Improved estimation procedures for a positive extreme value index

Thomas Louw Berning

Dissertation presented for the degree of

Doctor of Philosophy at

Stellenbosch University

Promoter: Prof. Tertius de Wet

December 2010

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the authorship owner thereof (unless to the extent explicitly otherwise stated) and I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature: …………………………………………………….. Date: 22 November 2010

Abstract

In extreme value theory (EVT) the emphasis is on extreme (very small or very large) observations. The crucial parameter when making inferences about extreme quantiles, is called the extreme value index (EVI). This thesis concentrates on only the right tail of the underlying distribution (extremely large observations), and specifically situations where the EVI is assumed to be positive. A positive EVI indicates that the underlying distribution of the data has a heavy right tail, as is the case with, for example, insurance claims data.

There are numerous areas of application of EVT, since there are a vast number of situations in which one would be interested in predicting extreme events accurately. Accurate prediction requires accurate estimation of the EVI, which has received ample attention in the literature from a theoretical as well as practical point of view.

Countless estimators of the EVI exist in the literature, but the practitioner has little information on how these estimators compare. An extensive simulation study was designed and conducted to compare the performance of a wide range of estimators, over a wide range of sample sizes and distributions.

A new procedure for the estimation of a positive EVI was developed, based on fitting the perturbed Pareto distribution (PPD) to observations above a threshold, using Bayesian methodology. Attention was also given to the development of a threshold selection technique.

One of the major contributions of this thesis is a measure which quantifies the stability (or rather instability) of estimates across a range of thresholds. This measure can be used to objectively obtain the range of thresholds over which the estimates are most stable. It is this measure which is used for the purpose of threshold selection for the proposed PPD estimator.

A case study of five insurance claims data sets illustrates how data sets can be analyzed in practice. It is shown to what extent discretion can/should be applied, as well as how different estimators can be used in a complementary fashion to give more insight into the nature of the data and the extreme tail of the underlying distribution. The analysis is carried out from the point of raw data, to the construction of tables which can be used directly to gauge the risk of the insurance portfolio over a given time frame.

Opsomming

Die veld van ekstreemwaardeteorie (EVT) is bemoeid met ekstreme (baie klein of baie groot) waarnemings. Die parameter wat deurslaggewend is wanneer inferensies aangaande ekstreme kwantiele ter sprake is, is die sogenaamde ekstreemwaarde-indeks (EVI). Hierdie verhandeling konsentreer op slegs die regterstert van die onderliggende verdeling (baie groot waarnemings), en meer spesifiek, op situasies waar aanvaar word dat die EVI positief is. ʼn Positiewe EVI dui aan dat die onderliggende verdeling ʼn swaar regterstert het, wat byvoorbeeld die geval is by versekeringseis data.

Daar is verskeie velde waar EVT toegepas word, aangesien daar ʼn groot aantal situasies is waarin mens sou belangstel om ekstreme gebeurtenisse akkuraat te voorspel. Akkurate voorspelling vereis die akkurate beraming van die EVI, wat reeds ruim aandag in die literatuur geniet het, uit beide teoretiese en praktiese oogpunte.

ʼn Groot aantal beramers van die EVI bestaan in die literatuur, maar enige persoon wat die toepassing van EVT in die praktyk beoog, het min inligting oor hoe hierdie beramers met mekaar vergelyk. ʼn Uitgebreide simulasiestudie is ontwerp en uitgevoer om die akkuraatheid van beraming van ʼn groot verskeidenheid van beramers in die literatuur te vergelyk. Die studie sluit ʼn groot verskeidenheid van steekproefgroottes en onderliggende verdelings in.

ʼn Nuwe prosedure vir die beraming van ʼn positiewe EVI is ontwikkel, gebaseer op die passing van die gesteurde Pareto verdeling (PPD) aan waarnemings wat ʼn gegewe drempel oorskrei, deur van Bayes tegnieke gebruik te maak. Aandag is ook geskenk aan die ontwikkeling van ʼn drempelseleksiemetode.

Een van die hoofbydraes van hierdie verhandeling is ʼn maatstaf wat die stabiliteit (of eerder onstabiliteit) van beramings oor verskeie drempels kwantifiseer. Hierdie maatstaf bied ʼn objektiewe manier om ʼn gebied (versameling van drempelwaardes) te verkry waaroor die beramings die stabielste is. Dit is hierdie maatstaf wat gebruik word om drempelseleksie te doen in die geval van die PPD beramer.

ʼn Gevallestudie van vyf stelle data van versekeringseise demonstreer hoe data in die praktyk geanaliseer kan word. Daar word getoon tot watter mate diskresie toegepas kan/moet word, asook hoe verskillende beramers op ʼn komplementêre wyse ingespan kan word om meer insig te verkry met betrekking tot die aard van die data en die stert van die onderliggende verdeling. Die analise word uitgevoer vanaf die punt waar slegs rou data beskikbaar is, tot op die punt waar tabelle saamgestel is wat direk gebruik kan word om die risiko van die versekeringsportefeulje te bepaal oor ʼn gegewe periode.

To my high school science teacher, Mr. Ben du Plessis

Acknowledgements

I would like to express my sincere gratitude towards the following persons:

- Our heavenly Father, for his undeserved blessing.
- Prof. Tertius de Wet, for his professional guidance. For patiently supporting and encouraging me when the going was tough. For long hours spent with me on research problems, teaching me how research is conducted, and painstakingly checking and refining my work. For much appreciated friendship.
- Ivona Contardo, for tremendous support, especially during the final stages of the work. For hours of proofreading the thesis, and for invaluable suggestions.
- My parents, Gert and Soma, for teaching me the value of discipline in my work, as well as in my personal life. For instilling in me from childhood a desire for knowledge and wisdom.
- My friend, George Harris, for always helping me to refocus on the bigger picture of my life.

Table of Contents

Abbreviations

AMSE	asymptotic mean square error
ERM	exponential regression model
EVI	extreme value index
EVT	extreme value theory
GEV	generalized extreme value
GPD	generalized Pareto distribution
hpd	highest posterior density
HSM	half-sample mode
MCMC	Markov Chain Monte Carlo
MDI	maximal data information
MLE	maximum likelihood estimator/estimate/estimation (depending on the context)
MSE	mean square error
PPD	perturbed Pareto distribution
RMSE	root mean square error

Notation

$F(x)$	If X is a continuous random variable with density function $f(x)$, $-\infty < x < \infty$, then the distribution function of X is denoted by $F(x)$, where $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du.$
$\bar{F}(x)$	The tail function or survival function of a continuous random variable X is denoted by $\bar{F}(x)$, where $\bar{F}(x) = 1 - F(x)$.
$Q(p)$	The inverse of the distribution function is called the quantile function and is denoted by $Q(p)$, where $Q(p) = \inf\{x \mid F(x) \geq p\}$.
$H_{k,n}$	The Hill estimator (or estimate) based on k excesses from an original sample of n observations.
$o(g(x))$	When considering the limit of $f(x)$ as $x \to c$, we write $f(x) = o(g(x))$ if and only if $f(x)/g(x) \to 0$ as $x \to c$.
$o(1)$	When considering the limit of $f(x)$ as $x \to c$, $f(x) = g(x)\big(1 + o(1)\big) \Rightarrow f(x)/g(x) \to 1$ as $x \to c$.
$x_{i,n}$	For a set of observations $x_1, x_2, \ldots, x_n$ the ordered observations are indicated as $x_{1,n} \leq x_{2,n} \leq \cdots \leq x_{n,n}$.
$\Phi(x)$	The distribution function of the standard normal distribution is denoted by $\Phi(x)$.
$\phi(x)$	The density function of the standard normal distribution is denoted by $\phi(x)$.
$X \overset{\mathcal{D}}{\approx} Y$	Random variables X and Y follow approximately the same distribution.

For notation regarding specific families of distributions, see *Distributions* below.

