (distribution) function which is in compliance with EVT in both tails, whilst bypassing the selection of a threshold. [Naveau et al. \(2016\)](#) begin deriving the EGPD by considering the probability integral transform described in [Result A.1](#). Random draws from a GPD can be obtained by defining

$$Z = H_{\gamma, \psi_t}^{-1}(U) \text{ for } U \sim \mathcal{U}(0, 1)$$

because $Z \sim GPD(\gamma, \psi_t)$ under this construction, as per [Result A.2](#). Flexibility can be added to the above scheme by simply replacing the uniform draws with $V = G^{-1}(U)$, where G is a

CHAPTER 3. EXTENDED GENERALISED PARETO DISTRIBUTION

continuous CDF defined on $[0, 1]$. Note that the previous formulation is a special case when $G(U) = U$. A family of richer random variables is obtained by defining:

$$Z = H_{\gamma, \psi_t}^{-1} \{G^{-1}(U)\} \quad (3.1)$$

The main problem now is to find a class of distributions for G that preserve the upper tail behaviour of a GPD with shape parameter γ . Furthermore, the CDF of Z , for values near zero, must behave like the function z^κ , with $\kappa > 0$, as discussed above. These constraints are met when the following conditions are fulfilled:

- (i) $\lim_{u \rightarrow 0} \frac{\bar{G}(1-u)}{u} = a$ for some finite $a > 0$.
- (ii) $\lim_{u \rightarrow 0} \frac{G\{\omega(u)\}}{G(u)} = b$ for some finite $b > 0$, and ω is an arbitrary positive function satisfying $\omega(u) = 1 + o(u)$, as $u \rightarrow 0$.
- (iii) $\lim_{u \rightarrow 0} \frac{G(u)}{u^\kappa} = c$ for some finite $c > 0$.

These conditions are derived in [Naveau et al. \(2016\)](#). Below is an explanation of each of the conditions. First, note that the CDF and survival function of Z as defined in (3.1), are $F(z) = G\{H_{\gamma, \psi_t}(z)\}$ and $\bar{F}(z) = \bar{G}\{H_{\gamma, \psi_t}(z)\}$, respectively. Thus, taking $u = \bar{H}_{\gamma, \psi_t}(z)$ in condition (i) has the implication that the ratio $\bar{F}(z)/\bar{H}_{\gamma, \psi_t}(z)$ converges to a constant as $z \rightarrow \infty$. Therefore, condition (i) ensures that the upper tail of Z is equivalent to a GPD tail, $\bar{H}_{\gamma, \psi_t}(z)$.

Consider the ratio $F(z)/G(z)$, which by using (3.1) can be expressed as:

$$\frac{F(z)}{G(z)} = \frac{G\{H_{\gamma, \psi_t}(z)\}}{G(z)} = \frac{G\{z[H_{\gamma, \psi_t}(z)/z]\}}{G(z)} = \frac{G\{\omega(u)\}}{G(u)}$$

where $\omega(u) = H_{\gamma, \psi_t}(u)/u$. Therefore, by condition (ii), the ratio $F(z)/G(z)$ converges to some non-null constant for small values of z . Thus, this condition implies that the CDF of Z , for low values, is driven by G . Moreover, condition (iii) ensures that G adopts the behaviour of a Weibull-type (i.e. short-tailed) GPD.

Following from these conditions, the EGP family is defined below ([Naveau et al., 2016](#)).

Definition 3.1 (EGPD Family of Distributions)

A distribution function F is a member of the Extended Generalised Pareto Distribution (EGPD) family if it can be expressed as

$$F(z) = G\{H_{\gamma, \psi_t}(z)\} \quad \forall z > 0$$

where H_{γ, ψ_t} is the CDF of a GPD with shape parameter γ and scale parameter ψ_t . G is the CDF of a continuous distribution on the unit interval. The corresponding PDF of a distribution in this family is

$$f(z) = g\{H_{\gamma, \psi_t}(z)\} \cdot h_{\gamma, \psi_t}(z) \quad \forall z > 0$$

▲

Four parametric families for the CDF G in the EGPD family, which meet conditions (i), (ii) and (iii) above have been studied in (Naveau et al., 2016):

(i) $G(u) = u^\kappa$ for $\kappa > 0$.

(ii) $G(u) = pu^{\kappa_1} + (1-p)u^{\kappa_2}$ for $0 < \kappa_1 \leq \kappa_2$ and $p \in [0, 1]$ which is a mixing probability for this mixture.

(iii) $G(u) = 1 - F_{\beta_{1/\delta, 2}} \{(1-u)^\delta\}$ for $\delta > 0$, where $F_{\beta_{1/\delta, 2}}$ is the CDF of a Beta distribution with parameters $1/\delta$ and 2.

(iv) $G(u) = \left[1 - F_{\beta_{1/\delta, 2}} \{(1-u)^\delta\}\right]^{\kappa/2}$ for $\delta, \kappa > 0$, where $F_{\beta_{1/\delta, 2}}$ is as given in (iii).

Model (i), the simplest of these parametric models, has been found to be the best performing. Model (iii) is sometimes comparable in performance to model (i), however, it suffers from the poor estimation of its skewness parameter, δ .

In practice, the EGPD is fitted to data, used to estimate the EVI and calculate tail measures. Furthermore, the definition of Z in (3.1) is useful in that it allows for an explicit formula for computing the quantiles of the EGPD:

$$z_p = \begin{cases} \frac{\psi_t}{\gamma} \left\{ [1 - G^{-1}(p)]^{-\gamma} - 1 \right\} & \text{if } \gamma > 0 \\ -\psi_t \log \{1 - G^{-1}(p)\} & \text{if } \gamma = 0 \end{cases}$$

These quantiles are then used to compute return levels, as is done in the classical tail models discussed in Chapter 2.

3.1 Semi-parametric EGPD models

Parametric forms of the function G in the EGPD family are convenient as they are fast to implement and simple to interpret. Furthermore, as seen in models (i) – (iv), there are few parameters to estimate when fitting these models (Naveau et al., 2016). However, besides these conveniences, there are no theoretical reasons for restricting the EGPD to these parametric forms.

Tencaliec et al. (2019) proposed a non-parametric form for G based on Bernstein polynomials. This choice is motivated by the fact that any continuous function on $[0, 1]$ can be approximated to some degree by Bernstein polynomials, as shown by Bernstein (1912). These polynomials are defined below.

Definition 3.2 (Bernstein polynomial)

The Bernstein polynomial approximation of a continuous function f defined on $[0, 1]$, at the point u , is defined as

$$B_m(f, u) = \sum_{i=0}^m f\left(\frac{i}{m}\right) \cdot b_{i,m}(u)$$

where the quantities

$$b_{i,m}(u) = \binom{m}{i} u^i (1-u)^{m-i}, u \in [0, 1]$$

are the non-negative Bernstein bases. Furthermore, $m \in \mathbb{N}$ is the degree of the Bernstein polynomial. ▲

Vitale (1975) studied the use of Bernstein polynomials in estimating density (distribution) functions defined on the unit interval. This forms the basis for choosing a non-parametric form for G in the EGPD. Since G is a CDF, a sample (from G) is necessary for computing its Bernstein polynomial approximation. Vitale proposed replacing the unknown function (G) by its Empirical Cumulative Distribution Function (ECDF), G_N . Therefore, the Bernstein polynomial approximation of G in line with Definition 3.2 is

$$B_m(G, u) = \sum_{i=0}^m G_N\left(\frac{i}{m}\right) \cdot b_{i,m}(u)$$

Since G is continuous, its PDF is obtained by computing $\frac{d}{du} B_m(G, u)$. Therefore, the PDF of G can be approximated by a Bernstein polynomial of degree m as (see Result B.1 for proof):

$$\hat{g}_{m,N}(u) = m \sum_{i=0}^{m-1} \left\{ G_N\left(\frac{i+1}{m}\right) - G_N\left(\frac{i}{m}\right) \right\} b_{i,m-1}(u)$$

Result B.2 shows that the above PDF can be equivalently, and more importantly, conveniently expressed as:

$$\hat{g}_{m,N}(u) = \sum_{k=1}^m \omega_{k,m} \beta_{k,m-k+1}(u) \tag{3.2}$$

where $\omega_{k,m} = G_N(k/m) - G_N((k-1)/m)$ are the mixing proportions based on the ECDF of G and $\beta_{k,m-k+1}$ is the density of a Beta distribution with parameters k and $m-k+1$. It is important to note that (3.2) is not a mixture of Betas, in the classical sense, because there is no latent component. This is because the number of components in the mixture (m), as well as the parameters of each Beta distribution (k and $m-k+1$), are known beforehand. Moreover, the mixing proportions, $\omega_{k,m}$, are straightforward to calculate when G_N is available.

The approximated CDF, $\widehat{G}_{m,N}(t)$, can be derived from (3.2) as follows:

$$\begin{aligned}
 \widehat{G}_{m,N}(u) &= \int_0^u \sum_{k=1}^m \omega_{k,m} \beta_{k,m-k+1}(t) dt \\
 &= \sum_{k=1}^m \omega_{k,m} \int_0^u \beta_{k,m-k+1}(t) dt \\
 &= \sum_{k=1}^m \omega_{k,m} F_{\beta_{k,m-k+1}}(u)
 \end{aligned} \tag{3.3}$$

where $F_{\beta_{k,m-k+1}}$ is the CDF of a Beta distribution with parameters k and $m-k+1$. Therefore, the EGPD family based on the Bernstein polynomial approximation is defined as:

$$F_{m,N}(z) = \widehat{G}_{m,N} \{H_{\gamma,\psi_t}(z)\} \text{ for } z > 0 \tag{3.4}$$

where $\widehat{G}_{m,N}$ is a CDF corresponding to $\widehat{g}_{m,N}$ as defined in (3.3). [Tencaliec et al. \(2019\)](#) show that the above specification meets conditions (i), (ii) and (iii) as required under the EGPD when $\omega_{m,m} > 0$ because:

(i) $\lim_{z \rightarrow 0} \frac{F_{m,N}(z)}{z^s} = \psi_t^{-s} \binom{m-1}{s-1} \omega_{s,m} > 0$ where s is the position of the first positive weight in ω . This satisfies conditions (ii) and (iii) of the EGPD family.

(ii) $\lim_{z \rightarrow \infty} \frac{\overline{F}_{m,N}(z)}{\overline{H}_{\gamma,\psi_t}(z)} = m\omega_{m,m} > 0$. This satisfies condition (i) of the EGPD family.

(iii) Let Y be a non-negative continuous random variable satisfying $P(Y > z+t|Y > t) \xrightarrow[t \rightarrow \infty]{} (1 + \frac{\tilde{\gamma}z}{\psi_t})^{-1/\tilde{\gamma}}$.

If $Y \sim EGPD$, i.e. $Y = H_{\gamma,\psi_t}^{-1} \{ \widehat{G}_{m,N}^{-1}(U) \}$ with $U \sim \mathcal{U}(0,1)$, and we have that

$$\lim_{u \rightarrow 0} \frac{\widehat{G}_{m,n}(1-u)}{u} > 0, \text{ then } \tilde{\gamma} = \gamma.$$

Condition (iii) above ensures that the upper tail behaviour of the EGPD model is governed by γ as is the case when a genuine EVT argument is used to justify the GPD.

With the theoretical background in place, the EGPD can now be used for inference. Algorithm 1 below, adapted from [\(Tencaliec et al., 2019\)](#), details how the EGPD model is fitted, given data (\mathbf{y}) and the degree of the Bernstein polynomial (m):

Algorithm 1: Estimation of the semi-parametric EGPD model, given m

1. Obtain initial values for $\theta^{(0)} = (\gamma^{(0)}, \psi_t^{(0)})$. [Tencaliec et al. \(2019\)](#) suggest fitting the parametric EGPD model (i) with $G(u) = u^\kappa$ to \mathbf{y} to obtain these values.
2. At the i^{th} iteration:
 - 2.1. Calculate the weights, $\omega^{(i)}$, as follows:
 - 2.1.1. Generate the pseudo-observations: $\hat{U} = H_{\gamma^{(i-1)}, \psi_t^{(i-1)}}(y)$.
 - 2.1.2. Calculate G_N , the ECDF based on \hat{U} .
 - 2.1.3. Calculate $\omega_{k,m} = G_N(k/m) - G_N((k-1)/m)$.
 - 2.1.4. Ensure the last weight is non-null. If $\omega_{m,m} = 0$, then:
 - 2.1.4.1. Calculate: $\omega_{m,m} = 1 - \hat{G}_{m,N}(1 - 1/m)$.
 - 2.1.4.2. Normalise the weights: $\omega = \omega / \sum \omega$.
 - 2.2. Generate the pseudo-observations: $\hat{Z} = H_{\gamma^{(i-1)}, \psi_t^{(i-1)}}(y)$.
 - 2.3. Calculate $\hat{V} = \hat{G}_{m,N}(\hat{Z})$ using the fitted weights $\omega^{(i)}$.
 - 2.4. Generate GPD variables, $\hat{X} = H_{\gamma^{(i-1)}, \psi_t^{(i-1)}}^{-1}(\hat{V})$, and compute the PWM estimates:

$$\hat{\gamma}^{(i)} = \frac{m_0 - 4m_1}{m_0 - 2m_1} \text{ and } \hat{\psi}_t^{(i)} = m_0 (1 - \hat{\gamma}^{(i)})$$

$$\text{where } m_0 = \overline{\hat{X}} \text{ and } m_1 = \frac{1}{N} \sum_{j=1}^N \frac{N-j}{N-1} \hat{X}_j \text{ with } \hat{X}_1 \leq \hat{X}_2 \leq \dots \leq \hat{X}_N.$$

3. Check for convergence: if $|\hat{\gamma}^{(i)} - \hat{\gamma}^{(i-1)}| < 10^{-3}$ the algorithm has converged, else repeat (2) above.

A non-parametric bootstrap can be used to obtain $100(1 - \alpha)\%$ confidence intervals as well as the standard errors of the parameters γ and ψ_t .