Distributions

Beta distribution

Notation: $X \sim Beta(a, b)$

Density function: $f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$

Distribution function: $F(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x w^{a-1}(1-w)^{b-1} dw$

Range: $0 \leq x \leq 1$

Parameters: $a > 0$; $b > 0$

EVI: $-\frac{1}{b}$

Burr distribution

Notation: $X \sim Burr(\beta, \tau, \lambda)$

Density function: $f(x) = \frac{\lambda\tau}{\beta}\left(\frac{\beta}{\beta+x^\tau}\right)^{\lambda+1} x^{\tau-1} = \frac{\lambda\tau}{\beta}\left(1+\frac{x^\tau}{\beta}\right)^{-\lambda-1} x^{\tau-1}$

Distribution function: $F(x) = 1 - \left(\frac{\beta}{\beta+x^\tau}\right)^{\lambda} = 1 - \left(1+\frac{x^\tau}{\beta}\right)^{-\lambda}$

Range: $x > 0$

Parameters: $\beta > 0$; $\tau > 0$; $\lambda > 0$

EVI: $\frac{1}{\lambda\tau}$

Exponential distribution

Notation: $X \sim exp(\lambda)$

Density function: $f(x) = \lambda e^{-\lambda x}$

Distribution function: $F(x) = 1 - e^{-\lambda x}$

Range: $x \geq 0$

Parameters: $\lambda > 0$

EVI: 0

F distribution

Notation: $X \sim F(m, n)$

Density function: $f(x) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} x^{m/2-1} \left(1 - \frac{m}{n} x\right)^{-(m+n)/2}$

Distribution function: $F(x) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} \int_0^x w^{m/2-1} \left(1 - \frac{m}{n} w\right)^{-(m+n)/2} dw$

Range: $x > 0$

Parameters: $m > 0; n > 0$

EVI: $\frac{2}{n}$

Fréchet distribution

Notation: $X \sim Fréchet(\alpha)$

Density function: $f(x) = \alpha x^{-\alpha-1} \exp(-x^{-\alpha})$

Distribution function: $F(x) = \exp(-x^{-\alpha})$

Range: $x > 0$

Parameters: $\alpha > 0$

EVI: $\frac{1}{\alpha}$

Gamma distribution

Notation: $X \sim \Gamma(\lambda, \alpha)$

Density function: $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$

Distribution function: $F(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^x w^{\alpha-1} e^{-\lambda w} dw$

Range: $x > 0$

Parameters: $\lambda > 0; \alpha > 0$

EVI: 0

Generalized extreme value (GEV) distribution $\gamma \neq 0$

Notation: $X \sim GEV(\gamma, \mu, \sigma)$

Density function: $f(x) = \frac{1}{\sigma}\left[1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\gamma}-1} \exp\left\{-\left[1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\gamma}}\right\}$

Distribution function: $F(x) = \exp\left\{-\left[1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\gamma}}\right\}$

Range: $\left\{x \middle| 1 + \gamma\left(\frac{x-\mu}{\sigma}\right) > 0\right\}$

Parameters: $\gamma \neq 0; -\infty < \mu < \infty; \sigma > 0$

EVI: γ

Generalized extreme value (GEV) distribution $\gamma = 0$

Notation: $X \sim GEV(0, \mu, \sigma)$

Density function: $f(x) = \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}\right) \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right)$

Distribution function: $F(x) = \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right)$

Range: $-\infty < x < \infty$

Parameters: $-\infty < \mu < \infty; \sigma > 0$

EVI: 0

Generalized Pareto distribution (GPD)

Notation: $X \sim GP(\gamma, \sigma)$

Density function: $f(x) = \frac{1}{\sigma}\left(1 + \frac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}-1}$

Distribution function: $F(x) = 1 - \left(1 + \frac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}}$

Range: $x \geq 0$

Parameters: $\gamma > 0; \sigma > 0$

EVI: γ

Inverse gamma distribution

Notation: $X \sim Inv\Gamma(\lambda, \alpha)$

Density function: $f(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\lambda/x}$

Distribution function: $F(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \int_0^x w^{-\alpha-1} e^{-\lambda/w} dw$

Range: $x > 0$

Parameters: $\lambda > 0; \alpha > 0$

EVI: $\frac{1}{\alpha}$

Loggamma distribution

Notation: $X \sim \log\Gamma(\lambda, \alpha)$

Density function: $f(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{-\lambda-1} (\log x)^{\alpha-1}$

Distribution function: $F(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \int_1^x w^{-\lambda-1} (\log w)^{\alpha-1} dw$

Range: $x > 1$

Parameters: $\lambda > 0; \sigma > 0$

EVI: $\frac{1}{\lambda}$

Logistic distribution

Notation: $X \sim Logistic$

Density function: $f(x) = \exp(x) / (1 + \exp(x))^2$

Distribution function: $F(x) = 1 - 1/(1 + \exp(x))$

Range: $-\infty < x < \infty$

EVI: 0

Lognormal distribution

Notation: $X \sim logN(\mu, \sigma^2)$

Density function: $f(x) = \frac{1}{\sigma\mu\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2\right)$

Distribution function: $F(x) = \frac{1}{\sigma\mu\sqrt{2\pi}} \int_0^x \exp\left(-\frac{1}{2}\left(\frac{\log(w)-\mu}{\sigma}\right)^2\right) dw$

Range: $x > 0$

Parameters: $-\infty < \mu < \infty; \sigma > 0$

EVI: 0

Normal distribution

Notation: $X \sim N(\mu, \sigma^2)$

Density function: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) = \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right)$

Distribution function: $F(x) = \int_{-\infty}^x f(w)dw = \Phi\left(\frac{x-\mu}{\sigma}\right)$

Range: $-\infty < x < \infty$

Parameters: $-\infty < \mu < \infty; \sigma > 0$

EVI: 0

Normal distribution $\mu = 0; \sigma = 1$ (Standard normal distribution)

Notation: $X \sim N(0,1)$

Density function: $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$

Distribution function: $\Phi(x) = \int_{-\infty}^x \phi(w)dw$

Range: $-\infty < x < \infty$

Parameters: $\mu = 0; \sigma = 1$

EVI: 0

Pareto distribution

Notation: $X \sim Pa(\alpha)$

Density function: $f(x) = \alpha x^{-\alpha-1}$

Distribution function: $F(x) = 1 - x^{-\alpha}$

Range: $x > 0$

Parameters: $\alpha > 0$

EVI: $\frac{1}{\alpha}$

Perturbed Pareto distribution (PPD)

Notation: $X \sim PPD(\gamma, \rho, c)$

Density function: $f(x) = \gamma^{-1} x^{-1/\gamma - 1}(1 - c + c(1-\rho)x^{\rho/\gamma})$

Distribution function: $F(x) = 1 - (1-c)x^{-1/\gamma} - cx^{-(1-\rho)/\gamma}$

Range: $x \geq 1$

Parameters: $\gamma > 0$; $\rho < 0$; $\rho^{-1} \leq c \leq 1$

EVI: γ

***t* distribution (inflected on positive half-line)**

Notation: $X \sim |t_n|$

Density function: $f(x) = \frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-\left(\frac{n+1}{2}\right)}$

Distribution function: $F(x) = \frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\int_0^x \left(1 + \frac{w^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} dw$

Range: $x \geq 0$

Parameters: $n > 0$

EVI: $\frac{1}{n}$

Uniform distribution

Notation: $X \sim U(a, b)$

Density function: $f(x) = \frac{1}{b-a}$

Distribution function: $F(x) = \frac{x-a}{b-a}$

Range: $a \leq x \leq b$

Parameters: $a < b$

EVI: -1

Weibull distribution

Notation: $X \sim Weibull(\lambda, \tau)$

Density function: $f(x) = \lambda\tau x^{\tau-1}\exp(-\lambda x^{\tau})$

Distribution function: $F(x) = 1 - \exp(-\lambda x^{\tau})$

Range: $x > 0$

Parameters: $\lambda > 0; \tau > 0$

EVI: 0

Chapter 1

Introduction

“The man who fears God will avoid all extremes.”

Ecclesiastes 7:18b NIV

The aim of this chapter is to give a brief introduction to extreme value theory (EVT), which will be expanded on in Chapter 2, and to pose the research question of this thesis. We will then expand on the contribution this thesis makes in the area of EVT. This chapter will be concluded by giving an outline of the chapters to follow.

1.1 Background and problem statement

In most situations when a data set is analyzed, the emphasis is on describing the center of the underlying distribution, aiming to give an idea of what typical observations from the underlying distribution are likely to be. This is usually done by calculating an estimate of a measure of central tendency, for example the mean, and constructing a confidence interval. In EVT the emphasis is on describing the tails of the underlying distribution. The aim is therefore not to examine typical observations, but extreme (very small or very large) observations. In this thesis we will concentrate on the right tail of the underlying distribution, that is examining extremely large observations, since most applications of EVT are motivated by the need to predict the probability of occurrence of large observations. However, EVT is equally applicable to situations where extremely small observations are of concern, even though examples of such situations are not encountered frequently.

The crucial parameter when making inferences about extreme quantiles is called the extreme value index (EVI). Estimating this parameter accurately goes hand-in-hand with accurate, useful and reliable inferences concerning extreme quantiles. To a large extent the aim of EVT is therefore the development of new estimators of the EVI which results in improvements with respect to inferences concerning extreme quantiles.

In this thesis we focus on estimating the EVI in the case where the EVI is assumed to be positive. This means that we assume that the underlying distribution of the data has a heavy right tail, as is the case with, for example, insurance claims data.

The aim of this study is threefold. The first aim is to develop an improved method of estimating the EVI in the case where it is assumed to be positive. The second is to conduct an extensive simulation study which compares the performance of the proposed estimator to a wide range of estimators in the literature. The third is to illustrate the use of the proposed estimator in practice, by means of a case study.

In working towards these aims, a number of contributions were made in the field of EVT, which will now be discussed.

1.2 Scope and contribution of the study

The following are the main points which will be addressed in this thesis:

1. The perturbed Pareto distribution (PPD) will receive special attention. One of the parameters of the PPD is the EVI. By fitting a PPD to observations above a threshold, we obtain an estimate of the EVI. We investigate different methods of estimation of the PPD parameters and also develop a threshold selection technique.
2. The relative performance of the PPD estimator and that of several estimators in the literature will be investigated by means of an extensive simulation study covering a wide range of distributions and sample sizes. It will be shown that the PPD estimator yields on average the lowest mean square error of all the estimators considered.
3. Given the complexity of the PPD estimator and the fact that it is computationally intensive, we suggest which estimators can be used as alternatives to the PPD estimator if the use of the PPD estimator is not feasible from a practical point of view.
4. A case study is presented, which illustrates how the estimators are used, and how inferences can be made in practice.

Given this background, we can summarize the contribution of the thesis in the field of EVT (for positive EVI) as the following:

1. Countless estimators of the EVI exist in the literature, but the practitioner has little information on how the performance of these estimators compare. Even though we by no means claim that our simulation study is exhaustive, it is quite extensive, covering a wide range of estimators, sample sizes and distributions. The results of the simulation study provide an idea of how estimators perform across distributions, as well as an indication of the region in which the threshold should be for a given sample size.
2. In the literature only small simulation studies are usually conducted, on a very limited set of sample sizes and distributions (because of practical considerations). There is also (without exception) no justification given for choosing the distributions from which data are simulated, and usually the distributions have no correspondence to reality, in the sense that the parameters chosen for those distributions can never represent the distribution of any real data. A further problem is that the performance of the proposed estimators is usually measured against the Hill estimator as the benchmark estimator, which is a poor estimator of the EVI. The design of the simulation study used in this thesis is thoroughly justified by

incorporating theoretical and practical considerations. This simulation study design can be used as a basis for other simulation studies.

3. We give a theoretical justification for using the PPD, and the estimation of the PPD parameters is discussed in detail with respect to Bayesian methodology. This includes a detailed description of how to apply the Gibbs sampler specifically for the purpose of estimating PPD parameters. The specific issues arising from the parameterization of the PPD are addressed from a theoretical, as well as a practical point of view. Computational aspects also receive attention, focussing on methods of approximation to speed up the Gibbs sampler. The result is a Gibbs sampler which performs extremely reliably.
4. A thorough description is given of how to use the results obtained from the Gibbs sampler in practice. This includes how to conduct inferences on the EVI, extreme quantiles and exceedance probabilities.
5. In EVT, parameter estimates are frequently calculated at a certain threshold. Only observations exceeding the threshold are included in the calculation of the parameter estimate. In most cases, the threshold is specified as a number k, which is the number of largest order statistics on which the estimate is based. For instance, for a sample of size $n = 1000$, one can choose the range of k as (say) $k = 5, 6, \dots, 900$. In situations where parameters can be calculated over a range of values of k, it is frequently specified that one should consider the estimates in a “stable” region. Even though it is intuitively clear what is meant, the choice of a stable region is very subjective. Also, one cannot compare estimators by means of a simulation study when the range of k should be chosen subjectively by the analyst. One of the major contributions of this thesis is a measure which quantifies the stability (or rather instability) of estimates across a range of thresholds. This measure can be used to objectively obtain the range of thresholds over which the estimates are most stable.
6. Applying the instability measure mentioned above, it was possible to develop a threshold selection procedure for estimating the EVI when fitting the PPD. This threshold selection technique yields an objective method of threshold selection, and a significant improvement on the method of simply choosing the threshold as a function of the sample size.
7. The estimator constructed in the abovementioned manner, is called the PPD estimator. Comparing the simulation results of the PPD estimator to the other estimators in the literature, it is clear that the PPD estimator is the method of choice. For the case of a positive EVI, we put the PPD estimator forward as a benchmark estimator against which the performance of other estimators can be measured. The PPD estimator performs well over a wide range of sample sizes and distributions.
8. Given the complexity of the PPD estimator and the fact that it is computationally extremely intensive, estimators are recommended which can be used as alternatives to the PPD estimator if the use of the PPD estimator is not feasible. These alternative estimators are recommended on the basis of their performance in the simulation study.

9. A case study of five insurance claims data sets illustrates how data sets can be analyzed in practice. It is shown to what extent discretion can/should be used, as well as how different estimators can be used in a complementary fashion to give more insight into the nature of the data and the extreme tail of the underlying distribution. The analysis is carried out from the point of raw data, to the construction of tables which can be used directly to gauge the risk of the insurance portfolio over a given time frame.

In the next section a chapter outline is given, indicating in which sections the various topics are addressed.

1.3 Organisation of the study

Chapter 2 provides an overview of extreme value theory (EVT). In Section 2.1 we present a simplified case study which introduces some key concepts in EVT. In Section 2.2 we describe the historical development of EVT. Section 2.2 also serves as a brief overview of the literature. In Section 2.3 the relevance of the field of EVT is highlighted by mentioning some areas of application. In Section 2.4 some well-known results from EVT are stated.

In Chapter 3 an estimator is developed, which can be viewed as a benchmark estimator of the EVI, to test the performance of other estimators against. In short, we determine the Bayesian estimate of the parameter γ (EVI) of the perturbed Pareto distribution (PPD). In Section 3.1 threshold models are discussed, the Hill estimator and PPD defined, and it is shown in which sense the maximum likelihood estimator of the parameter γ of the PPD is a generalization of the Hill estimator. Section 3.2 covers Bayesian estimation of the PPD parameters. In Section 3.3 we provide the methodology, design and results of a simulation study performed to assess the performance of the estimator at a range of thresholds. An adaptive threshold selection technique, which improves the accuracy of estimation of the EVI, is developed in Section 3.4. In Section 3.5 details of the programming methodology applied to ensure sound simulation results, are discussed.

The first sections of Chapter 4 are devoted to comparing the performance of the PPD estimator to that of a wide range of estimators in the literature. In Section 4.8 a modification of the PPD estimator is proposed for samples of sizes $n = 2000$ and $n = 5000$, which leads to an improved PPD estimator. Since the PPD estimator is computationally extremely involved, we also propose estimators which can be used as alternatives to the PPD estimator. Section 4.9 presents results concerning the bias of some of the estimators considered earlier in the chapter.

Chapter 5 presents the analysis of case study data as an illustration of how the techniques described in the earlier chapters can be used in practice. The data are claims data from five insurance portfolios of a South African short term insurer.

Finally, Chapter 6 will present the conclusions of this study, and identify some areas which provide scope for further research.

Chapter 2

Overview of extreme value theory

The aim of this chapter is to give an overview of extreme value theory (EVT). In the first section we present a simplified case study which introduces some key concepts in EVT. Section 2.2 covers some aspects of the historical development of EVT, and also serves as a brief overview of the literature. In Section 2.3 we highlight the relevance of the field of EVT by mentioning some areas of application. In the final section we provide some well-known results from EVT in preparation for Chapter 3, in which we start addressing the problem of extreme value index estimation.

2.1 Case study: Norwegian fire insurance

In this section a simplified case study in the field of insurance is considered in order to give a feel for the nature of the analyses one performs in an application of EVT. This case study demonstrates the need for tools other than that provided by classical statistical methods.

The data consist of a set of 375 claims paid by Storbränder, a Norwegian fire insurance company, during 1980. We use this data set for a case study, since Norwegian fire insurance data sets are real data that have received ample attention in the literature as benchmark data sets. This particular set lends itself well to illustrate basic EVT concepts.

The format in which the data are given, is shown in Table 2.1.1 below. Claim amounts are in thousands of crowns. This is the unit that will be used throughout this section and will not be stated explicitly again.

Claim No.	**Date claim occurred**	**Claim amount**
1	1980-01-01	1 200
2	1980-01-04	500
3	1980-01-06	500
:	:	:
375	1980-12-29	685

Table 2.1.1 Claims paid by Storbränder during 1980.

The observations range from a minimum of 80 to a maximum of 12 725.

Clearly, in an insurance context, it is not only typical claims sizes that are of concern, but also large claims, since large claims can have a disastrous impact on the bottom line of the insurer and can even cause financial ruin. Recent examples of events that caused excessively large insurance claims include the Sumatra tsunami and Hurricane Katrina.

The actuary consequently needs to examine large claim events and can typically ask the following questions:

a) What is the probability that a claim larger than 4000 will occur?

b) What size claim will be exceeded with probability 0.1?

In the case of these two questions classical statistical techniques suffice, even though we are not interested in central but extreme observations. This is because of the fact that the number of observations is large enough and the value of 4000 is well within the range of observed values. To illustrate how these questions can be answered, we follow the most straightforward approach available.

In order to answer the first question, we examine the data and note that 21 of the claims exceeded 4000. If we denote a random claim by X, the required probability can be estimated as $P(X > 4000) \approx 21/375 = 0.056$.

In order to answer question b), we need the value (quantile) x^* such that $P(X > x^*) = 0.1$. This value is the 90th percentile of the underlying distribution and is approximately ordered observation number $0.9(375 + 1) = 338.4$. We have to calculate the average of ordered observations 338 and 339. From the data we find $x_{338,375} = 3000$ and $x_{339,375} = 3000$. We get $x^* \approx 3000$. We estimate that a claim of size 3000 will be exceeded with a probability of approximately 0.1.