3.1.1 Estimation of the degree of the Bernstein polynomial

[Algorithm 1](#) details the fitting of the EGPD model when the degree of the Bernstein polynomial, m , is known. In practice, this value is not known and is thus determined through Least Squares Cross Validation (LSCV). This method finds the appropriate value for m as that value which minimises ([Tencaliec et al., 2019](#)):

$$\begin{aligned} & \int_y [\hat{f}_{m,N}(y) - f(y)]^2 dy \\ &= \int_y \hat{f}_{m,N}^2(y) dy - 2 \int_y \hat{f}_{m,N}(y) f(y) dy + \int_y f^2(y) dy \end{aligned}$$

The last term in the above expression does not depend on m and is thus left out of subsequent calculations. Furthermore, [Silverman \(1986\)](#) showed that sample data can be used to estimate the second term:

$$2 \int_y \widehat{f}_{m,N}^2(y) f(y) dy \approx \frac{2}{N} \sum_{j=1}^N \widehat{f}_{m,N}^{(-j)}(Y_j)$$

with $\widehat{f}_{m,N}^{(-j)}$ being the approximated PDF corresponding to the CDF in [\(3.3\)](#), but calculated based on all the data except Y_j . The LSCV criterion can thus be written as:

$$LSCV(m) = \int_0^\infty \widehat{f}_{m,N}^2(y) dy - \frac{2}{N} \sum_{j=1}^N \widehat{f}_{m,N}^{(-j)}(Y_j)$$

Recall that $\widehat{f}_{m,N}(y) = \widehat{g}_{m,N} \{H_{\gamma,\psi_t}(y)\} \cdot h_{\gamma,\psi_t}(y)$. Define $u = H_{\gamma,\psi_t}(y) = 1 - \left(1 + \frac{\gamma y}{\psi_t}\right)^{-1/\gamma}$, then we can rewrite

$$\widehat{f}_{m,N}(y) = \frac{1}{\psi_t} \widehat{g}_{m,N}(u) \cdot (1-u)^{1+\gamma}$$

Therefore, consider a change of variable in the integral in the calculation of $LSCV(m)$ from y to $u = H_{\gamma,\psi_t}(y)$. The new bounds are:

$$y = 0 \implies u = H_{\gamma,\psi_t}(0) = 0 \text{ and } y = \infty \implies u = H_{\gamma,\psi_t}(\infty) = 1$$

and we can write $du = \frac{1}{\psi_t} \left(1 + \frac{\gamma y}{\psi_t}\right)^{-1+\gamma} dy = \frac{1}{\psi_t} (1-u)^{1+\gamma} dy$. This implies that the term in the integrand is

$$\widehat{f}_{m,N}^{(2)}(y) = \frac{1}{\psi_t^2} \widehat{g}_{m,N}^2(u) \cdot \{(1-u)^{1+\gamma}\}^2 = \frac{1}{\psi_t} \widehat{g}_{m,N}^2(u) \cdot (1-u)^{1+\gamma} \frac{du}{dy}$$

Therefore, we express $LSCV(m)$ as follows:

$$\begin{aligned} LSCV(m) &= \int_0^1 \frac{1}{\psi_t} \widehat{g}_{m,N}^2(u) \cdot (1-u)^{1+\gamma} du - \frac{2}{N\psi_t} \sum_{j=1}^N \widehat{g}_{m,N}^{(-j)} \{H_{\gamma,\psi_t}(Y_j)\} \cdot h_{\gamma,\psi_t}(Y_j) \\ &= \frac{1}{\psi_t} \left\{ \int_0^1 \widehat{g}_{m,N}^2(u) \cdot (1-u)^{1+\gamma} du - \frac{2}{N} \sum_{j=1}^N \widehat{g}_{m,N}^{(-j)}(Z_j) \cdot (1-Z_j)^{1+\gamma} \right\} \end{aligned}$$

where $Z_j = H_{\gamma,\psi_t}(Y_j)$. Applying the formula derived by [Leblanc \(2010\)](#) leads to $\psi_t LSCV(m)$ being written as:

$$\frac{m^2}{2m+\gamma} \mathbf{G}^T \mathbf{A} \mathbf{G} - \frac{2}{N-1} \left\{ \sum_{j=1}^N \widehat{g}_{m,N}(Z_j) \cdot (1-Z_j)^{1+\gamma} - \frac{1}{N} \sum_{j=1}^N \beta_{k_j+1, m-k_j}(Z_j) \cdot (1-Z_j)^{1+\gamma} \right\}$$

where $k_j = [mZ_j]$, $\mathbf{G} : m \times 1$ is a vector with elements $(g_k) = G_N(k/m) - G_N((k-1)/m)$, and $\mathbf{A} : m \times m$ is a matrix with elements:

$$(a_{i,j}) = \frac{\binom{m-1}{i} \binom{m-1}{j}}{\binom{2(m-1)}{i+j}} \text{ for } i, j = 0, 1, \dots, m-1$$

In practice a sequence of values of m is chosen and $LSCV(m)$ is calculated for each value in the sequence. The appropriate m is then chosen as the value with the corresponding least LSCV.

The rest of this chapter will evaluate the ability of the EGPD to estimate the EVI through a simulation study. A case study will also be conducted to examine the suitability of the EGPD in modelling the entire range of data and quantile estimation.

3.2 Simulation study

A comprehensive simulation study is carried out to investigate the performance of the semi-parametric EGPD model in estimating the extreme value index (EVI) of data arising from distributions in the Gumbel, Fréchet and Weibull domains of attraction. Model performance is measured in terms of the bias and Mean Squared Error (MSE) of the estimator, as defined below:

$$\begin{aligned} \text{Bias}(\hat{\gamma}) &= \mathbb{E}(\hat{\gamma}) - \gamma \\ \text{MSE}(\hat{\gamma}) &= \mathbb{E}[(\hat{\gamma} - \gamma)^2] \end{aligned}$$

3.2.1 Design

The simulation study is designed as follows:

- 1000 samples of size 1000 each are taken from 3 distributions in the Fréchet, Gumbel and Weibull domains of attraction, respectively.
- For each sample, the EVI ($\hat{\gamma}_k$) is estimated for the threshold exceedances:

$$Y_{N-k+j:N} - Y_{N-k:N}, \quad j = 1, 2, \dots, k$$

for each $k = 2, 3, \dots, N$, via the EGPD and the following well-known models:

- (i) GPD, (ii) Hill estimator and (iii) EPD for distributions in the Fréchet domain of attraction.
- GPD for distributions in the Gumbel and Weibull domains of attraction.
- The performance for each estimator is measured by Bias and MSE where for each k , $\text{Bias}(\hat{\gamma}_k)$ and $\text{MSE}(\hat{\gamma}_k)$ is calculated as the average across the 1000 samples.

Table 3.1 below details the distributions which are used in the simulation study, along with the chosen parameters. The chosen values of the EVI for the different domains of attraction are also given in the table. A comprehensive list of distributions in the different domains and how they are parameterised to achieve specific values of the EVI can be found in Tables 2.1, 2.2 and 2.3 in [Beirlant et al. \(2004\)](#).

Distribution	Chosen Parameters
<i>Distributions in the Fréchet domain ($\gamma = 1/2$)</i>	
Fréchet(α)	$\alpha = 2$
Burr XII(λ, τ, η)	$\lambda = 1, \tau = 2, \eta = 1$
Half- $t(\nu)$	$\nu = 2$
<i>Distributions in the Gumbel domain ($\gamma = 0$)</i>	
Weibull(θ, τ)	$\theta = 1, \tau = 3/2$
Gamma(θ, κ)	$\theta = 2, \kappa = 3/2$
log-Normal(μ, σ^2)	$\mu = 0, \sigma^2 = 1$
<i>Distributions in the Weibull domain ($\gamma = -1/2$)</i>	
Beta (α, β)	$\alpha = 1, \beta = 2$
Reversed Burr (t_1, t_2)	$t_1 = 1, t_2 = 2$
GPD(γ, ψ_t)	$\gamma = -1/2, \psi_t = 1$

Table 3.1: Distributions in the Fréchet, Gumbel and Weibull domains of attraction, used in the simulation study.

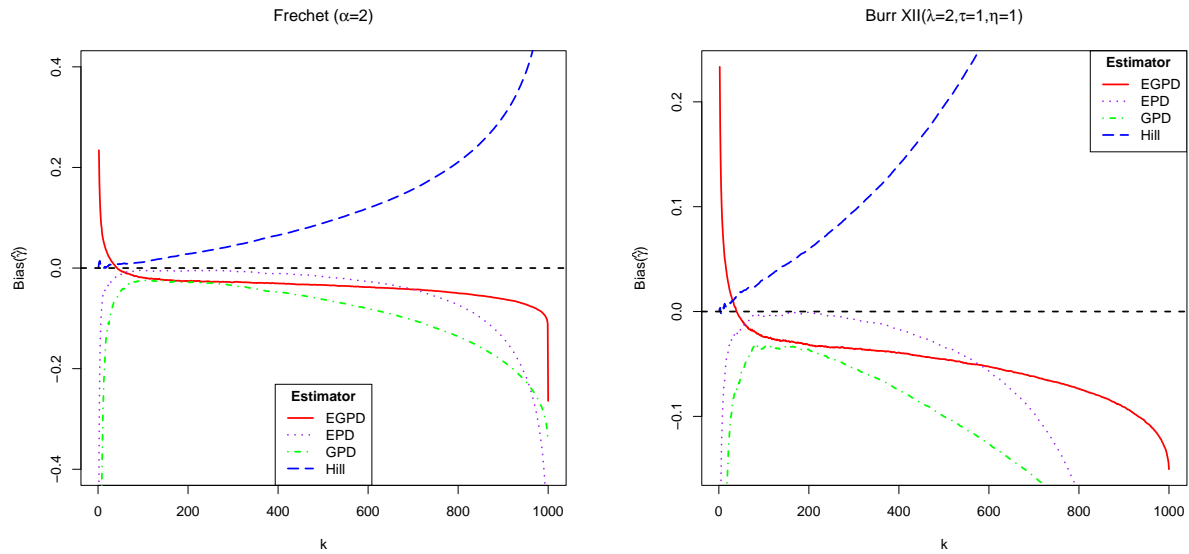
3.2.2 Results

The appropriate values of the degree of the Bernstein polynomial (m) are found through LSCV as follows: for each domain; the value of m is calculated for each of the 1000 samples from each of the chosen distributions in Table 3.1. The final value of m in each domain is then determined as the median of these values, as this measure is insensitive to the outliers that may occur due to sample variation. The results of this process are as follows: $m = 30$ in the Fréchet domain, $m = 40$ in the Gumbel domain and $m = 95$ in the Weibull domain. These are the values of m used in the estimation of the EVI.

Next, the bias and MSE are given for each of the considered models in the different domains of attraction.

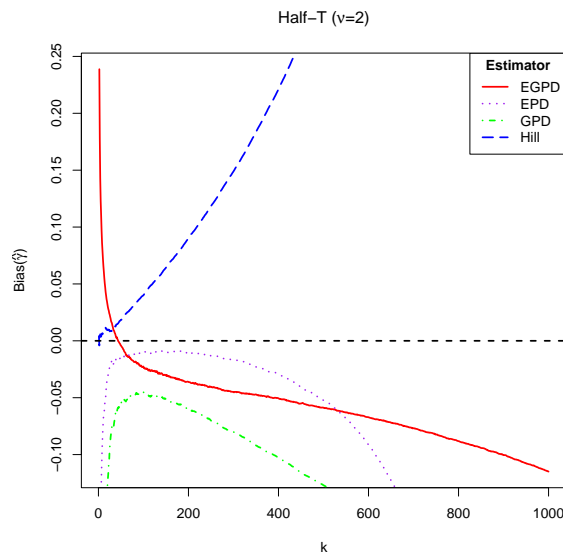
3.2.2.1 Distributions in the Fréchet domain

Figure 3.1 below illustrates the bias of the estimators of the EVI for distributions in the Fréchet domain. The EVI is taken to be $1/2$.



(a) Fréchet($\alpha = 2$) distribution

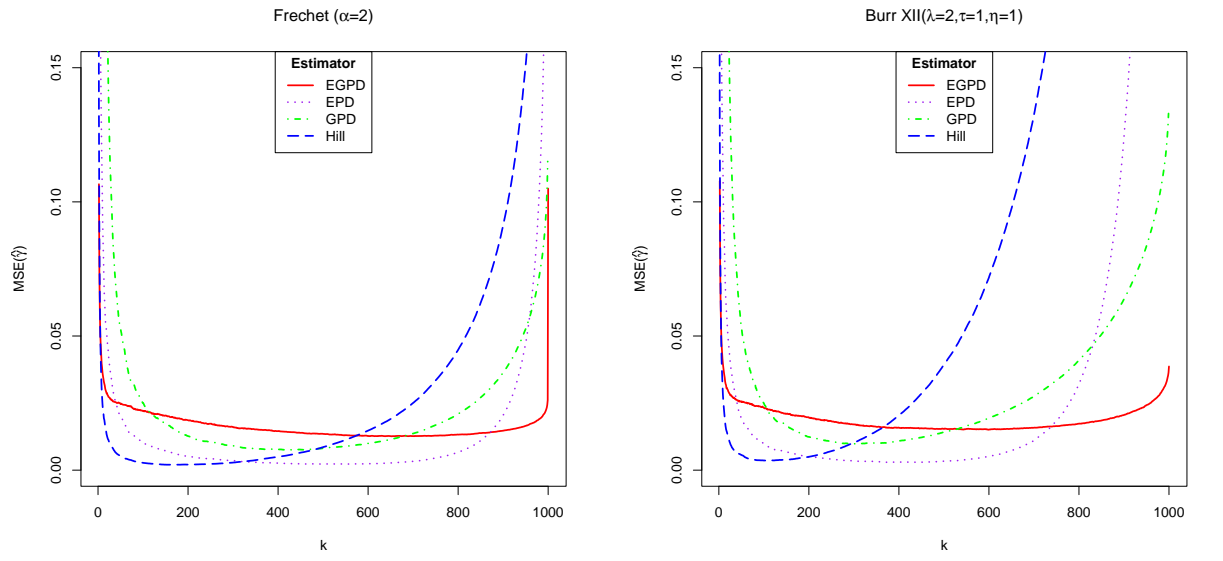
(b) Burr XII($\lambda = 1, \tau = 2, \eta = 1$) distribution



(c) half-T($\nu = 2$) distribution

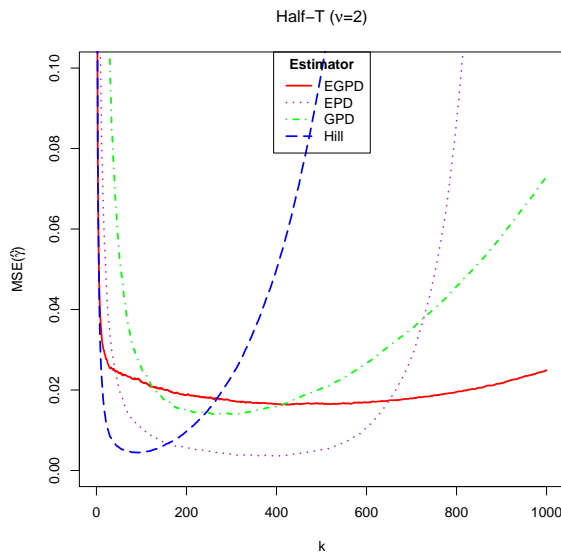
Figure 3.1: Bias($\hat{\gamma}_k$) for distributions in the Fréchet domain, i.e. Fréchet($\alpha = 2$), Burr XII($\lambda = 1, \tau = 2, \eta = 1$) and half-T($\nu = 2$).