These methods break down in the situation where, for example, the following questions need to be answered:

c) What is the probability that a claim is larger than $20\,000$?

d) What size claim will be exceeded with probability 0.001?

The required probability in question c) cannot be approximated in the same fashion as we did previously, since the maximum claim size in the data is $12\,725$. There are no observations exceeding $20\,000$. Using the same technique will yield 0 as an estimate of the probability. Clearly, a claim of that size need not necessarily be impossible.

To answer question d), using the same method as before, would require ordered observation $0.999(375 + 1) = 375.6$, that is the average of ordered observation 375 and 376. But, since we have only 375 observations, this value cannot be calculated.

An alternative method, one which allows us to extrapolate beyond the data, is required. This is where EVT comes in.

In order to help us identify the type of distribution we are dealing with, let us first consider a histogram of the data, given in Figure 2.1.1 below.

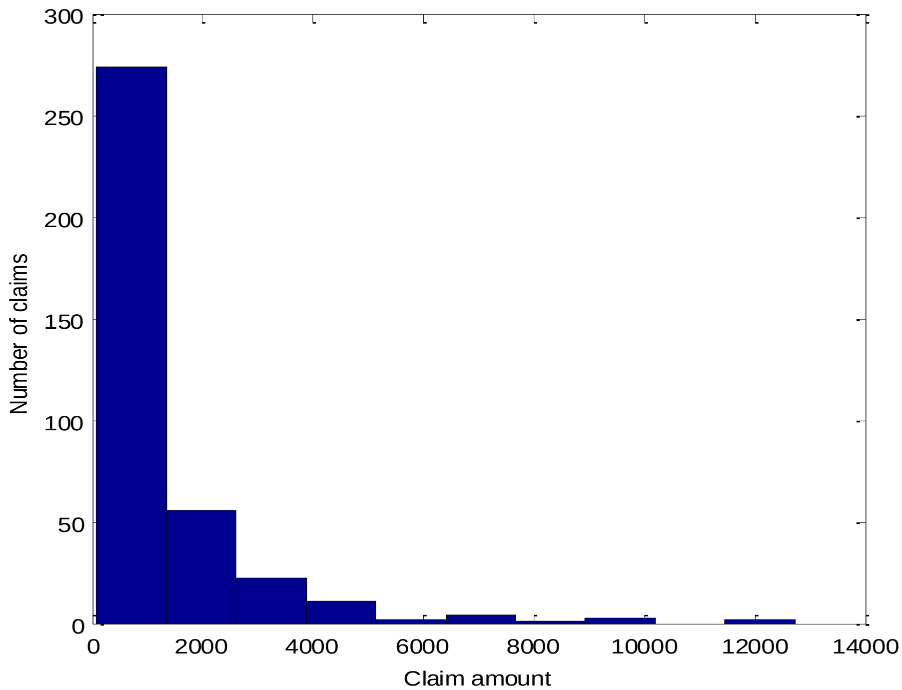

Figure 2.1.1 Histogram of the Storbränder data.

The salient feature of the underlying distribution is immediately evident from the graph, namely that it is a so-called heavy-tailed distribution. The exact mathematical definition of heavy-tailed distributions will follow in Section 2.4.2. Suffice it to say at this point that it means that some very large observations are likely to occur, something which is not uncommon for insurance claims data.

Note that questions c) and d) posed by the actuary can be answered if a plausible model for the underlying distribution is available. The availability of such a model permits the calculation of tail probabilities and quantiles.

One method of establishing the plausibility of a proposed model is by constructing a Q-Q plot of the data. A Q-Q plot is a graphical tool used to assess goodness of fit, that is whether observations $x_1, x_2, \dots, x_n$ of a random variable X support F as the underlying distribution.

The empirical quantiles are $x_{1,n} \leq x_{2,n} \leq \cdots \leq x_{n,n}$, with associated probabilities $1/n$, $2/n$, …, n/n since $P(X \leq x_{i,n}) \approx i/n$. Consequently, if the distribution of X is F, a plot of $\left(Q(i/n)\,;x_{i,n}\right)$, $i = 1,2,\dots,n$, where Q is the quantile function associated with F, should be approximately linear with slope 1 and intercept 0.

Since we are comparing a discontinuous function with a continuous function and in order to avoid overflow at n/n, a continuity correction should be applied. Instead of i/n, $(i-0.5)/n$ or $i/(n+1)$ should be used. In this thesis $i/(n+1)$ will be used.

The formal definition of the Q-Q plot follows:

Definition 2.1.1 Q-Q plot

A Q-Q plot is a graphical tool used to visually assess goodness of fit. A plot of $\left(Q\left(i/(n+1)\right); x_{i,n}\right)$, $i = 1,2,\ldots,n$, *should be approximately linear if* $x_1, x_2, \ldots, x_n$ *are from a distribution with quantile function* Q.

An advantage of Q-Q plots is the fact that location and scale parameters of the model we fit need not be known in advance and can even be estimated from the Q-Q plot. Consider, for example, the normal Q-Q plot. If we have observations $x_1, x_2, \ldots, x_n$ from a normal distribution with mean μ and variance σ^2, it can easily be shown that a plot of $\left(\Phi^{-1}\left(i/(n+1)\right); x_{i,n}\right)$, $i = 1, 2, \ldots, n$, should be approximately linear with slope σ and intercept μ, where $\Phi(\cdot)$ is the standard normal distribution function. This leads to the following definition:

Definition 2.1.2 Normal Q-Q plot

Given observations $x_1, x_2, \ldots, x_n$, *a plot of* $\left(\Phi^{-1}\left(i/(n+1)\right); x_{i,n}\right)$, $i = 1, 2, \ldots, n$, *is called the normal Q-Q plot of the data, where* $\Phi(\cdot)$ *is the standard normal distribution function. If observations* $x_1, x_2, \ldots, x_n$ *are from a normal distribution with mean* μ *and variance* σ^2, *the plot should be approximately linear with slope* σ *and intercept* μ.

Keep in mind that, to answer the questions posed, interest centers mainly on the right tail of the distribution and not on the distribution on its entire support. Therefore only observations above a certain threshold need to be considered. Fortunately it is known from EVT that, if the underlying distribution is heavy-tailed and the threshold is large enough, the tail of the distribution above the threshold is approximately that of a Pareto tail. This result will be formally stated in Section 3.1.2. Before we continue, let us first define the Pareto distribution and its corresponding Q-Q plot.

Definition 2.1.3 Pareto distribution

If a random variable X is distributed Pareto with parameter $1/\gamma$, *denoted* $X \sim Pa(1/\gamma)$, *then* $\bar{F}(x) = x^{-1/\gamma}$ *and* $f(x) = (1/\gamma)x^{-1/\gamma-1}$, *where* $x \geq 1$ *and* $\gamma > 0$.

The parameterization of the Pareto is in terms of $1/\gamma$, since γ is the extreme value index (EVI). This will be shown formally in Section 2.4.4.

If $X \sim Pa(1/\gamma)$, it can easily be shown that $Y = \log X \sim exp(1/\gamma)$. It follows that an exponential Q-Q plot of the log of the observations is equivalent to a Pareto Q-Q plot. For an $exp(1/\gamma)$ distribution the quantile function is $Q(p) = -\gamma \log(1-p)$. This leads to the following definition of the Pareto plot (Pareto Q-Q plot):

Definition 2.1.4 Pareto plot

Given observations $x_1, x_2, \ldots, x_n$, a plot of $\left(-\log\left(1 - i/(n+1)\right); \log\left(x_{i,n}\right)\right)$, $i = 1,2, \ldots, n$, is called the Pareto plot of the data. If observations $x_1, x_2, \ldots, x_n$ are from a $Pa(1/\gamma)$ distribution, the plot should be approximately linear with slope γ.

As noted above, if the underlying distribution is heavy-tailed and the threshold is large enough, the tail of the distribution (the distribution above a threshold) can be approximated by a Pareto tail. This implies that the Pareto plot will eventually become linear as the sizes of the observations increase.

Let us now return to our case study and see how the Pareto plot assists us in modeling the distribution.

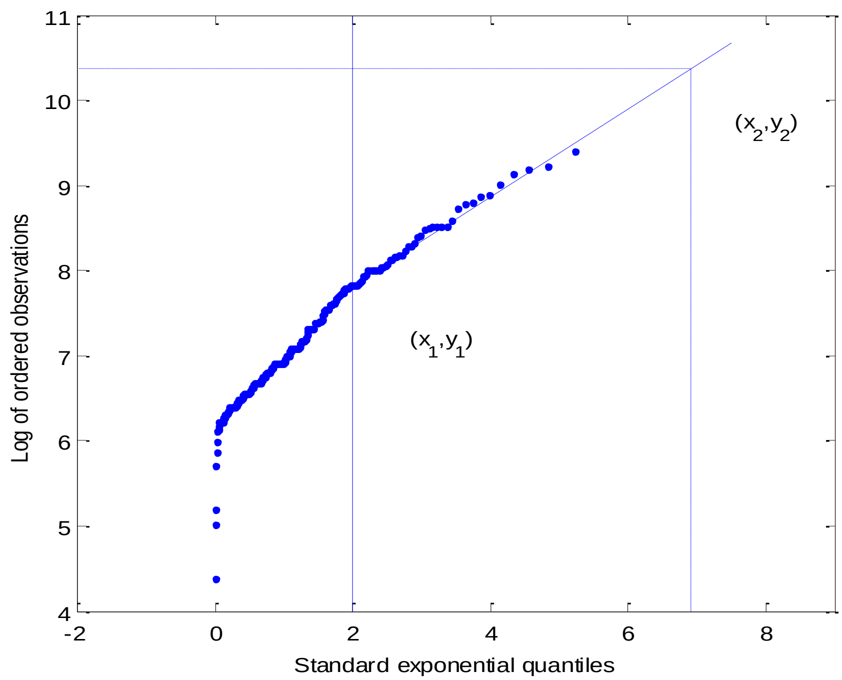

Figure 2.1.2 Pareto plot of the Storbränder data.

Examining Figure 2.1.2, we see that it becomes approximately linear in the vicinity of $x = 2$. Solving for i in the equation $x = -\log(1 - i/376) = 2$, we obtain $i \approx 325$. The observation associated with that point on the graph therefore corresponds more or less with observation

$x_{325,375}$ = 2500 (yielding x_1 = $-\log(1 - 325/376)$ = 1.9978 and y_1 = $\log 2500$ = 7.824). The slope from that point onwards is estimated as 0.5183. This estimate is called the Hill estimate and will be discussed in Section 3.1.3. The solid vertical line indicates the threshold of 1.9978 from where we believe the plot becomes approximately linear.

If we now assume that we have a reliable choice of the threshold (large enough) and a reliable estimate of the slope from that point onwards, we can extend the diagonal line beyond the data. This is in effect a method of inference concerning large quantiles, even larger than we have observed in the data.

We are now able to answer the previously posed questions. Consider question d) again:

What size claim will be exceeded with probability 0.001?

To answer this question, we calculate x_2 = $-\log 0.001$ = 6.9078 (indicated by the dashed vertical line) and from that y_2 = y_1 + $0.5183(x_2 - x_1)$ = 10.3691 (indicated by the dashed horizontal line). The size claim in question is then estimated as $e^{10.3691}$ = $31\,859.957$.

Consider question c) again:

What is the probability that a claim larger than $20\ 000$ will occur?

We calculate $y_3 = \log 20000$ = 9.9035. This yields x_3 = x_1 + $(y_3 - y_1)/0.5183$ = 6.0095. Solving $-\log(1-p) = 6.0095$ gives the probability in question, namely $1-p$ = 0.0025.

These fundamental questions have prompted much research over the years. An overview of some of the milestones in the development of EVT will now be presented.

2.2 Historical development of EVT

We will now present some highlights in the historical development of EVT, leaving the more technical issues to Section 2.4, which will include some definitions and theorems.

Heavy-tailed distributions are frequently encountered in practice and therefore the problem of finding appropriate models for heavy-tailed data is quite old. The need arose for distributions of which the tail decays slower than an exponentially decaying tail, the latter of which the exponential and normal distributions are examples. This led Pareto (1897) to define the most basic distribution which satisfies this requirement, namely a distribution with a polynomially decaying tail, with tail function $\bar{F}(x) = x^{-\alpha}$, where α > 0. Pareto used this distribution to model various income distributions.

Zipf (1949) fitted the Pareto distribution to a wide variety of phenomena. This includes examples from business, economics, commerce, industry, communication, travel, traffic, sociology, psychology, music, politics, and warfare. He shows graphically that an approximate Pareto fit is convincing, especially in the tails.

Another approach can also be followed. Instead of trying to find a model that fits data, either on its entire support or just in the tails, one can attempt to find the distribution of the maximum (or minimum) of the data. It is exactly in this area that the first significant contribution to the field of EVT was made by Fisher and Tippett (1928) with their paper entitled *On the estimation of the frequency distributions of the largest or smallest member of a sample*. The problem of finding the limiting distribution of the maximum of a series of random variables was also later solved by Gnedenko (1943).

The Fisher-Tippett Theorem (given in Section 2.4.1) proves a remarkable result. Simply put, the theorem states that, if the distribution of the (normalized) maximum of a sequence of random variables converges, it always converges to the generalized extreme value distribution, regardless of the underlying distribution F.

Note that the result is in nature very similar to the central limit theorem, which states that the normalized mean of a sample of independent, identically distributed random variables with common distribution function F, tends to the normal distribution (under some assumptions).

The next important contribution to EVT was the book *Statistics of extremes* by Gumbel (1958), which was considered the main reference work for application of EVT in the field of engineering for a number of years. This publication also reveals that in the early stage of the development of the field, the attempt was to approach EVT using results from central limit theory.

Concurrent with development of EVT, was the development of the theory of regular variation in mathematics, which later played a crucial role in EVT. The modern period of the theory of regular variation started with Karamata (1930). The developments in the field of regular variation from the beginning up till 2007, as well as its impact on EVT, are described by Bingham (2007) in his paper *Regular variation and probability: the early years*. Bingham was also a co-author of the book entitled *Regular variation* (1987), a comprehensive text on the subject.

During the years which followed *Statistics of extremes* by Gumbel (1958), theoretical interest in the subject of EVT was limited, until interest was revived by the pivotal doctoral dissertation by L. de Haan (1970). De Haan refined the result of Fisher and Tippet (1928) and also refined Karamata's results on regular variation in order to provide a rigorous mathematical framework for key results in EVT.

De Haan obtained his doctorate at the University of Amsterdam and it is no coincidende that the first substantial theoretical interest in the field of EVT had its genesis in the Netherlands. One of the first major applications of EVT concerned the dike construction in the Netherlands. It is self-evident that the accurate prediction of extreme water levels is crucial. Constructing the dikes unnecessarily high will involve spending an enormous amount of money which could be better applied to other infrastructural developments, whereas constructing it too low may lead to a catastrophy, like the flood of 1953. Nowadays it is required that the low-land countries (Belgium and the Netherlands) have dikes built as high as the 10^4 year flood discharge level. This means that the dike should be so high that the water level exceeds it only once in 10 000 years on average.

Since the publication of De Haan's dissertation (1970), interest in EVT abounded. From the 1990s onwards the field exploded with numerous papers and dissertations every year. One of the main reasons for the increased interest in EVT is the wide range of applications of it in finance, insurance and related fields.

Since 1970 interest has centered on the main themes of EVT, namely construction of estimators of the EVI, estimation of large quantiles, threshold selection techniques, different methods of estimation (including Bayesian methods), application of results of the theory of regular variation in an attempt to reduce bias of estimators and derive new models, expanding the application of EVT to regression, multivariate settings, time series, etc.

Among the vast collection of papers and books on the subject, the following are among the most noteworthy:

- The book *Extremes and related properties of random sequences and processes* by Leadbetter, Lindgren and Rootzén (1983), had an enormous impact, especially in time series by the relaxtion of conditions from independent variables to stationary sequences. They also developed a much broader characterization of extremal behaviour.
- The review article by Coles and Powell (1996), entitled *Bayesian methods in extreme value modelling: a review and new developments*, provides an instructive explanation of the application of the Bayesian paradigm in EVT and a literature review of important Bayesian publications in EVT up to 1996.
- No literature review would be complete without mentioning *Modelling extremal events for insurance and finance* by Embrechts, Klüppelberg and Mikosch (1997). This is an excellent book which presents an extensive overview of mathematical theory as well as applications, mainly in insurance and finance.
- The paper *Estimating the tail index* by Csörgő and Viharos (1998) is a more technical review of the subject of tail index (EVI) estimation, and features an extensive literature review, referring to 138 publications.
- The 2001 book by Coles, *An introduction to statistical modeling of extreme values*, is an excellent book with which to start the journey into EVT. The arguments flow nicely and the concepts are well explained, keeping technical details to a minimum.
- A more recent publication *Statistics of extremes* is a comprehensive book on the subject of extremes. Written by Beirlant, Goegebeur, Segers and Teugels (2004), it is understandable that this can be regarded as one of the main texts in extremes. It is an indispensable reference

for any practitioner of EVT. The book covers an enormous number of theoretical and practical aspects of EVT, referring to approximately 400 publications.

- *Extreme value theory: an introduction* by De Haan and Ferreira (2006) gives a rigorous introduction to the field of extremes, also focussing on the probablistic and theoretical side of EVT up to the cutting edge of inference for extreme values.

The brief overview of EVT literature that was given in this section will be expanded upon in the sections and chapters which follow. Since a major part of this thesis concerns itself with comparison of different estimators, much detail on what was mentioned above will be added as we proceed.