Figure 3.2 below illustrates the MSE of the estimators of the EVI for distributions in the Fréchet domain. The EVI is taken to be $1/2$.



(a) Fréchet($\alpha = 2$) distribution

(b) Burr XII($\lambda = 1, \tau = 2, \eta = 1$) distribution

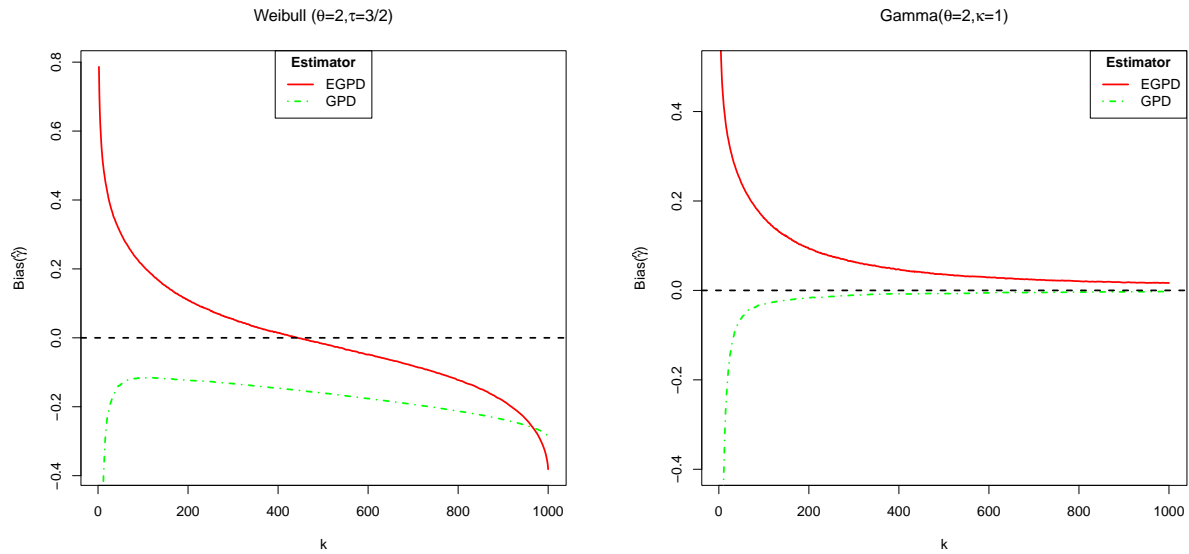


(c) half-T($\nu = 2$) distribution

Figure 3.2: $MSE(\hat{\gamma}_k)$ for distributions in the Fréchet domain, i.e. Fréchet($\alpha = 2$), Burr XII($\lambda = 1, \tau = 2, \eta = 1$) and half-T($\nu = 2$).

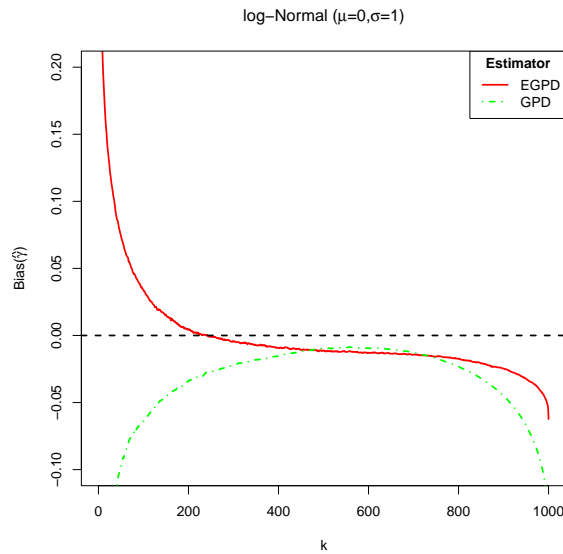
3.2.2.2 Distributions in the Gumbel domain

Figure 3.3 below illustrates the bias of the estimators of the EVI for distributions in the Gumbel domain. The EVI is zero (0) in this domain, by definition.



(a) Weibull($\theta = 1, \tau = 3/2$) distribution

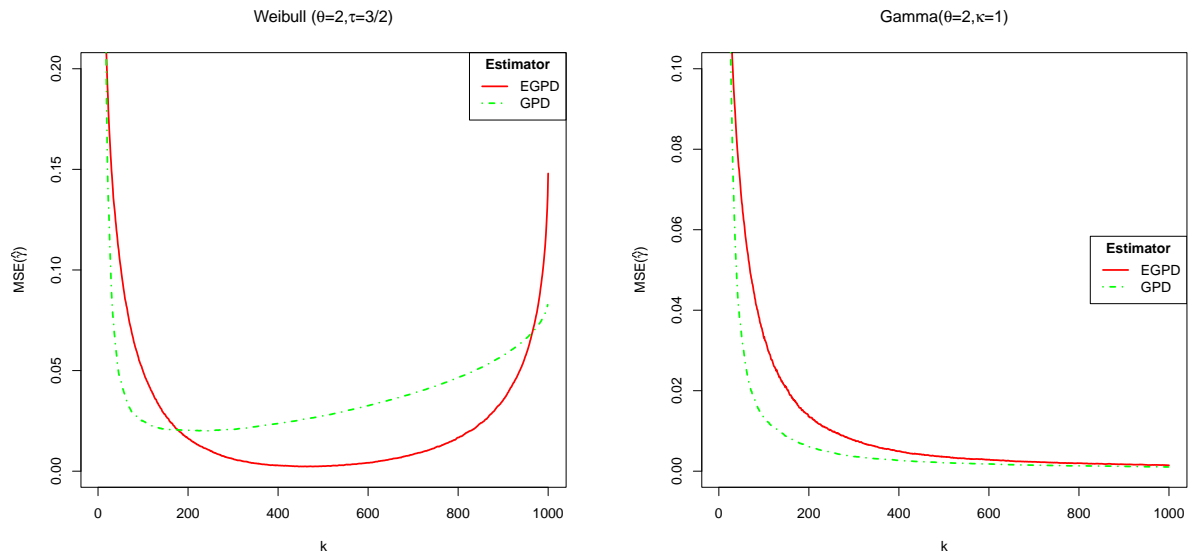
(b) Gamma($\theta = 2, \kappa = 3/2$) distribution



(c) log-Normal($\mu = 0, \sigma^2 = 1$) distribution

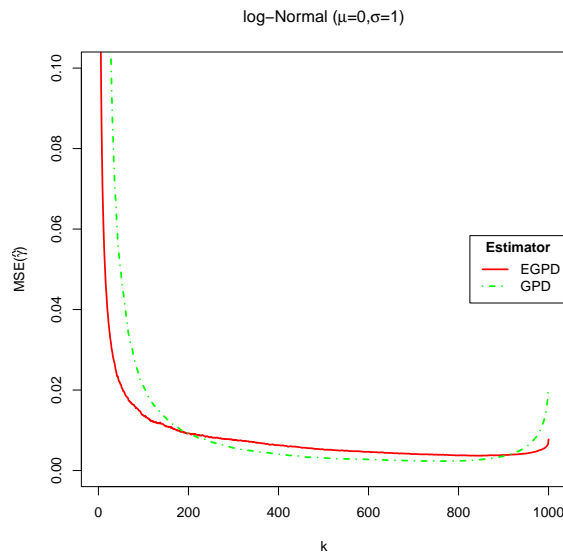
Figure 3.3: Bias($\hat{\gamma}_k$) for distributions in the Gumbel domain, i.e. Weibull($\theta = 1, \tau = 3/2$), Gamma($\theta = 2, \kappa = 3/2$) and log-Normal($\mu = 0, \sigma^2 = 1$).

Figure 3.4 below illustrates the MSE of the estimators of the EVI for distributions in the Gumbel domain. The EVI in this domain is 0.



(a) Weibull($\theta = 1, \tau = 3/2$) distribution

(b) Gamma($\theta = 2, \kappa = 3/2$) distribution

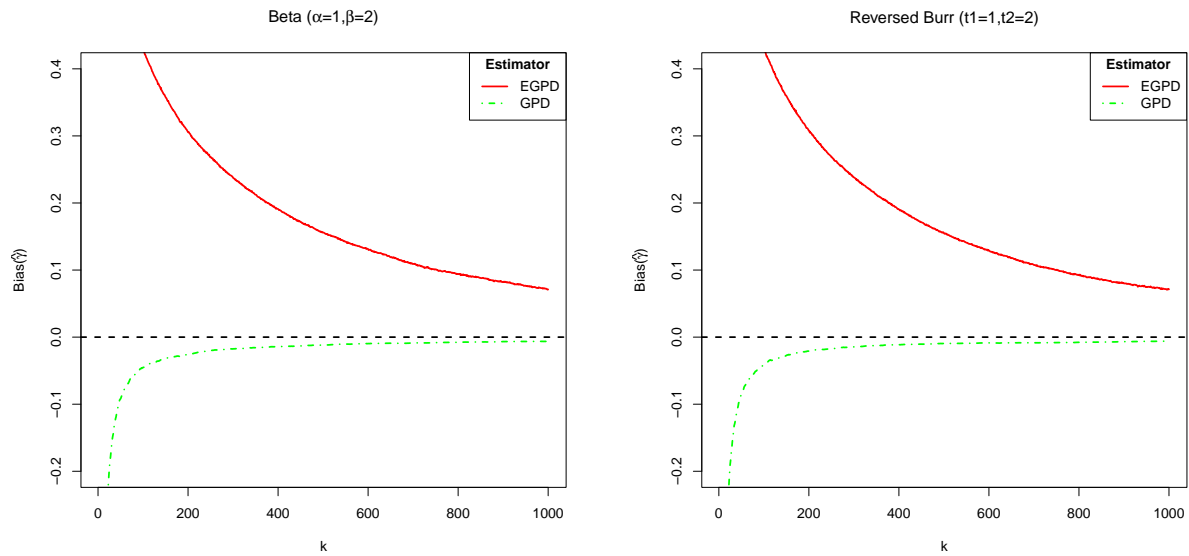


(c) log-Normal($\mu = 0, \sigma^2 = 1$) distribution

Figure 3.4: $MSE(\hat{\gamma}_k)$ for distributions in the Gumbel domain, i.e. Weibull($\theta = 1, \tau = 3/2$), Gamma($\theta = 2, \kappa = 3/2$) and log-Normal($\mu = 0, \sigma^2 = 1$).

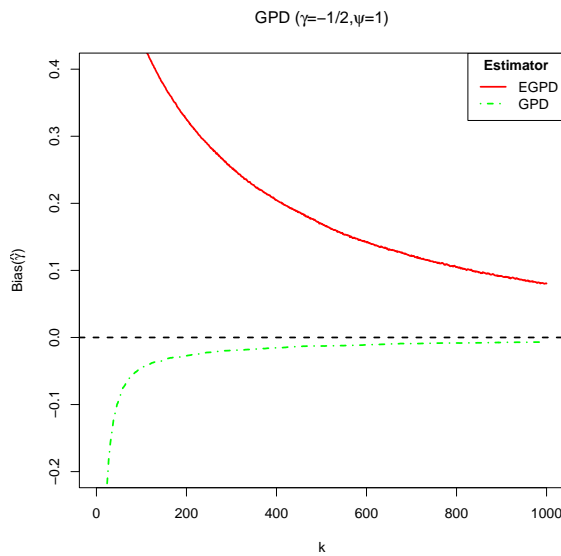
3.2.2.3 Distributions in the Weibull domain

Figure 3.5 below illustrates the bias of the estimators of the EVI for distributions in the Weibull domain. The EVI is taken to be $-1/2$.



(a) Beta($\alpha = 1, \beta = 2$) distribution

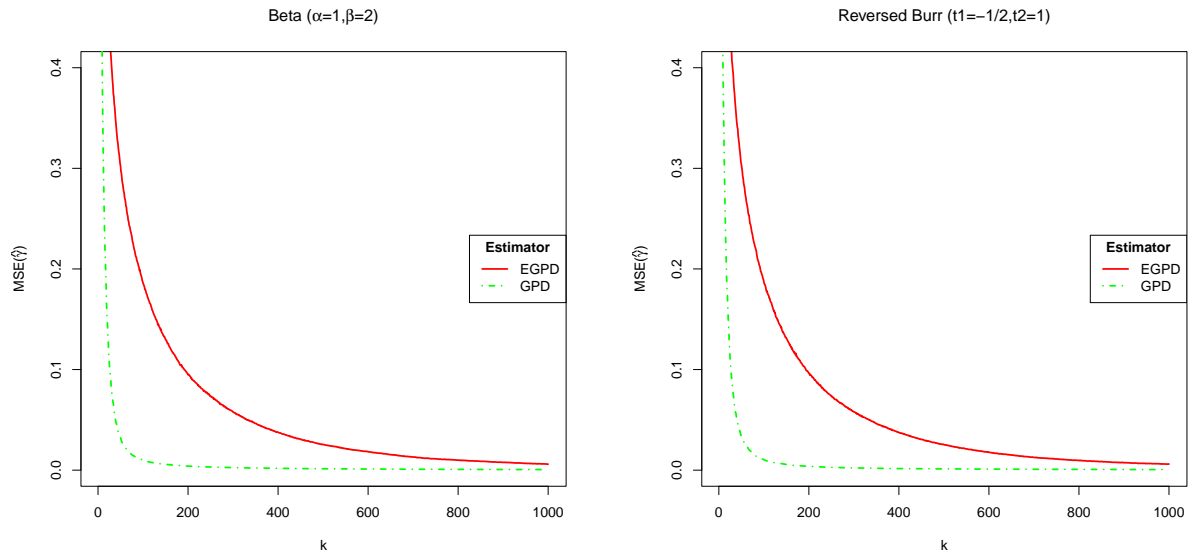
(b) Reversed Burr($t_1 = 1, t_2 = 2$) distribution



(c) GPD($\gamma = -1/2, \psi_t = 1$) distribution

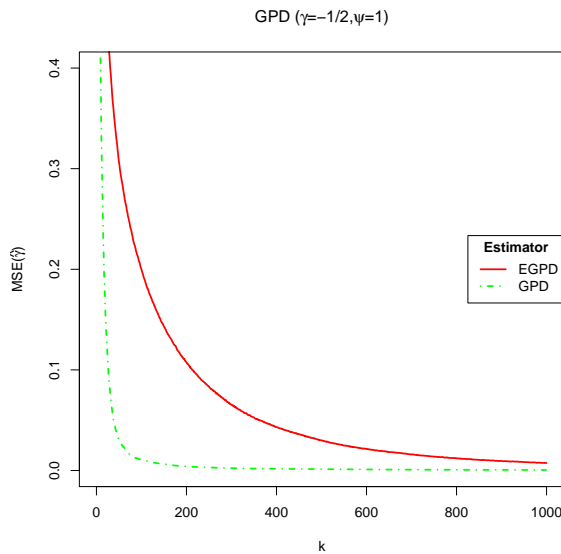
Figure 3.5: Bias($\hat{\gamma}_k$) for distributions in the Weibull domain, i.e. Beta($\alpha = 1, \beta = 2$), Reversed Burr($t_1 = 1, t_2 = 2$) and GPD($\gamma = -1/2, \psi_t = 1$).