We will, however, for practical reasons limit ourselves to aspects of the literature relevant to our study. In Section 2.4 specifically, we will provide the theoretical background we need in order to justify, construct and test the performance of the PPD estimator in a meaningful way. The PPD estimator is our benchmark estimator which we propose and examine in Chapter 3.

Given all the contributions made in the literature, EVT has become an essential tool in the analysis of data from a wide range of practical applications, which will be discussed in the next section.

2.3 Areas of application

We mentioned in Section 2.2 the areas Zipf (1949) applied extreme value theory (EVT) to by fitting the Pareto distribution, namely business, economics, commerce, industry, communication, travel, traffic, sociology, psychology, music, politics, and warfare.

Further areas of application will now be mentioned, with a brief description of each. The list below contains only a few applications, and the descriptions do not provide a great amount of detail. Other texts in the literature can be consulted for more extensive discussions on a wide variety of applications, such as Beirlant et al. (2004).

Hydrology: EVT is used in flood frequency analysis, where it is of interest to estimate the T-year flood discharge, which is the water level exceeded every T years on average. An example of the use of EVT in this domain of application is the dike constructions in the Netherlands. Another parameter of interest is rainfall intensity, which is used to model water course systems, urban drainage and water runoff. Here again it is important to model extreme events accurately.

Environmental research and meteorology: Meteorological data, such as ozone concentration, daily maximum temperatures and wind speeds, are of interest in order to predict situations where extreme situations might have negative consequences.

Finance applications: Risk management at a commercial bank is intended to guard against risk or loss due to fall in prices of financial assets held or issued by the bank. EVT is used to calculate the maximum (extreme) losses that can occur in a given time period. The value at risk (VaR) of a portfolio needs to be estimated, which is the level below which the future portfolio will drop with only a specified small probability.

Insurance applications: This is one of the most prominent areas of applications of EVT, since it is important to model the probability of excessively large claims.

Geological and seismic analysis: Application of extreme value statistics in geology can be found in the magnitudes of and losses resulting from earthquakes, and in diamond sizes and values, amongst others.

Metallurgy: Here EVT is used to model metal fatigue, which causes failure of a metallic component under extreme stress.

It is clear that the acceleration of interest in EVT is justified, since it provides a framework which can be used to answer crucial and relevant questions in numerous areas of application.

Some well-known results from EVT will now be presented, in preparation for Chapter 3, in which we start addressing the problem of extreme value index estimation.

2.4 Basic results from EVT

The aim of this section is to state some basic results in extreme value theory, which expand on some of the topics mentioned in Section 2.2 and will enable us to discuss the PPD estimator of Chapter 3 in a meaningful way.

We will specify precisely what is meant by heavy-tailed distributions. Some well-known results from EVT which characterize heavy-tailed distributions will be stated, together with examples of such distributions.

2.4.1 The generalized extreme value distribution

We begin by discussing the generalized extreme value distribution, which appears as a limiting distribution in the Fisher-Tippett Theorem (Beirlant et al., 2004).

Theorem 2.4.1 Fisher-Tippet Theorem

Let $M_n = \max\{X_1, X_2, \dots, X_n\}$, where $X_1, X_2, \dots, X_n$ is a sequence of independent, identically distributed random variables with common distribution function F.

If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$P((M_n - b_n)/a_n \le z) \to G(z) \text{ as } n \to \infty$$

for a non-degenerate distribution function G, then G is a member of the generalized extreme value family, with distribution function

$$G(z) = \exp\left\{-\left[1+\gamma\left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\gamma}}\right\}, \qquad (\mathbf{2.1})$$

defined on $\left\{z \middle| 1+\gamma\left(\frac{z-\mu}{\sigma}\right)>0\right\}$, where $-\infty<\mu<\infty$, $\sigma>0$ and $-\infty<\gamma<\infty$.

The case where $\gamma = 0$ is interpreted as the limit of (2.1) as $\gamma \to 0$, which yields

$$G(z) = \exp\left(-\exp\left(-\frac{z-\mu}{\sigma}\right)\right),$$

defined on $-\infty<z<\infty$, where $-\infty<\mu<\infty$ and $\sigma>0$. ■

The term *non-degenerate distribution* refers to a distribution which does not have all its mass in one point. As an example of a degenerate limiting distribution, the most common example is the limiting distribution of

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ as } n \to \infty,$$

which has all its mass in the point $E(X)$, if $E(X)$ exists.

The distribution specified by (2.1) is called the generalized extreme value (GEV) distribution. The strength of Theorem 2.4.1 lies in the fact that, if the limiting distribution exists, it belongs to the same family of distributions regardless of the distribution of F. Also, Theorem 2.4.1 holds for a very wide range of continuous distributions F and certainly for all distributions that will be mentioned in this thesis. In fact Theorem 2.4.1 holds for all distributions with a continuous, strictly increasing distribution function F, and even this assumption can be relaxed further when proving Theorem 2.4.1 in its full generality. Refer to Section 2.1 in Beirlant et al. (2004) for more on the relaxation of assumptions.

From Theorem 2.4.1 it is clear that it is natural to fit the GEV distribution to maxima of long sequences of observations. In applications this result will often be applied to block maxima. As an example of this, consider for instance yearly maxima of wind speeds. Each maximum is the maximum of a long sequence (365 observations) of daily maxima. The GEV distribution is then fitted to the observed yearly maxima.

It should be noted that the daily maxima are most probably not independent and identically distributed, since stronger winds will occur certain times of the year. However, even though the formal justification for the GEV distribution is invalid, it is reasonable to assume that the yearly maxima are identically distributed, hence fitting the GEV is still appropriate. For a more detailed discussion, see Coles (2001).

The apparent difficulty that the normalizing contants are unknown in pratice is easily resolved. Assuming n is large enough for the GEV to provide a reasonable approximation of the normalized maximum $(M_n - b_n)/a_n$, then

$$P((M_n - b_n)/a_n \leq y) \approx G(y) \Rightarrow P(M_n \leq a_n y + b_n) \approx G(y).$$

Letting $z = a_n y + b_n$, yields $P(M_n \leq z) \approx G((z - b_n)/a_n)$. In equation (2.1) $(z - \mu)/\sigma$ becomes $(z - (b_n + \mu a_n))/(a_n \sigma)$.

Therefore, the distribution of M_n is also approximately GEV with location and scale parameters $\mu^* = b_n + \mu a_n$ and $\sigma^* = a_n \sigma > 0$, respectively.

2.4.2 Three classes of distributions

It has been shown that the crucial parameter that distinguishes between classes of distributions to which the distribution F belongs, is the parameter γ in Theorem 2.4.1. This parameter is called the *extreme value index* (EVI).

The EVI distinguishes the class to which F belongs in the following way:

$\gamma < 0$: The (extremal) Weibull class

For $\gamma < 0$, F is said to belong to the (extremal) Weibull class. These distributions have a finite right endpoint of support.

Examples include the following distributions: uniform and beta.

$\gamma = 0$: The Gumbel class

For $\gamma = 0$, F is said to belong to the Gumbel class. These distributions have an infinite right endpoint of support and the right tail of these distributions decays exponentially.

Examples include the following distributions: normal, Weibull, exponential, gamma and lognormal.

$\gamma > 0$: The Fréchet-Pareto class

For $\gamma > 0$, F is said to belong to the Fréchet-Pareto class. These distributions are called *heavy-tailed*. They have an infinite right endpoint of support and the right tail of these distributions decays polynomially.

Examples include the following distributions: Pareto, Fréchet, Burr, F, inverse gamma and loggamma.

The term *domain of attraction* is sometimes used. The set of distributions attracted to G (the GEV distribution of Theorem 2.4.1) is said to fall in the domain of attraction of G. For instance, the Burr distributions falls in the domain of attraction of G with $\gamma > 0$.

2.4.3 The generalized Pareto distribution

Given a sequence of independent, identically distributed random variables $X_1, X_2, \dots, X_n$ with common distribution function F, it seems wasteful to just consider the maximum, since there might be more order statistics that contain information on the tail of the underlying distribution.

A natural progression from the maximum is to consider the distribution of the k largest order statistics. This leads to a number of complexities and quite cumbersome mathematical derivations. We will not pursue this avenue further here. For more information on this subject, see De Haan and Ferreira (2006), Coles (2001) and Beirlant et al. (2004).

Another approach, which we apply in this thesis, is to consider threshold models. The idea is to specify a threshold t above which observations are regarded as being in the tail of the distribution F.

The following theorem states a well-known result in EVT. It is stated in Coles (2001), Beirlant et al. (2004), and Embrechts et al. (1997), amongst others.

Theorem 2.4.3

Let X be a random variable with distribution function F. Then, for large enough t, the distribution of $Z = X - t$ conditional on $X \geq t$ is approximately

$$G(z) = 1 - \left(1 + \frac{\gamma z}{\sigma}\right)^{-\frac{1}{\gamma}} \qquad \mathbf{(2.2)}$$

defined on $\left\{z \middle| z \geq 0; 1 + \frac{\gamma z}{\sigma} > 0\right\}$, where $-\infty < \gamma < \infty$ and $\sigma > 0$.

The case $\gamma = 0$ is interpreted as the limit obtained when letting $\gamma \to 0$, which yields

$$G(z) = 1 - \exp(-z/\sigma)$$

defined on $z > 0$, where $\sigma > 0$. ■

The distribution specified by (2.2) is called the generalized Pareto distribution (GPD) for real γ. The restriction $\gamma > 0$ is sometimes applied to the GPD. See for instance Beirlant et al. (2004) and Section 2.4.4. It should therefore always be clear whether *GPD* refers to the GPD where $\gamma > 0$ or where $-\infty < \gamma < \infty$. In this thesis we will always assume *GPD* to refer to the case where $\gamma > 0$.

2.4.4 The Fréchet-Pareto class

This is the class of heavy-tailed distributions, to be considered in this thesis. They occur frequently in practice, for example insurance claims data, sizes of files transferred form a web-server, earthquake magnitudes, and diamond values.

It has been mentioned that examples of distributions in the Fréchet-Pareto class include the Pareto, Fréchet, Burr, F, inverse gamma and loggamma. We are now going to consider some of the properties of these distributions. In order to do this, we first need some definitions and results from the theory of regular variation. See for instance Geluk, De Haan, Resnick and Starica (1997).

Definition 2.4.4.1 Regular variation

Let X be a random variable with distribution function F concentrated on $[0,\infty)$ and tail function $\overline{F} = 1 - F$.

$\overline{F}$ is said to be regularly varying with index $-1/\gamma, \gamma > 0$, if

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-\frac{1}{\gamma}} \qquad (\mathbf{2.3})$$

for all $x > 0$.

The right-hand side of (2.3) can be recognized as the tail function of the Pareto distribution, which is why distributions satisfying (2.3) are said to be of *Pareto type.*

Distributions satisfy (2.3) if and only if they belong to the Fréchet-Pareto class (De Haan and Ferreira, 2006).

Hence, the following statements are equivalent for a distribution with distribution function F and tail function $\overline{F} = 1 - F$:

- F is heavy-tailed.
- The EVI of F is positive.
- F belongs to the Fréchet-Pareto class.
- F is in the domain of attraction of the GEV distribution with $\gamma > 0$.
- $\overline{F}$ is regularly varying with index $-1/\gamma, \gamma > 0$.
- F is of Pareto type.

Definition 2.4.4.2 Second order regular variation

A regularly varying tail function $\overline{F}$ is said to be second order regularly varying with first order index $-1/\gamma$, $\gamma > 0$, and second order index ρ/γ, $\rho \leq 0$, if a function $A(t)$ exists which tends to 0 *and is ultimately of constant sign as $t \to \infty$, such that*

$$\lim_{t\to\infty} \left(\frac{\overline{F}(tx)}{\overline{F}(t)} - x^{-1/\gamma}\right)(A(t))^{-1} = H(x)$$

for all $x > 0$, where $H(x) = dx^{-1/\gamma} logx$ if $\rho = 0$ and $H(x) = dx^{-1/\gamma}\frac{x^{\rho/\gamma}-1}{\rho/\gamma}$ if $\rho < 0$, with $-\infty < d < \infty$.

Note that Geluk et al. (1997) specify that $d \neq 0$. De Haan and Ferreira (2006) also specify $d \neq 0$. This restriction is not applied in Beirlant et al. (2004).

Whether or not d may be zero is of theoretical rather than practical significance. Theoretically we should have $d \neq 0$, since if $d = 0$, ρ is undefined, and consequently the second order index is also undefined.

The advantage of allowing d to be zero, is that the Pareto distribution can be included as a special case. For the Pareto distribution $\overline{F}(tx)/\overline{F}(t) = x^{-1/\gamma}$, and therefore $d = 0$. The Pareto distribution can then also be seen as a special case of the perturbed Pareto distribution (refer to Section 3.1.4) when $c = 0$, which can only be the case if $d = 0$.

The strange parameterization in Definitions 2.4.4.1 and 2.4.4.2 eases interpretation. In these definitions γ is called the *first order parameter* (also the EVI) and ρ is called the *second order parameter*. Definitions 2.4.4.1 and 2.4.4.2 will be used to derive estimators for the EVI. See Sections 3.1.2 and 3.1.4.

Table 2.4.4 lists some of the Pareto type distributions, together with their corresponding first order and second order parameters:

Distribution	**Notation**	$\mathbf{1-F(x)}$	**Conditions**	$\boldsymbol{\gamma}$	$\boldsymbol{\rho}$
Pareto	$Pa(\alpha)$	$x^{-\alpha}$	$x \geq 1$ $\alpha > 0$	$\frac{1}{\alpha}$	undef.
generalized Pareto	$GP(\gamma, \sigma)$	$\left(1+\frac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}}$	$x \geq 0$ $\gamma, \sigma > 0$	γ	$-\gamma$
Burr	$Burr(\beta, \tau, \lambda)$	$\left(\frac{\beta}{\beta+x^{\tau}}\right)^{\lambda}$	$x > 0$ $\beta, \tau, \lambda > 0$	$\frac{1}{\lambda\tau}$	$-\frac{1}{\lambda}$
Fréchet	$Fréchet(\alpha)$	$1-\exp(-x^{-\alpha})$	$x > 0$ $\alpha > 0$	$\frac{1}{\alpha}$	-1
t with n d.f.	$\lvert t_n \rvert$	$\frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\int_x^{\infty} g(w)dw$ $g(w) = \left(1+\frac{w^2}{n}\right)^{-\left(\frac{n+1}{2}\right)}$	$x \geq 0$ $n > 0$	$\frac{1}{n}$	$-\frac{2}{n}$
loggamma	$\log\Gamma(\lambda, \alpha)$	$\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\int_x^{\infty} g(w)dw$ $g(w) = w^{-\lambda-1}(\log w)^{\alpha-1}$	$x > 1$ $\lambda, \alpha > 0$	$\frac{1}{\lambda}$	0

Table 2.4.4 Pareto type distributions

Similar tables appear in Beirlant et al. (2004) and Dierckx (2000).

Note that the t distribution is defined on $(-\infty, \infty)$, which poses a problem, since by definition F should be defined on $[0, \infty)$. We therefore inflect the distribution on the positive half-line, denoting it by $|t_n|$.

A special class of Pareto type distributions, called the *Hall class*, will now be defined (Dierckx, 2000 and Beirlant et al., 2004). If a distribution belongs to the Hall class, it simplifies obtaining the first and second order parameters considerably, as will be seen in Section 2.4.5.

Definition 2.4.4.3 Hall class of distributions

If F is of Pareto type with first order parameter $\gamma > 0$ and second order parameter $\rho \leq 0$, F is said to belong to the Hall class of distributions if

$$\overline{F}(x) = Cx^{-\frac{1}{\gamma}}\left[1 + Dx^{\frac{\rho}{\gamma}}(1 + o(1))\right]$$

as $x \to \infty$, where $C > 0$ and $-\infty < D < \infty$.

All distributions in the Hall class are second order regularly varying (Beirlant et al., 2004).

Not all Pareto type distributions are second order regularly varying, but one has to go out of one's way to construct examples of such distributions. De Haan and Ferreira (2006) give an example of such a distribution in Exercise 2.7 on p. 61:

$$\overline{F}(x) = x^{-1}[1 + x^{-1}\exp(\sin(\log x))]$$

Most common Pareto type distributions are second order regularly varying, and certainly all Pareto type distributions mentioned in this thesis are second order regularly varying.

For all practical purposes one can assume that all Pareto type distributions are first and second order regularly varying. This is a crucial assumption, since the estimator constructed in Chapter 3 is derived from this assumption.

2.4.5 The Burr distribution

Of the distributions listed in Table 2.4.4, the Burr distribution deserves some extra attention.

The Burr distribution can be reparameterized in terms of β, γ and ρ. Solving for λ and τ from the equations $\gamma = 1/\lambda\tau$ and $\rho = -1/\lambda$ in Table 2.4.4, yields $\lambda = -1/\rho$ and $\tau = -\rho/\gamma$. Therefore, specifying β, γ and ρ completely specifies the Burr distribution. The notation *Burr*(β, γ, ρ) will indicate that this parameterization is used.

It should be clear that, from the distributions listed in Table 2.4.4, the Burr distribution is the most flexible, in the sense that γ and ρ can be specified independently. This property makes the Burr distribution popular to use when simulating data.

As an example of how to obtain γ and ρ for a given distribution, we derive the formulas for γ and ρ for the Burr distribution:

For the Burr distribution we have $\overline{F}(x) = \left(\frac{\beta}{\beta + x^\tau}\right)^\lambda = \left(1 + \frac{x^\tau}{\beta}\right)^{-\lambda} = \left(\frac{x^\tau}{\beta}\right)^{-\lambda}\left(1 + \frac{\beta}{x^\tau}\right)^{-\lambda}$.