Figure 3.6 below illustrates the MSE of the estimators of the EVI for distributions in the Weibull domain. The EVI is taken to be $-1/2$.



(a) Beta($\alpha = 1, \beta = 2$) distribution

(b) Reversed Burr($t_1 = 1, t_2 = 2$) distribution



(c) GPD($\gamma = -1/2, \psi_t = 1$) distribution

Figure 3.6: $MSE(\hat{\gamma}_k)$ for distributions in the Weibull domain, i.e. Beta($\alpha = 1, \beta = 2$), Reversed Burr($t_1 = 1, t_2 = 2$) and GPD($\gamma = -1/2, \psi_t = 1$).

3.2.3 Discussion of results

The EGPD model generally yields more stable estimates of the EVI when compared to the Hill, EPD and GPD in the Fréchet domain. From Figure 3.1, the Hill estimator has the least bias for very small values of k (i.e. the number of threshold exceedances), however, this bias is inflated as k increases. This estimator is thus only suitable for use when modelling very few threshold exceedances. The EPD model recorded the lowest bias for values of $k \in (100, 400)$, that is when modelling 10% – 40% of the data. This model thus allows for lower thresholds than the Hill estimator. The EGPD model has the lowest bias as k increases, thus allowing for modelling larger portions of the data. The bias of this model does however show a sharp increase when modelling more than about 99% of the data. Furthermore, although the EGPD had a lower bias for large k , an increase in bias with an increase in k was observed in the case of the Burr XII and half-T distributions. The GPD consistently performed poorer than all the other models. The MSE of the EGPD model in Figure 3.1 is more stable and lowest for large k when the data is generated from the Burr XII and half-T distributions. The EPD model reports the lowest MSE for moderate values of k , while the Hill has the lowest MSE for small values of k .

Consider next the results pertaining to the Gumbel domain. Figure 3.3 shows that the EGPD reduces bias in the estimation of the EVI over the GPD in the case of the Weibull and log-Normal distributions. Thus, a larger portion of the data can be modelled via the EGPD while maintaining a low bias. In the case of the Gamma distribution, however, the GPD reported the lowest bias across all values of k . Figure 3.4 shows the EGPD consistently reporting the lower MSE compared to the GPD for the Weibull and log-Normal distributions. In the case of the Gamma distribution, however, the GPD reports lower MSE for small values of k , but the MSEs of the two models are equivalent when more than half of the data is used to estimate the EVI. Both Figures 3.3 and 3.4 once again demonstrate the increase in bias and variance of the EGPD model when more than 99% of the data is used to fit the model.

Figure 3.5 shows that in the Weibull domain, the EGPD model has decreasing bias as the proportion of data modelled increases. This plot also shows however that the GPD consistently has a lower bias than the EGPD. Figure 3.6 shows that the GPD also has lower MSE than the EGPD across all considered distributions in the Weibull domain.

The EGPD model can thus be fitted to a larger portion of the data than the other models it was compared to, and this will reduce bias in estimating the EVI in the Fréchet and Gumbel domains. This simulation study has, however, not demonstrated any merit to using this model over existing models for data coming from distributions in the Weibull domain.

3.3 Case study

A comparative study is conducted to investigate the suitability of the EGPD as both a bulk model and a tail model. The data for this analysis is a record of 228 rainfall amounts (in mm) recorded at 92 weather stations in France for the Fall (Autumn) season. Particular focus is placed on Chartes weather station (48.46 Lat, 1.5 Lon), following the direction taken by [Tencaliec et al. \(2019\)](#) where the data is obtained.

The EGPD is compared to the classic 3-parameter GPD tail model with: the threshold (t), shape (γ) and scale (ψ_t) parameters. The parameterised tail fraction approach to the Gamma-GPD mixture model proposed by [Behrens et al. \(2004\)](#) is also considered as an alternative full distribution model. Recall from (2.3), the CDF of this model is:

$$P(Y \leq y|t, \theta, \kappa, \gamma, \psi_t, \phi_t) = \begin{cases} (1 - \phi_t) \cdot \frac{F(y|\theta, \kappa)}{F(t|\theta, \kappa)} & \text{for } y \leq t \\ (1 - \phi_t) + \phi_t \cdot H_{\gamma, \psi_t}(y) & \text{for } y > t \end{cases}$$

where F is the CDF of a Gamma distribution with shape and scale parameters given by κ and θ , respectively. The tail fraction, $\phi_t = P(Y > t)$, is estimated as the sample proportion of observations above the threshold, t . The choice of parametric bulk is motivated by the fit of the Gamma distribution to the positive rainfall amounts. The Gamma Q-Q plot in Figure 3.7 below demonstrates that the Gamma distribution fits the bulk of the data well, but fits rather poorly in the upper tail.

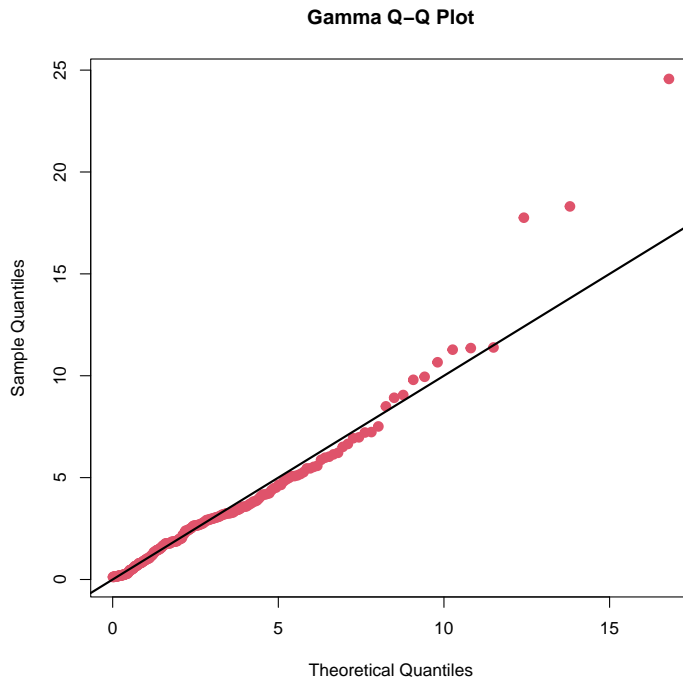


Figure 3.7: Gamma Q-Q plot for positive rainfall (mm) at Chartes

There are 228 rainfall amounts (mm) recorded at Chartes weather station, however, the modelling is done on the 222 positive rainfall amounts. Furthermore, a small jitter is added to the data to counter the excessive rounding thereof. The three models are fitted as follows:

1. GPD tail model is fitted via ML estimation using the `evmix` package in R. The threshold is taken to be the 85% quantile (i.e. $t = 4.9687$), and 34 threshold exceedances are used to fit the model.
2. The Gamma-GPD mixture model is also fitted using via ML estimation using the `evmix` package.
3. The EGPD is fitted using Algorithm 1.

Table 3.2 below gives the estimated GPD parameters under the various models. The Gamma scale and shape parameters in the Gamma-GPD mixture model are estimated to be $\hat{\theta} = 1.0992$ and $\hat{\kappa} = 2.6019$, respectively. The threshold for this model is $\hat{t} = 5.4567$, and the tail fraction is $\hat{\phi}_t = 0.1171$.

Shape (γ)	Scale (ψ_t)
<i>GPD tail model</i>	
0.3504	2.2770
<i>Gamma-GPD mixture model</i>	
0.2016	3.0870
<i>EGPD model with $m = 5$</i>	
0.1022 (0.05, 0.2446)	2.4448 (2.1577, 2.9417)

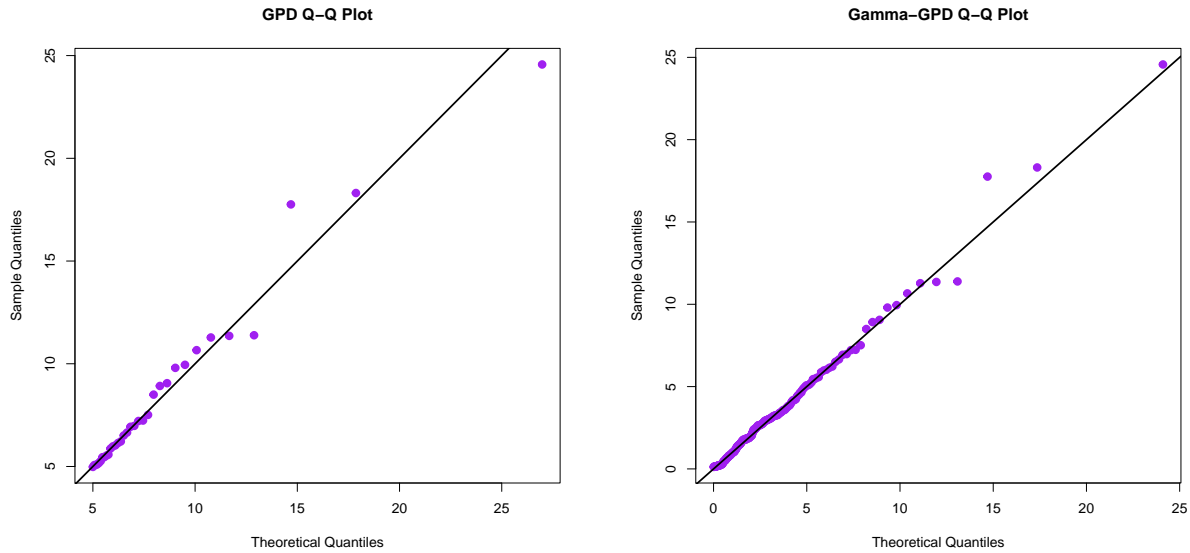
Note: the 95% confidence intervals for the parameters under the EGPD model are based on a non-parametric bootstrap of 500 samples.

Table 3.2: Estimated shape and scale parameters of the GPD, under the GPD tail fit, Gamma-GPD mixture model and the EGPD.

It is evident from Table 3.2 above that all three models have consistently estimated the tail index (shape parameter) as being positive, indicating a heavy-tailed underlying process. The EGPD model has reported the lowest estimate for the upper tail shape parameter, but it does however have a wide 95% confidence interval for this parameter. The GPD tail model has estimated a very high value of the shape parameter, compared to the other two models. The Gamma-GPD model has overestimated the scale parameter ψ_t compared to the other models.

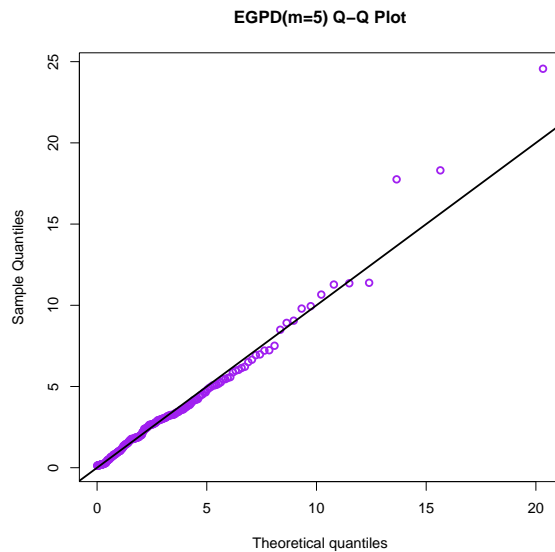
Figure 3.8 below gives Q-Q plots for the fitted models. From this, it is evident that all three models have been an overall good fit for the data. The EGPD has captured the upper tail as well as the GPD tail model has. Furthermore, this model has also captured the bulk of the

data well. This has also been observed in the Gamma-GPD mixture model. The strength of the EGPD over this model is that it is a simpler model with only 3 parameters versus the 6 parameters of the Gamma-GPD mixture.



(a) GPD tail model

(b) Gamma-GPD mixture model



(c) EGPD model

Figure 3.8: Q-Q plots for positive rainfall (mm) at Chartes, under the GPD tail, Gamma-GPD mixture and EGPD models.

Figure 3.9 below gives the fitted densities of the three models overlaid on the data, which is visibly heavy-tailed. Model fit results are similar to the conclusions reached in the Q-Q plots in Figure 3.8. The Gamma-GPD mixture model is discontinuous at the threshold. A continuity constraint can be imposed on this model but has been found in the literature to have a negligible effect on inference.

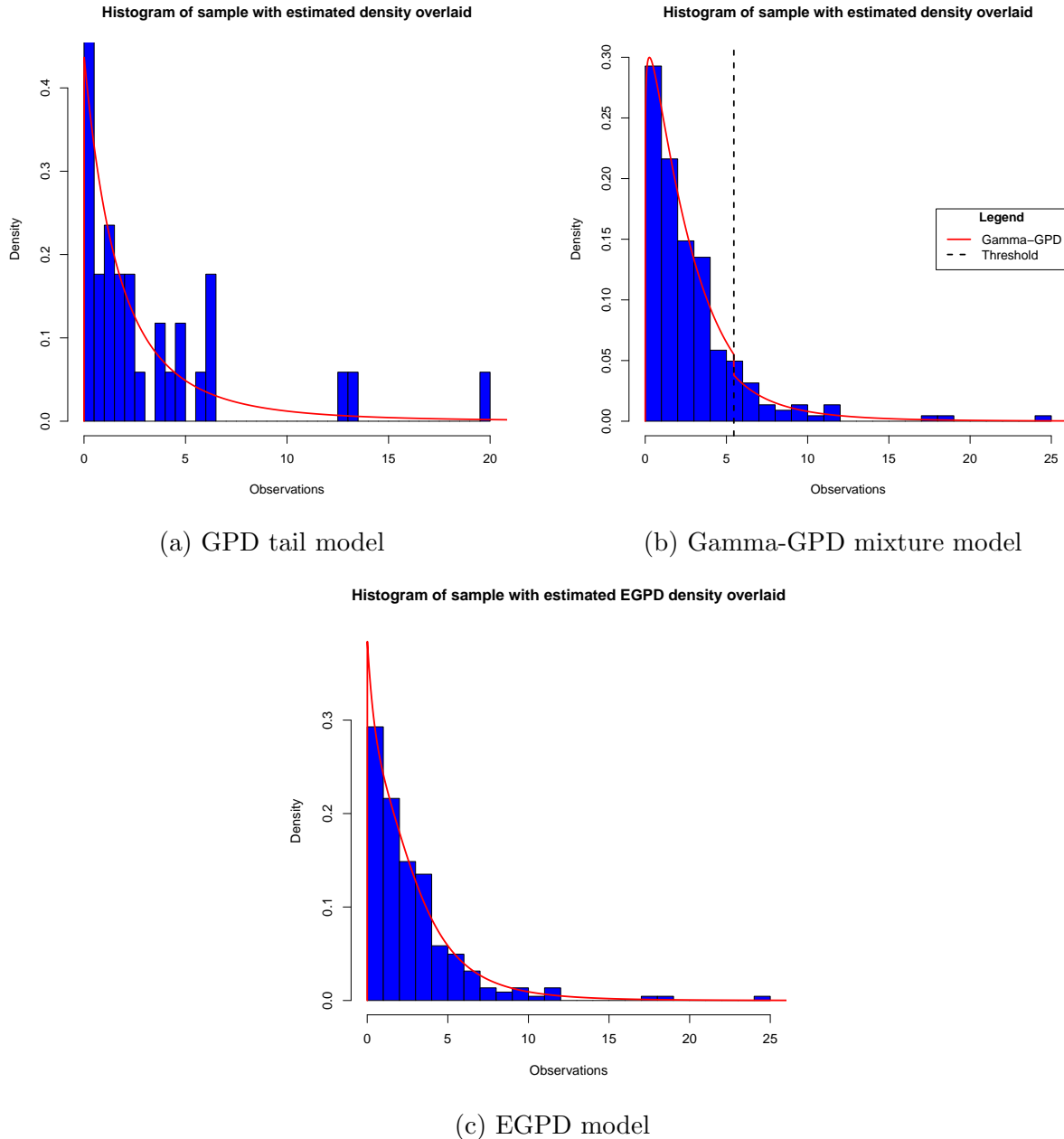


Figure 3.9: Histogram of positive rainfall (mm) at Chartes with Fitted GPD, Gamma-GPD and EGPD ($m = 5$) densities overlaid.

Figure 3.10 below gives the estimated return levels of rainfall at Chartes weather station. The blue dotted-lines represent the 90%, 95% and 99% quantiles of the data, respectively (from bottom to top).

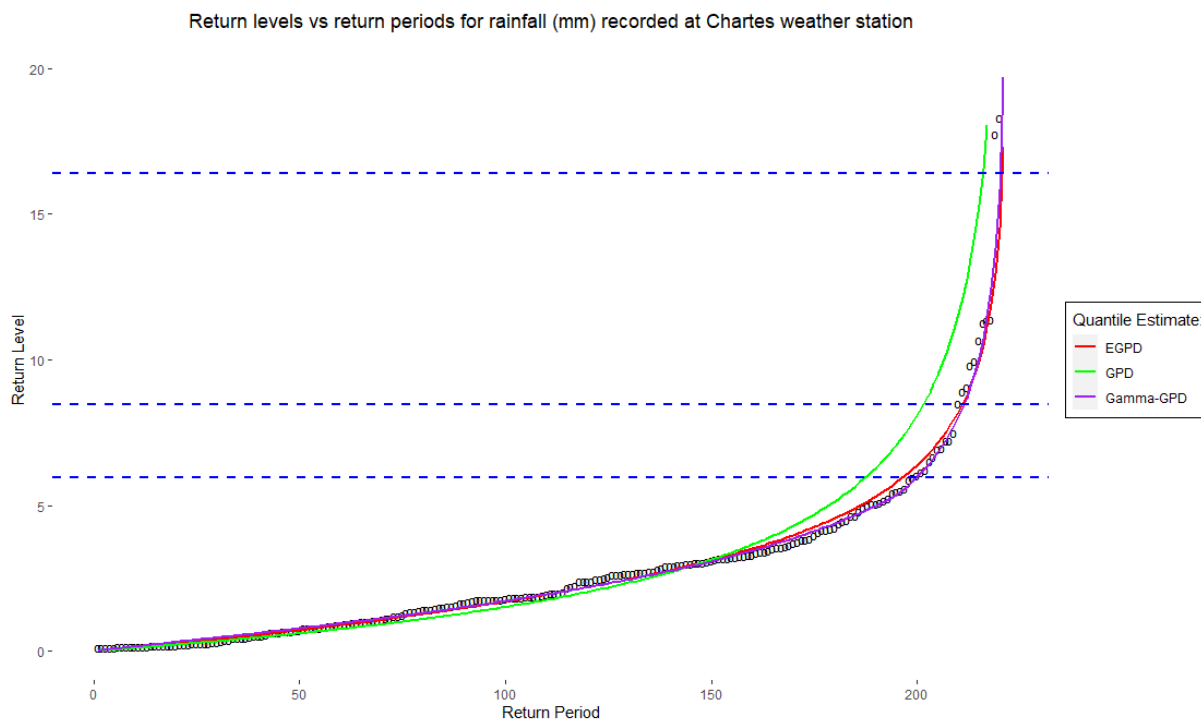


Figure 3.10: Estimates of return levels for rainfall (mm) recorded at Chartes weather station.

From Figure 3.10 above, it is evident that the Gamma-GPD and EGPD models perform well in estimating quantiles for the rainfall (mm) at Chartes. The estimates yielded by these models are similar in the lower tail and again beyond the 95% quantile. The EGPD model, however, estimates the quantiles well only up to the 99% quantile. Also evident from the plot is that the GPD tail model consistently overestimates extreme quantiles above the chosen threshold ($t = 4.9687$).

The above figure thus validates the suitability of the EGPD model as both bulk and tail model. Furthermore, this model has the advantage of being a parsimonious model, as it only used 3 parameters compared to the 6 parameters of the Gamma-GPD model.

3.4 Conclusion

This chapter discussed the Extended Generalised Pareto Distribution (EGPD) model developed to model the full range of data characterised by the presence of extreme values. Section 3.1

detailed the semi-parametric EGPD model based on Bernstein polynomials. The model-fitting procedure as well as a data-driven approach to estimating the degree of the Bernstein polynomial were also described in detail.

A comprehensive simulation study was carried out in section 3.2. This study demonstrated that, when compared to other models, the EGPD model can be more efficiently fitted to larger portions of data arising from distributions in the Fréchet domains. In the Gumbel domain, the performance of the EGPD varied depending on the underlying distribution. This model was not the best-performing model in the Weibull domain.

Section 3.3 was dedicated to a detailed case study which showed that the EGPD can model the tail of a distribution as well as the classic GPD tail model does. Furthermore, this model can also model the bulk of the distribution well, while bypassing threshold selection and being simpler and parsimonious compared to parametric mixture models.

The rest of this mini-dissertation will study this model only in the Fréchet domain, where it performed well.

Chapter 4

Semi-parametric Extended Pareto Distribution

Heavy-tailed distributions are defined as those distributions which have heavier tails than the exponential distribution (Nair et al., 2022). Let Y be a random variable with CDF F . Mathematically, Y is said to have a heavy-tailed distribution if:

$$\lim_{y \rightarrow \infty} \frac{1 - F(y)}{e^{-\lambda y}} = \lim_{y \rightarrow \infty} \frac{\bar{F}(y)}{e^{-\lambda y}} = \infty \text{ for } \lambda > 0$$

In essence, it is clear that $e^{-\lambda y} \rightarrow 0$ as $y \rightarrow \infty$; while this is also true of the tail (\bar{F}) of a heavy-tailed distribution, it happens much slower than for the exponential distribution.

Heavy-tailed distributions, in the context of EVT, are in the Fréchet domain of attraction. The simulation study that was carried out in section 3.2 of Chapter 3 demonstrated the ability of the EGPD to efficiently estimate the EVI for larger portions of data than other considered models in the Fréchet domain. This study, however, also showed a sharp increase in the bias of the EGPD model when working with more than 99% of the data. The cause of this is that the GPD tail model sometimes estimates negative values of the EVI. One possible solution for this would be to restrict the EVI in the model so that $\gamma \geq 0$ as in Tencaliec et al. (2019). The objective of this chapter is to consider another adjustment to the EGPD specifically for heavy-tailed data with the goal of reducing bias in the estimation of the EVI.

4.1 Pareto substitution

The Pareto distribution is the limiting distribution for the relative excesses (Y/t) of data arising from distributions in the Fréchet domain ($\gamma > 0$), as previously mentioned in Chapter 2. Furthermore, it is documented in the literature that the Pareto distribution and its extensions

are better POT models than the GPD (Beirlant et al., 2009, 2022). We thus propose substituting the Pareto distribution as the tail model in the EGPD, in place of the GPD. This new semi-parametric extended Pareto distribution is characterised by its CDF:

$$F_{m,N}(z) = \widehat{G}_{m,N} \{H_{\gamma,z_m}(z)\} \text{ for } z > z_m > 1 \quad (4.1)$$

where $\widehat{G}_{m,N}$ is a CDF corresponding to $\widehat{g}_{m,N}$ as defined in (3.3), H_{γ,z_m} is the CDF of a $Pareto(z_m, 1/\gamma)$ distribution and z_m is the minimum value that the random variable can take. The constraints on $\widehat{G}_{m,N}$ that need to be in place to ensure that $F_{m,N}$ is part of the EGPD family defined in Definition 3.1 were given in Chapter 3. This was done assuming a GPD tail model. We can, however, show that the Pareto distribution is a special case of the non-standard GPD as follows:

$$\begin{aligned} H_{\gamma,\psi_t,\mu}(z) &= 1 - \left[1 + \gamma \left(\frac{z - \mu}{\psi_t} \right) \right]^{-1/\gamma} \\ &= 1 - \left[1 + \gamma \left(\frac{z}{\psi_t} - \frac{\psi_t}{\gamma\psi_t} \right) \right]^{-1/\gamma} \quad \text{by taking } \mu = \psi_t/\gamma \\ &= 1 - \left[1 + z \frac{\gamma}{\psi_t} - \frac{\gamma}{\gamma} \right]^{-1/\gamma} \\ &= 1 - \left[1 + \frac{z}{z_m} - 1 \right]^{-1/\gamma} \quad \text{by defining } z_m := \psi_t/\gamma \\ &= 1 - \left(\frac{z}{z_m} \right)^{-1/\gamma} \\ &= H_{\gamma,z_m} \end{aligned}$$

Therefore, if the last weight is non-null (i.e. $\omega_{m,m} > 0$), then we have:

(i) $\lim_{z \rightarrow 0} \frac{F_{m,N}(z)}{z^s} = (\gamma z_m)^{-s} \binom{m-1}{s-1} \omega_{s,m} > 0$. This follows from the above derivation ($\psi_t = \gamma z_m$) and as proven in Tencaliec et al. (2019).

(ii) $\lim_{z \rightarrow \infty} \frac{\bar{F}_{m,N}(z)}{\bar{H}_{\gamma,z_m}(z)} = m\omega_{m,m} > 0$. See Result B.3 for proof.

(iii) Let Y be a non-negative continuous random variable satisfying $P(Y/t > z | Y > t) \xrightarrow[t \rightarrow \infty]{} (y/y_m)^{-1/\tilde{\gamma}}$. If Y is a semi-parametric EPD variable, i.e. $Y = H_{\gamma,y_m}^{-1} \left\{ \widehat{G}_{m,N}^{-1}(U) \right\}$ with $U \sim \mathcal{U}(0, 1)$, and we have that $\lim_{u \rightarrow 0} \frac{\widehat{\bar{G}}_{m,n}(1-u)}{u} > 0$, then $\tilde{\gamma} = \gamma$. See Result B.4 for proof.

These three conditions above confirm the semi-parametric EPD in (4.1) is a member of the EGPD family. Furthermore, the last of these conditions, (iii), ensures that the upper tail

behaviour of this model is governed by γ , as would be the case when using a genuine EVT argument.

With the theoretical background in place, the semi-parametric EPD can now be used for inference. Algorithm 2 below details how the semi-parametric EPD model is fitted, given data (\mathbf{y}) and the degree of the Bernstein polynomial (m). This algorithm is adapted from (Tencaliec et al., 2019) and adjusted accordingly for the Pareto distribution.

Algorithm 2: Estimation of the semi-parametric EPD model, given m

1. Obtain an initial value for the EVI, for $\gamma^{(0)}$ by fitting the Pareto distribution to \mathbf{y} .
2. At the i^{th} iteration:
 - 2.1. Calculate the weights, $\omega^{(i)}$, as follows:
 - 2.1.1. Generate the pseudo-observations: $\hat{U} = H_{\gamma^{(i-1)}, y_m^{(i-1)}}(\mathbf{y})$.
 - 2.1.2. Calculate G_N , the ECDF based on \hat{U} .
 - 2.1.3. Calculate $\omega_{k,m} = G_N(k/m) - G_N((k-1)/m)$.
 - 2.1.4. Ensure the last weight is non-null. If $\omega_{m,m} = 0$, then:
 - 2.1.4.1. Calculate: $\omega_{m,m} = 1 - \hat{G}_{m,N}(1 - 1/m)$.
 - 2.1.4.2. Normalise the weights: $\omega = \omega / \sum \omega$.
 - 2.2. Generate the pseudo-observations: $\hat{Z} = H_{\gamma^{(i-1)}, y_m^{(i-1)}}(\hat{U})$.
 - 2.3. Calculate $\hat{V} = \hat{G}_{m,N}(\hat{Z})$ using the fitted weights $\omega^{(i)}$.
 - 2.4. Generate PD variables, $\hat{X} = H_{\gamma^{(i-1)}, y_m^{(i-1)}}^{-1}(\hat{V})$, and compute the ML estimate, as derived in Result B.5:

$$\hat{\gamma}^{(i)} = \frac{1}{N} \sum_{i=1}^N \log \frac{\hat{X}_i}{\min \{ \hat{X}_1, \hat{X}_2, \dots, \hat{X}_N \}}$$

3. Check for convergence: if $|\hat{\gamma}^{(i)} - \hat{\gamma}^{(i-1)}| < 10^{-3}$ the algorithm has converged, else repeat (2) above.
-

Similar to the EGPD, the non-parametric bootstrap approach can be taken to obtain a $100(1 - \alpha)\%$ confidence interval as well as the standard error of the EVI, γ .

The value of m in Algorithm 2 above is obtained using LSCV as follows:

$$LSCV(m) = \int_0^\infty \hat{f}_{m,N}^2(y) dy - \frac{2}{N} \sum_{j=1}^N \hat{f}_{m,N}^{(-j)}(Y_j)$$

where $\hat{f}_{m,N}(y) = \hat{g}_{m,N} \{H_{\gamma,y_m}(y)\} \cdot h_{\gamma,y_m}(y)$. Define $u = H_{\gamma,y_m}(y) = 1 - \left(\frac{y}{y_m}\right)^{-1/\gamma}$, then we can rewrite

$$\hat{f}_{m,N}(y) = \frac{1}{\gamma y_m} \hat{g}_{m,N}(u) \cdot (1-u)^{1+\gamma}$$

Following in the same direction as in section 3.1.1, we can express $LSCV(m)$ as:

$$LSCV(m) = \frac{1}{\gamma y_m} \left\{ \int_0^1 \hat{g}_{m,N}^2(u) \cdot (1-u)^{1+\gamma} du - \frac{2}{N} \sum_{j=1}^N \hat{g}_{m,N}^{(-j)}(Z_j) \cdot (1-Z_j)^{1+\gamma} \right\}$$

where $Z_j = H_{\gamma,y_m}(Y_j)$. Applying the formula derived by [Leblanc \(2010\)](#) leads to $\gamma y_m LSCV(m)$ being written as:

$$\frac{m^2}{2m + \gamma} \mathbf{G}^T \mathbf{A} \mathbf{G} - \frac{2}{N-1} \left\{ \sum_{j=1}^N \hat{g}_{m,N}(Z_j) \cdot (1-Z_j)^{1+\gamma} - \frac{1}{N} \sum_{j=1}^N \beta_{k_j+1, m-k_j}(Z_j) \cdot (1-Z_j)^{1+\gamma} \right\}$$

with k_j , \mathbf{G} and \mathbf{A} as defined in section 3.1.1. In order to obtain the appropriate value of m , a sequence of values of m is chosen and $LSCV(m)$ is calculated for each value in the sequence. The appropriate m is then chosen as the value with the corresponding least LSCV.

The rest of this chapter will evaluate the ability of the semi-parametric EPD to estimate the EVI through a simulation study. A case study will also be conducted to demonstrate the practical usefulness of this model.

4.2 Simulation study

A simulation study is carried out to investigate the performance of the semi-parametric Extended Pareto Distribution (from here on "EGPD (Pareto)") model in estimating the extreme value index (EVI) of data arising from distributions in the Fréchet domain of attraction. Model performance is measured in terms of the bias and MSE of the estimator.

4.2.1 Design

The simulation study is designed as follows:

- 1000 samples of size 1000 each are taken from 4 distributions in the Fréchet domain of attraction.
- For each sample, the EVI ($\hat{\gamma}_k$) is estimated for the threshold exceedances:

$$Y_{N-k+j:N} - Y_{N-k:N}, \quad j = 1, 2, \dots, k$$

for each $k = 2, 3, \dots, N$. Estimation methods are detailed below.

- The performance for each estimator considered is measured by Bias and MSE where for each k , $Bias(\hat{\gamma}_k)$ and $MSE(\hat{\gamma}_k)$ is calculated as the average across the 1000 samples.

The following distributions are used in the study: Pareto($\alpha = 4$), Pareto($\alpha = 2$), Fréchet($\alpha = 4$) and Fréchet($\alpha = 2$). This parameterisation will show how the estimators behave under different tail-heaviness. The EGPD (Pareto) is compared to the original EGPD model in the case of Pareto data. EVI estimators for heavy-tailed distributions are expected to perform very well in the case of exact Pareto data, and thus this will be a good evaluation of the suitability of the EGPD (Pareto) model (McNeil, 1999). For the Fréchet distributions, the EGPD (Pareto) model is compared to the EPD, GPD and Hill estimators.

4.2.2 Results

The results of the simulation study are given below. Plot of Bias and MSE are given for each of the considered distributions. The results are discussed in subsection 4.2.3

4.2.2.1 Pareto distributions

Figure 4.1 below illustrates the bias and MSE of the estimators of the EVI for the Pareto($\alpha = 4$) and Pareto($\alpha = 2$) distributions. The corresponding EVIs are $1/4$ and $1/2$, respectively. The EGPD (Pareto) model is given the alias "S-EPD" in the plots.

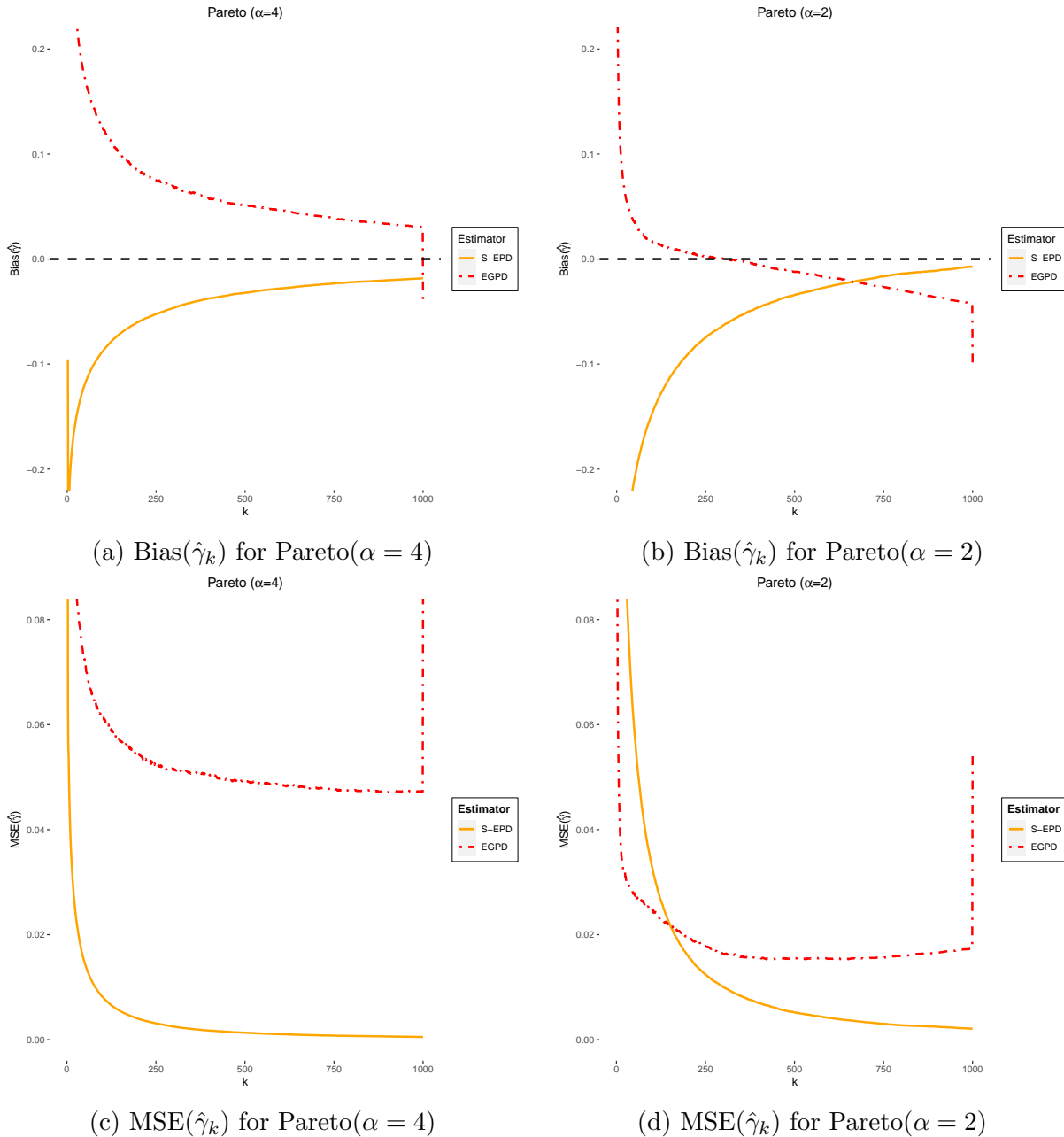


Figure 4.1: Bias($\hat{\gamma}_k$) and MSE($\hat{\gamma}_k$) for Pareto($\alpha = 4$) and Pareto($\alpha = 2$) distributions.

4.2.2.2 Fréchet distributions

Figure 4.2 below illustrates the bias and MSE of the estimators of the EVI for the Fréchet($\alpha = 4$) and Fréchet($\alpha = 2$) distributions. The corresponding EVIs are $1/4$ and $1/2$, respectively. The EGD (Pareto) model is given the alias "S-EPD" in the plots.

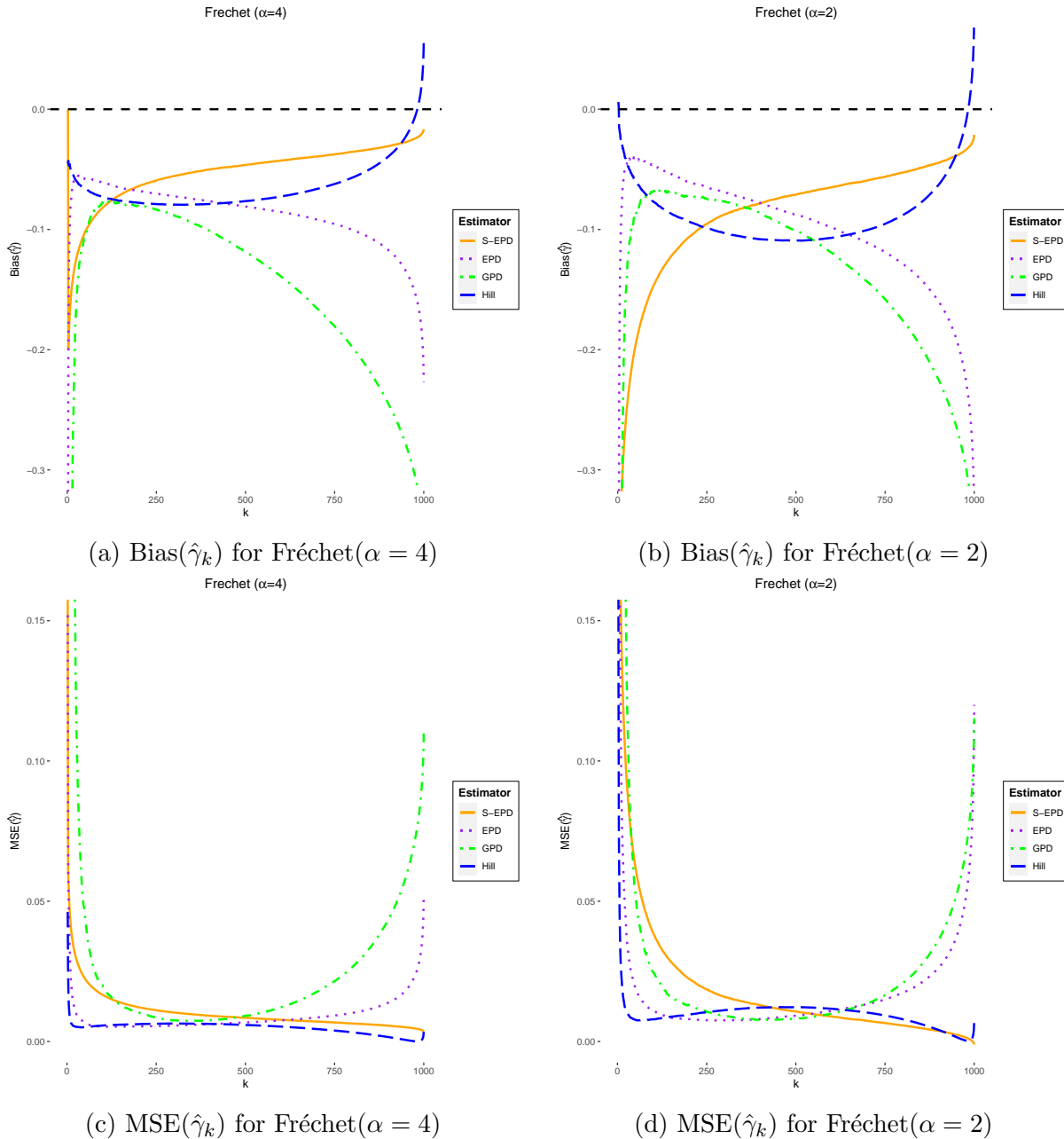


Figure 4.2: Bias($\hat{\gamma}_k$) and MSE($\hat{\gamma}_k$) for Fréchet($\alpha = 4$) and Fréchet($\alpha = 2$) distributions.

4.2.3 Discussion of results

The results in Figure 4.1 are discussed first. The EGPD (Pareto) shows more stable estimates of the EVI across k . Furthermore, this model reduces bias in the estimation of the EVI as k increases, that is, as larger portions of the data are modelled. Evidently, the Pareto substitution also reduces the MSE of the EGPD model. Moreover, there is a sharp increase (in absolute value) in the bias of the EGPD model when more than 99% of the data, however, the EGPD (Pareto) does not suffer from this.

In the case of Fréchet distribution in Figure 4.2, the Hill estimator has the least bias for small k . The EPD allows for larger portions of data to be modelled than the GPD and Hill estimators. The EGPD (Pareto) model, however, shows the most stable estimates of the EVI and reduces bias in the estimation more than the other models. All models save for the GPD, show higher bias and MSE as tail-heaviness increases.

The EGPD (Pareto) model can thus be fitted to a larger portion of the data than the other models it was compared to, and this will reduce bias in estimating the EVI in the Fréchet domain. This result is consistent regardless of the tail-heaviness of the underlying distribution. This study has also shown that the Pareto substitution reduces the variance in the original EGPD model.

4.3 Case study

A case study is carried out on the Secura Belgian Re data of 371 automobile insurance claims in excess of €1,200,000. These claims were filed over the period from 1988 to 2001, and have been adjusted for inflation. This data is publicly available in the `ReIns` package in R, and was previously analysed in [Beirlant et al. \(2004\)](#) and [Beirlant et al. \(2009\)](#). The primary objective of the study done on this data is to price an excess-of-loss reinsurance contract with a retention limit (operational priority) of R . The re-insurer in this type of contract pays out only on claims exceeding the priority level, and the contract is thus priced in accordance with this limit.

The net premium for an excess-of-loss reinsurance contract with operational priority R is calculated as

$$\Pi(R) := \mathbb{E}[(Y - R)_+] = e(R) \cdot \bar{F}(R)$$

where $e(\cdot)$ is the mean excess function defined in [Chapter 2](#). In this study, $\Pi(R)$ is first calculated using a simple non-parametric approach which uses the following data-driven formula for the net premium:

$$\Pi(R) = \hat{e}_N(R) \cdot \bar{F}_N(R)$$

where $\hat{e}_N(\cdot)$ is the sample mean excess function defined in [Chapter 2](#), and F_N is the ECDF.

We also consider the EGPD (Pareto), EPD and Hill estimators. [Beirlant et al. \(2004\)](#) showed that the net premium, for Pareto-type models, can be calculated as follows:

$$\Pi(R) = \frac{R}{\frac{1}{\hat{\gamma}} - 1} \cdot \frac{k}{N} \cdot \left(\frac{R}{Y_{N-k:N}} \right)^{-\frac{1}{\hat{\gamma}}}$$

where k is the number of observations over the threshold, $Y_{N-k:N}$. The above formula is used to calculate $\Pi(R)$ for the Hill and EPD estimators. The net premium for the EGPD (Pareto) model is estimated as:

$$\Pi(R) = \frac{R}{\frac{1}{\hat{\gamma}} - 1} \bar{F}_{m,N}(R)$$

The calculation of the net premium for the Pareto-type models depends on the estimated EVI. [Figure 4.3](#) below shows the trajectories of the EVI estimates for the models considered above, as well as the GPD for consistency.

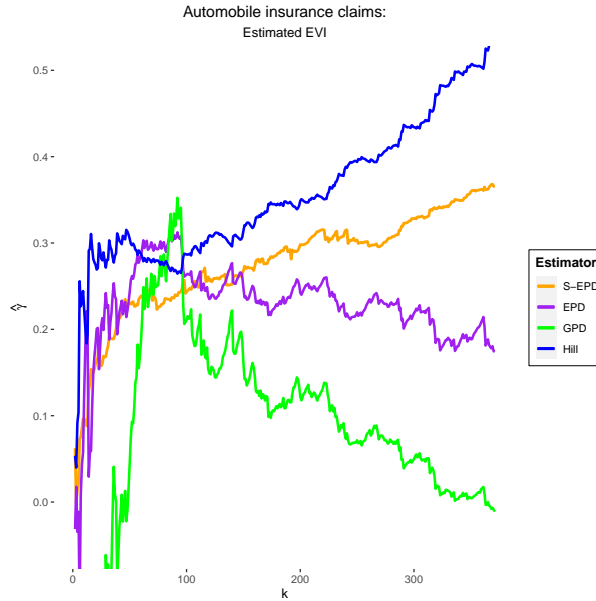


Figure 4.3: Estimates of the EVI for the Secura Belgian Re data.

The EPD and Hill estimators show the most stability around $k = 95$ with an estimated EVI of around 0.3. The EGPD model shows relative stability in the EVI estimates for small changes in k .

Table 4.1 below gives the estimated net premium for different priority levels, using the non-parametric estimator, the Hill and EPD estimators with $k = 95$ and $Y_{N-k:N} = \text{€}2,580,026$, as well as the EGPD (Pareto) model. Notice that the non-parametric estimator cannot extrapolate beyond the observed range of data, because it relies on the empirical CDF.

R	Non-parametric	Hill	EGPD (Pareto)	EPD
3 000 000	161 728.11	163 793.14	165 966.71	191 948.09
3 500 000	108 837.24	108 214.64	112 747.48	132 641.17
4 000 000	74 696.34	75 570.89	81 138.10	96 303.53
4 500 000	53 312.31	55 057.08	61 404.52	72 611.21
5 000 000	35 888.04	41 474.22	48 403.56	56 402.68
7 500 000	1 074.50	13 941.06	22 882.08	21 336.32
10 000 000	–	6 432.13	17 487.11	10 704.79

Table 4.1: Estimated net premium, in €, calculated for varying priority levels, R for the Secura Belgian Re data.

The non-parametric estimator yields conservative premiums, while the EPD yields the highest premiums. However, for $R \geq \text{€}7,500,000$ the EGPD (Pareto) model yields the highest estimate, because of its heavier right tail. Contrary to this, the non-parametric estimator calculates very low premiums for large R , owing to the small number of observations exceeding these values. The results in Table 4.1 thus show the strength of extreme value models in extrapolation.

The estimators considered above are also used to calculate exceedance probabilities, $p = P(Y > R)$ for large R . Table 4.2 below gives the calculated values of p when $R = \text{€}7,000,000$.

Estimator	$P(Y > R)$
Non-parametric	0.0081
Hill	0.0065
EPD	0.0075
EGPD (Pareto)	0.0094 (0.0062, 0.0154)

Note: the 95% confidence interval for the EGPD model is based on a non-parametric bootstrap of 500 samples.

Table 4.2: Estimated exceedance probabilities, $p = P(Y > R)$ for $R = \text{€}7,000,000$, for the Secura Belgian Re data.

The EGPD (Pareto) model allows for more observations in the right tail than the other estimators, as can be seen in the estimated exceedance probability and corresponding 95% confidence interval in Table 4.2.

4.4 Conclusion

This chapter introduced the Pareto substitution as the tail model in the EGPD, thus defining the EGPD (Pareto) model. Motivation for the Pareto substitution and an algorithm to fit the model to data was also given.

A simulation study of the EGPD (Pareto) model's performance in EVI estimation was conducted in section 4.2. This study showed that the Pareto substitution not only reduces bias in the estimation of the EVI, but it also corrected for the erroneous negative estimates of the EVI that were calculated by the EGPD model in the case of heavy-tailed data.

Section 4.3 was a case study on the Secura Belgian Re data. The EGPD model was compared to the EPD, Hill and non-parametric estimator in estimating the EVI of the data, calculating a

net premium and estimating an exceedance probability. The EGPD model had relatively stable estimates of the EVI and proved to extrapolate better than the other models considered.

Chapter 5

Conclusion

This research aimed to investigate extreme value mixture models, in particular, the Extended Generalised Pareto Distribution (EGPD) and its extension for Fréchet-type data.

[Chapter 2](#) provided an overview of EVT, detailing tail modelling and the threshold selection problem in POT modelling. Extreme value mixture models were also discussed in this chapter.

The EGPD was defined in [Chapter 3](#), in both its parametric and non-parametric forms. An extensive Monte Carlo simulation study was conducted to investigate this model's effectiveness as an estimator of the Extreme Value Index (EVI). This study showed that the EGPD allowed for larger portions of data to be modelled than its contemporaries, while maintaining a low bias, for data arising from distribution in the Fréchet and Gumbel domains. The EGPD was not found to make an improvement over existing EVI estimators in the Weibull domain of attraction. A case study done on rainfall data demonstrated the ability of the EGPD to model the bulk and tail of a distribution while bypassing threshold selection, and do so well.

[Chapter 4](#) introduced the Pareto distribution tail model substitution in the EGPD for the case of heavy-tailed data. This model was shown to be a valid member of the EGPD family and its compliance with EVT was illustrated. A comprehensive simulation study was carried out, which showed a tenuous reduction in the variance of the EVI estimator when using the Pareto distribution as opposed to the GPD as the tail model in the EGPD. The Pareto substitution also corrected for the inflated bias of the EVI estimator when modelling more than 99% of the data. A case study was also considered to illustrate the practical usefulness of the EGPD (Pareto) model in calculating the net premium and exceedance probabilities in excess-of-loss reinsurance contracts.

5.1 Future work

Future research work on this topic could investigate:

- adjusting the EGPD model for the case of incomplete data. This includes when data is truncated or randomly right-censored.
- Bayesian methods of parameter estimation, as these methods have been shown in the literature to reduce both bias and variance of the EVI estimators for the GPD and its extensions.

Bibliography

- [1] R Aviv. An extreme-value theory approximation scheme in reinsurance and insurance-linked securities. *ASTIN Bulletin: The Journal of the IAA*, 48(3):1157–1173, 2018.
- [2] R Baker. Lebesgue measure on $r^{\wedge\{\infty\}}$. *Proceedings of the American Mathematical Society*, 113(4):1023–1029, 1991.
- [3] AA Balkema and L De Haan. Residual life time at great age. *The Annals of Probability*, 2(5):792–804, 1974.
- [4] CN Behrens, HF Lopes, and D Gamerman. Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling*, 4(3):227–244, 2004.
- [5] J Beirlant, Y Goegebeur, J Segers, and JL Teugels. *Statistics of extremes: Theory and applications*, volume 558. John Wiley & Sons, 2004.
- [6] J Beirlant, E Joossens, and J Segers. Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference*, 139(8):2800–2815, 2009.
- [7] J Beirlant, G Maribe, P Naveau, and A Verster. Bias reduced peaks over threshold tail estimation. *REVSTAT-Statistical Journal*, 20(3):277–304, 2022.
- [8] S Bernstein. Démonstration du théorème de weierstrass fondée sur le calcul des probabilités. *Communications of the Kharkov Mathematical Society*, 13(1):1–2, 1912.
- [9] NH Bingham, CM Goldie, and JL Teugels. *Regular variation*. Cambridge University Press, 1989.
- [10] LV Bortkiewicz. Das helmertsche verteilungsgesetz für die quadratsumme zufälliger beobachtungsfehler. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 2(5):358–375, 1922.
- [11] A Charpentier and E Flachaire. Pareto models for risk management. *Recent econometric techniques for macroeconomic and financial data*, pages 355–387, 2021.
- [12] G Chryssolouris, V Subramanian, and M Lee. Use of extreme value theory in engineering decision making. *Journal of manufacturing Systems*, 13(4):302–312, 1994.
- [13] L de Haan and A Ferreira. *Extreme value theory: An introduction*, volume 3. Springer, 2006.
- [14] EL Dodd. The greatest and the least variate under general laws of error. *Transactions of the American Mathematical Society*, 25(4):525–539, 1923.
- [15] P Embrechts, C Klüppelberg, and T Mikosch. *Modelling extremal events: For insurance and finance*, volume 33. Springer Science & Business Media, 1997.

- [16] RA Fisher and LHC 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.
- [17] M Fréchet. Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Math. Polon.*, 6: 93–116, 1927.
- [18] A Frigessi, O Haug, and H Rue. A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, 5:219–235, 2002.
- [19] B Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, pages 423–453, 1943.
- [20] BM Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, pages 1163–1174, 1975.
- [21] Y Hu. *Extreme value mixture modelling with simulation study and applications in finance and insurance*. PhD thesis, University of Canterbury, 2013.
- [22] AF 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.
- [23] ML Juncosa. The asymptotic behavior of the minimum in a sequence of random variables. *Duke Mathematical Journal*, 16(4):609–618, 1949.
- [24] MR Leadbetter. On a basis for 'peaks over threshold' modeling. *Statistics & Probability Letters*, 12(4):357–362, 1991.
- [25] A Leblanc. A bias-reduced approach to density estimation using bernstein polynomials. *Journal of Nonparametric Statistics*, 22(4):459–475, 2010.
- [26] D Maposa, AM Seimela, C Sigauke, and JJ Cochran. Modelling temperature extremes in the limpopo province: Bivariate time-varying threshold excess approach. *Natural Hazards*, 107(3):2227–2246, 2021.
- [27] AJ McNeil. Extreme value theory for risk managers. *Departement Mathematik ETH Zentrum*, 12(5):217–37, 1999.
- [28] AJ McNeil, R Frey, and P Embrechts. *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton University Press, 2015.
- [29] SH Mkhandi, RK Kachroo, and TAG Gunasekara. Flood frequency analysis of southern africa: Identification of regional distributions. *Hydrological Sciences Journal*, 45(3):449–464, 2000.
- [30] S Nadarajah and JT Shiau. Analysis of extreme flood events for the pachang river, taiwan. *Water Resources Management*, 19(4):363–374, 2005.

- [31] J Nair, A Wierman, and B Zwart. *The fundamentals of heavy tails: properties, emergence and estimation*. Cambridge University Press, 2022.
- [32] P Naveau, R Huser, P Ribereau, and A Hannart. Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, 52(4): 2753–2769, 2016.
- [33] I Papastathopoulos and JA Tawn. Extended generalised pareto models for tail estimation. *Journal of Statistical Planning and Inference*, 143(1):131–143, 2013.
- [34] J Pickands III. Statistical inference using extreme order statistics. *The Annals of Statistics*, pages 119–131, 1975.
- [35] BW Silverman. *Density Estimation for Statistics and Data Analysis*, volume 26. CRC press, 1986.
- [36] RL Smith. Extreme value theory. *Handbook of Applicable Mathematics*, 7:437–471, 1990.
- [37] H Tabari. Extreme value analysis dilemma for climate change impact assessment on global flood and extreme precipitation. *Journal of Hydrology*, 593:125932, 2021.
- [38] P Tencaliec, A-C Favre, P Naveau, C Prieur, and G Nicolet. Flexible semiparametric generalized pareto modeling of the entire range of rainfall amount. *Environmetrics*, 31(2):e2582–1, 2019.
- [39] SLG Vicente. *Extreme value theory: An application to sports*. PhD thesis, Universidade de Lisboa, 2012.
- [40] RA Vitale. A bernstein polynomial approach to density function estimation. In *Statistical Inference and Related Topics*, pages 87–99. Elsevier, 1975.
- [41] L von Mises. Neue beiträge zum problem der sozialistischen wirtschaftsrechnung. *Archive für Socialwissenschaft und Socialpolitik*, 51:488–500, 1923.
- [42] M Vyskocil and J Koudelka. Extreme value theory for operational risk in insurance: a case study. *Journal of Operational Risk*, 16(4), 2021.
- [43] L Zheng and T Sayed. Application of extreme value theory for before-after road safety analysis. *Transportation Research Record*, 2673(4):1001–1010, 2019.
- [44] L Zheng, K Ismail, and X Meng. Freeway safety estimation using extreme value theory approaches: A comparative study. *Accident Analysis & Prevention*, 62:32–41, 2014.

Appendices

Appendix A

Results stated without proof

Result A.1 (Probability Integral Transform)

Let Y be a continuous random variable with CDF F which is a continuous function. Define the random variable $X = F(Y)$. Then $X \sim \mathcal{U}(0, 1)$, i.e. X is uniformly distributed over the interval $[0, 1]$. ▲

The probability integral transform in [Result A.1](#) above is used in the result below in order to generate random variables from any continuous population with an invertible CDF.

Result A.2

Let Y be a continuous random variable with CDF F . Suppose F is continuous and one-to-one, such that F^{-1} exists. Define the random variable $Z = F^{-1}(U)$, with $U \sim \mathcal{U}(0, 1)$. Then Z has the same distribution as Y , i.e. $P(Z \leq z) = F(z)$. ▲

Appendix B

Derivations

Result B.1

Let G be a continuous CDF defined on the unit interval, with Bernstein polynomial approximation

$$B_m(G, u) = \sum_{i=0}^m G_N \left(\frac{i}{m} \right) \cdot b_{i,m}(u).$$

The approximation of the corresponding PDF is

$$\widehat{g}_{m,N}(u) = m \sum_{i=0}^{m-1} \left\{ G_N \left(\frac{i+1}{m} \right) - G_N \left(\frac{i}{m} \right) \right\} b_{i,m-1}(u)$$

▲

Proof

The PDF corresponding to G is obtained by differentiation as follows:

$$\begin{aligned} \widehat{g}_{m,N}(u) &= \frac{d}{du} B_m(G, u) \\ &= \frac{d}{du} \left\{ \sum_{i=0}^m G_N \left(\frac{i}{m} \right) \cdot b_{i,m}(u) \right\} \\ &= \frac{d}{du} \left\{ \sum_{i=0}^m G_N \left(\frac{i}{m} \right) \binom{m}{i} u^i (1-u)^{m-i} \right\} \\ &= \sum_{i=0}^m G_N \left(\frac{i}{m} \right) \binom{m}{i} [i u^{i-1} (1-u)^{m-i} - (m-i) u^i (1-u)^{m-i-1}] \end{aligned} \quad (\text{B.1})$$

The first term in the brackets in (B.1) is equal to 0 when $i = 0$. Therefore, the first term of $\widehat{g}_{m,N}(u)$ can be rewritten as:

$$\sum_{i=0}^m G_N \left(\frac{i}{m} \right) \binom{m}{i} [i u^{i-1} (1-u)^{m-i}]$$

APPENDIX B. DERIVATIONS

$$\begin{aligned}
 &= \sum_{i=0}^{m-1} G_N \left(\frac{i+1}{m} \right) \binom{m}{i+1} (i+1) u^{i+1-1} (1-u)^{m-i-1} \\
 &= \sum_{i=0}^{m-1} G_N \left(\frac{i+1}{m} \right) \frac{m!}{(i+1)!(m-i-1)!} (i+1) u^i (1-u)^{m-i-1} \\
 &= \sum_{i=0}^{m-1} G_N \left(\frac{i+1}{m} \right) \frac{m[(m-1)!]}{(i+1)(i)!(m-i-1)!} (i+1) u^i (1-u)^{m-i-1} \\
 &= m \sum_{i=0}^{m-1} G_N \left(\frac{i+1}{m} \right) \binom{m-1}{i} u^i (1-u)^{m-1-i} \\
 &= m \sum_{i=0}^{m-1} G_N \left(\frac{i+1}{m} \right) b_{i,m-1}(u)
 \end{aligned}$$

Similarly, the second term in the brackets in (B.1) is equal to 0 when $i = m$. Therefore, the second term of $\hat{g}_{m,N}(u)$ can be rewritten as:

$$\begin{aligned}
 &\sum_{i=0}^m G_N \left(\frac{i}{m} \right) \binom{m}{i} [(m-i) u^i (1-u)^{m-i-1}] \\
 &= \sum_{i=0}^{m-1} G_N \left(\frac{i}{m} \right) \binom{m}{i} (m-i) u^i (1-u)^{m-i-1} \\
 &= \sum_{i=0}^{m-1} G_N \left(\frac{i}{m} \right) \frac{m[(m-1)!]}{i!(m-i)[(m-i-1)!]} (m-i) u^i (1-u)^{m-i-1} \\
 &= m \sum_{i=0}^{m-1} G_N \left(\frac{i}{m} \right) \binom{m-1}{i} u^i (1-u)^{m-1-i} \\
 &= m \sum_{i=0}^{m-1} G_N \left(\frac{i}{m} \right) b_{i,m-1}(u)
 \end{aligned}$$

Putting these two terms together yields

$$\hat{g}_{m,N}(u) = m \sum_{i=0}^{m-1} \left\{ G_N \left(\frac{i+1}{m} \right) - G_N \left(\frac{i}{m} \right) \right\} b_{i,m-1}(u)$$

▲

APPENDIX B. DERIVATIONS

Result B.2

The Bernstein polynomial approximation of the PDF of the unknown distribution function G :

$$\widehat{g}_{m,N}(u) = m \sum_{i=0}^{m-1} \left\{ G_N \left(\frac{i+1}{m} \right) - G_N \left(\frac{i}{m} \right) \right\} b_{i,m-1}(u)$$

can be equivalently expressed as:

$$\widehat{g}_{m,N}(u) = \sum_{k=1}^m \omega_{k,m} \beta_{k,m-k+1}(u)$$

where $\omega_{k,m} = G_N(k/m) - G_N((k-1)/m)$ and $\beta_{k,m-k+1}$ is the density of the Beta distribution with parameters k and $m - k + 1$. ▲

Proof

The approximation of the density function is

$$\widehat{g}_{m,N}(u) = m \sum_{i=0}^{m-1} \left\{ G_N \left(\frac{i+1}{m} \right) - G_N \left(\frac{i}{m} \right) \right\} b_{i,m-1}(u)$$

In the above expression, put $k = i + 1$, which implies $i = k - 1$. Then, the PDF can be written as:

$$\begin{aligned} \widehat{g}_{m,N}(u) &= m \sum_{k=1}^m \left\{ G_N \left(\frac{k}{m} \right) - G_N \left(\frac{k-1}{m} \right) \right\} b_{k-1,m-1}(u) \\ &= m \sum_{k=1}^m \left\{ G_N \left(\frac{k}{m} \right) - G_N \left(\frac{k-1}{m} \right) \right\} \binom{m-1}{k-1} u^{k-1} (1-u)^{m-1-(k-1)} \\ &= m \sum_{k=1}^m \left\{ G_N \left(\frac{k}{m} \right) - G_N \left(\frac{k-1}{m} \right) \right\} \frac{(m-1)!}{(k-1)!(m-1-k+1)!} u^{k-1} (1-u)^{m-k} \\ &= \sum_{k=1}^m \left\{ G_N \left(\frac{k}{m} \right) - G_N \left(\frac{k-1}{m} \right) \right\} \frac{m(m-1)!}{(k-1)!(m-k)!} u^{k-1} (1-u)^{(m-k+1)-1} \\ &= \sum_{k=1}^m \left\{ G_N \left(\frac{k}{m} \right) - G_N \left(\frac{k-1}{m} \right) \right\} \frac{m!}{(k-1)!(m-k)!} u^{k-1} (1-u)^{(m-k+1)-1} \\ &= \sum_{k=1}^m \left\{ G_N \left(\frac{k}{m} \right) - G_N \left(\frac{k-1}{m} \right) \right\} \frac{\Gamma(m+1)}{\Gamma(k)\Gamma(m-k+1)} u^{k-1} (1-u)^{(m-k+1)-1} \\ &= \sum_{k=1}^m \left\{ G_N \left(\frac{k}{m} \right) - G_N \left(\frac{k-1}{m} \right) \right\} \frac{\Gamma\{(m-k+1)+k\}}{\Gamma(k)\Gamma(m-k+1)} u^{k-1} (1-u)^{(m-k+1)-1} \\ &= \sum_{k=1}^m \left\{ G_N \left(\frac{k}{m} \right) - G_N \left(\frac{k-1}{m} \right) \right\} \frac{1}{B(k, m-k+1)} u^{k-1} (1-u)^{(m-k+1)-1} \\ &= \sum_{k=1}^m \omega_{k,m} \beta_{k,m-k+1}(u) \end{aligned}$$

where $\omega_{k,m} = G_N(k/m) - G_N((k-1)/m)$ and $\beta_{k,m-k+1}$ is the density of the Beta distribution with parameters k and $m - k + 1$. ▲

APPENDIX B. DERIVATIONS

Result B.3

The upper tail of the semi-parametric extended Pareto distribution in (4.1) is equivalent to the tail of the Pareto distribution when $\omega_{m,m} > 0$, i.e.

$$\lim_{z \rightarrow \infty} \frac{\bar{F}_{m,N}(z)}{\bar{H}_{\gamma,z_m}(z)} = m\omega_{m,m} > 0$$

▲

Proof

$$\begin{aligned} \lim_{z \rightarrow \infty} \frac{\bar{F}_{m,N}(z)}{\bar{H}_{\gamma,z_m}(z)} &= \lim_{z \rightarrow \infty} \frac{\bar{\widehat{G}}_{m,N}\{H_{\gamma,z_m}(z)\}}{\bar{H}_{\gamma,z_m}(z)} \\ &= \lim_{u \rightarrow 0} \frac{\bar{\widehat{G}}_{m,N}(1-u)}{u} \text{ where } u = \bar{H}_{\gamma,z_m}(z) \\ &= \lim_{u \rightarrow 0} \frac{1 - \widehat{G}_{m,N}(1-u)}{u} \\ &= \lim_{u \rightarrow 0} \widehat{g}_{m,N}(1-u) \text{ by l'Hopital's rule} \\ &= \widehat{g}_{m,N}(1) \\ &= \sum_{k=1}^m \omega_{k,m} \beta_{k,m-k+1}(1) \text{ from (3.2)} \\ &= \sum_{k=1}^m \omega_{k,m} \times \frac{\Gamma(m+1)}{\Gamma(k)\Gamma(1)} \times 0^{m-k} \\ &= \omega_{m,m} \frac{\Gamma(m+1)}{\Gamma(m)} \\ &= \omega_{m,m} \frac{m\Gamma(m)}{\Gamma(m)} \\ &= m\omega_{m,m} > 0 \end{aligned}$$

▲

APPENDIX B. DERIVATIONS

Result B.4

Let Y be a non-negative continuous random variable satisfying $P(Y/t > z|Y > t) \xrightarrow{t \rightarrow \infty} (y/y_m)^{-1/\tilde{\gamma}}$.

If Y is a semi-parametric EPD variable, i.e. $Y = H_{\gamma, y_m}^{-1} \left\{ \widehat{G}_{m, N}^{-1}(U) \right\}$ with $U \sim \mathcal{U}(0, 1)$, and we have that $\lim_{u \rightarrow 0} \frac{\overline{\widehat{G}}_{m, n}(1-u)}{u} > 0$, then $\tilde{\gamma} = \gamma$. ▲

Proof

$$\begin{aligned}
 P(Y/t > z|Y > t) &= P(Y > zt|Y > t) \\
 &= \frac{P(Y > zt, Y > t)}{P(Y > t)} \\
 &= \frac{P\left(H_{\gamma, y_m}^{-1} \left\{ \widehat{G}_{m, N}^{-1}(U) \right\} > zt\right)}{P\left(H_{\gamma, y_m}^{-1} \left\{ \widehat{G}_{m, N}^{-1}(U) \right\} > t\right)} \\
 &= \frac{P\left(\widehat{G}_{m, N}^{-1}(U) > H_{\gamma, y_m}(zt)\right)}{P\left(\widehat{G}_{m, N}^{-1}(U) > H_{\gamma, y_m}(t)\right)} \\
 &= \frac{P\left(U > \widehat{G}_{m, N} \left\{ H_{\gamma, y_m}(zt) \right\}\right)}{P\left(U > \widehat{G}_{m, N} \left\{ H_{\gamma, y_m}(t) \right\}\right)} \\
 &= \frac{1 - \widehat{G}_{m, N} \left\{ H_{\gamma, y_m}(zt) \right\}}{1 - \widehat{G}_{m, N} \left\{ H_{\gamma, y_m}(t) \right\}} \\
 &= \frac{\overline{\widehat{G}}_{m, N} \left\{ H_{\gamma, y_m}(zt) \right\}}{\overline{\widehat{G}}_{m, N} \left\{ H_{\gamma, y_m}(t) \right\}} \\
 &= \frac{\overline{\widehat{G}}_{m, N}(1-u)}{\overline{\widehat{G}}_{m, N}(1-u^*)} \text{ with } u = \overline{H}_{\gamma, y_m}(zt) \text{ and } u^* = \overline{H}_{\gamma, y_m}(t) \\
 &= \frac{\overline{\widehat{G}}_{m, N}(1-u)}{u} \times \frac{u^*}{\overline{\widehat{G}}_{m, N}(1-u^*)} \times \frac{u}{u^*}
 \end{aligned}$$

As $t \rightarrow \infty$, we have that $u \rightarrow 0$ and $u^* \rightarrow 0$. Using the product limit law and l'Hopital's rule we get that as $t \rightarrow \infty$:

$$\begin{aligned}
 P(Y/t > z|Y > t) &\rightarrow \frac{\widehat{g}_{m, N}(1-u)}{\widehat{g}_{m, N}(1-u^*)} \times \frac{u}{u^*} \\
 &\rightarrow \frac{\widehat{g}_{m, N}(1)}{\widehat{g}_{m, N}(1)} \times \frac{u}{u^*} \\
 &= \frac{u}{u^*}
 \end{aligned}$$

APPENDIX B. DERIVATIONS

$$\begin{aligned} &= \frac{(zt/y_m)^{-1/\gamma}}{(t/y_m)^{-1/\gamma}} \\ &= \left(\frac{zt \cdot y_m}{t \cdot y_m} \right)^{-1/\gamma} \\ &= z^{-1/\gamma} = \bar{H}_{\gamma,1}(z) \end{aligned}$$

This implies that for large t , $\tilde{\gamma} \rightarrow \gamma$.

▲

APPENDIX B. DERIVATIONS

Result B.5

Let Y be a random variable with a Pareto distribution with parameters $\alpha = 1/\gamma$ and y_m , i.e.

$$h_{\gamma, y_m}(y|\alpha, y_m) = \frac{\alpha y_m^\alpha}{y^{\alpha+1}} \text{ for } y > y_m > 0.$$

Let $\mathbf{y} = (y_1, y_2, \dots, y_N)^T$ be a sample from this population. The ML estimate for γ is

$$\hat{\gamma} = \frac{1}{N} \sum_{i=1}^N \log \frac{y_i}{y_m}$$

.

▲

Proof

The log-likelihood function for γ and y_m is derived as follows:

$$\begin{aligned} \mathcal{L}(\gamma, y_m|\mathbf{y}) &= \prod_{i=1}^N h_{\gamma, y_m}(y_i|\gamma, y_m) \\ &= \prod_{i=1}^N \frac{1}{\gamma} y_m^{1/\gamma} y_i^{-\frac{1+\gamma}{\gamma}} \\ \therefore l(\gamma|\mathbf{y}) = \log \mathcal{L}(\gamma|\mathbf{y}) &= \sum_{i=1}^N \left\{ \log \gamma^{-1} + \frac{1}{\gamma} \log y_m - \frac{1+\gamma}{\gamma} \log y_i \right\} \\ &= -N \log \gamma + \frac{N}{\gamma} \log y_m - \frac{1+\gamma}{\gamma} \sum_{i=1}^N \log y_i \end{aligned}$$

Differentiating this function with respect to γ yields:

$$\frac{\partial}{\partial \gamma} l(\gamma|\mathbf{y}) = -\frac{N}{\gamma} - \frac{1}{\gamma^2} + \frac{N}{\gamma^2} \sum_{i=1}^N \log y_i = -\frac{N}{\gamma} + \frac{1}{\gamma^2} \sum_{i=1}^N \log \frac{y_i}{y_m}$$

The ML estimate for γ is obtained by setting the partial derivative to zero:

$$\begin{aligned} \frac{N}{\gamma} &= \frac{1}{\gamma^2} \sum_{i=1}^N \log \frac{y_i}{y_m} \\ N\gamma &= \sum_{i=1}^N \log \frac{y_i}{y_m} \\ \therefore \hat{\gamma} &= \frac{1}{N} \sum_{i=1}^N \log \frac{y_i}{y_m} \end{aligned}$$

▲