Since (from Taylor's formula) we have $(1+x)^{\alpha} \approx 1+\alpha x$ if $x \approx 0$, we have for x large

$$\overline{F}(x) = \beta^{\lambda} x^{-\lambda\tau}[1 + -\lambda\beta x^{-\tau}(1+o(1))].$$

Therefore, the Burr distribution belongs to the Hall class, where $\gamma = 1/\lambda\tau$, and $\rho = -1/\lambda$ (since $-\tau = \rho/\gamma = \rho\lambda\tau \Rightarrow \rho = -1/\lambda$).

Conclusion

The Norwegian fire insurance case study pointed out some ideas:

- Classical statistical tools are inadequate when we need to make inferences concerning quantiles beyond the scope of the data. This is where results from EVT need to be used.
- We have seen the prominent role played by the Pareto distribution, and the interpretation of the EVI as the ultimate slope of the Pareto plot.
- Two pertinent questions in EVT arose, namely the choice of threshold (beyond which we assume to be in the tail of the distribution) and the estimation of the EVI, given the threshold.

Section 2.2 presented an overview of the historical development of EVT, and mentioned some of the major publications in the literature.

In Section 2.3 we showed that the growth of interest in the field is justified by mentioning various areas of application.

In Section 2.4 the following was covered:

- The Fisher-Tippett Theorem was given, which shows the GEV distribution as the limiting distribution of the normalized maximum. Issues surrounding the use of the result in practice were also discussed.
- We also introduced the concept of threshold models, where not only the maximum of the data is considered, but all observations above a threshold. The GPD was given as an example of such a model.
- It was shown that distributions fall into one of three classes, namely the (extremal) Weibull class ($\gamma < 0$), the Gumbel class ($\gamma = 0$) or the Fréchet-Pareto class ($\gamma > 0$).
- We paid special attention to the Fréchet-Pareto class, which contains the heavy-tailed distributions with $\gamma > 0$. It was seen that distributions in this class are first and second order regularly varying and we defined the Hall class of distributions as a special subclass.
- The Burr distribution received some special mention because of its flexibility. We saw that for the Burr distribution one can specify the first and second order paramters (γ and ρ) independently.

From this foundation of basic results from EVT, we can now progress towards the construction of an estimator.

Chapter 3

The perturbed Pareto distribution estimator

In this chapter we develop an estimator, which we can be viewed as a benchmark estimator of EVI, to test the performance of other estimators against.

We will define the perturbed Pareto distribution (PPD), which is a heavy-tailed distribution, having as one of its parameters the EVI γ. We determine the Bayesian estimate of γ as our estimate of the EVI. Despite the simplicity of this basic idea, there are a number of technical and numerical difficulties which needed to be addressed.

In Section 3.1 threshold models will be discussed, as well as the most famous estimator of the EVI, namely the Hill estimator. The PPD will be defined in Section 3.1, and it will be shown in which sense the maximum likelihood estimator of the parameter γ is a generalization of the Hill estimator.

Section 3.2 covers Bayesian estimation of the PPD parameters. The section begins with a brief overview of Bayesian theory, followed by a motivation for the choice of prior distribution, and a detailed description of the Gibbs sampler. In the remainder of Section 3.2 it is shown how the results from the Gibbs sampler can be used for the purpose of making inferences.

Section 3.3 presents the methodology, design and results of a simulation study performed to assess the performance of the estimator at a range of thresholds.

Section 3.3 presents the methodology, design and results of a simulation study performed to assess the performance of the estimator referred to above at a range of thresholds.

In Section 3.4 we develop an adaptive threshold selection technique based on the stability of the estimates over a range of thresholds, and show through simulation that this technique improves on the performance of estimation at a fixed threshold.

Section 3.5 discusses details of the programming methodology applied to ensure sound simulation results.

3.1 Threshold models

In Section 2.4.3 we have already encountered a threshold model, namely the generalized Pareto distribution (GPD). In this section we will look at some other threshold models, how they are derived and how the parameters of the models are estimated.

3.1.1 Additive and multiplicative excesses

Let $X_1, X_2, \dots, X_n$ be a sequence of independent, identically distributed random variables with common distribution function F concentrated on $[0, \infty)$. There are two ways in which one can define the threshold above which one assumes to be in the tail of the distribution. One is to specify the threshold itself, denoted by t. The other is to specify the number of observations exceeding the threshold, denoted by k. These approaches are equivalent and the relation between t and k will be taken as $t = X_{n-k,n}$. Note that, since we only work with distributions with support on the positive half-line, the threshold is always positive.

There are also two ways in which excesses can be determined for a given threshold t, namely additively and multiplicatively. If X is distributed F and we define Z as the additive excess given a threshold t, then $Z = X - t$ given $X \geq t$. If we define Z as the multiplicative excess given a threshold t, then $Z = X/t$ given $X \geq t$.

If the number of excesses is k, we determine the k excesses $Z_1, Z_2, \dots, Z_k$ and fit the appropriate model to these excesses. Since it does not make sense to fit a distribution to one observation, the restriction $2 \leq k \leq n - 1$ will always be applied.

In the case of additive excesses $Z_i = X_{n-i+1,n} - X_{n-k,n}$, $i = 1,2, \dots, k$. Here $Z_i \geq 0$ for all $i = 1,2, \dots, k$.

In the case of multiplicative excesses $Z_i = X_{n-i+1,n}/X_{n-k,n}$, $i = 1,2, \dots, k$. Here $Z_i \geq 1$ for all $i = 1,2, \dots, k$.

Also note that for both definitions of excesses (additive and multiplicative) we always have $Z_1 \geq Z_2 \geq \dots \geq Z_k$.

3.1.2 The Pareto distribution as limiting distribution

In Section 2.1 (Norwegian fire insurance case study) we stated that the right tail of a heavy-tailed distribution is approximately that of a Pareto tail if the threshold is large enough. (That was the reason for examining the linearity of the Pareto plot.) We will now show this result formally.

Let X be a random variable with distribution function F, where F is a heavy-tailed distribution concentrated on $[0, \infty)$. Let Z be the multiplicative excess X/t given $X \geq t$ for a positive threshold t. Let $\overline{F}_t$ be the tail function of Z.

Recall the definition of regular variation (Definition 2.4.4.1):

$\overline{F}$ is said to be regularly varying with index $-1/\gamma, \gamma > 0$, if

$$\lim_{t \to \infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-\frac{1}{\gamma}}$$

for all $x > 0$.

Since all heavy-tailed distributions have regularly varying tail functions, we have

$$\begin{aligned} \overline{F}_t(z) &= P(X/t > z | X \geq t) = P(X > tz | X \geq t) \\ &= P(X > tz)/P(X \geq t) \quad \text{(since } t > 0 \text{ and } z \geq 1\text{)} \\ &= \overline{F}(tz)/\overline{F}(t) \to z^{-1/\gamma} \end{aligned}$$

as $t \to \infty$ for all $z > 0$ (and in particular for all $z \geq 1$), where $\gamma > 0$.

Therefore, for heavy-tailed distributions the distribution of multiplicative excesses over a threshold t tends to a Pareto distribution as t becomes large.

Note that we have plotted the (log of) ordered observations when constructing the Pareto plot, and not the (log of) multiplicative excesses. The reason for this is that as far as the linearity of the Pareto plot and its ultimate slope are concerned, constructing the plot from orderered observations is equivalent to constructing the plot from multiplicative excesses.

To understand why this is a feature of the Pareto plot, consider the following:

Suppose we have observations $x_{1,n} \leq x_{2,n} \leq \cdots \leq x_{n,n}$ from a heavy-tailed distribution with EVI $\gamma > 0$. We know from Definition 2.1.4 that a Pareto plot is a plot of

$$\left(-\log\left(1 - i/(n+1)\right); \log\left(x_{i,n}\right)\right), i = 1,2, \dots, n.$$

Consider now two consecutive points in the Pareto plot: points i and $i + 1$, where $1 < i < n$. The slope of the line connecting points i and $i + 1$ is $\Delta y / \Delta x$, where

$$\Delta x = -\log\left(1 - (i+1)/(n+1)\right) + \log\left(1 - i/(n+1)\right) \text{ and}$$

$$\Delta y = \log\left(x_{i+1,n}\right) - \log\left(x_{i,n}\right) = \log\left(x_{i+1,n}/t_i\right) - \log\left(x_{i,n}/t_i\right),$$

for any $0 < t_i < x_{i,n}$.

We have

$$\Delta y = \log(z_{i+1}) - \log(z_i),$$

where z_i and z_{i+1} are multiplicative excesses obtained when dividing observations $x_{i,n}$ and $x_{i+1,n}$, respectively, by the threshold t_i. If we choose $t_i = x_{i-1,n}$, then t_i increases as i increases.

The fact that Δx and Δy (from point i to $i + 1$ on the plot) are the same for a Pareto plot based on observations and a Pareto plot based on multiplicative excesses, implies that the resulting Pareto plots are the same, except for the starting point. The Pareto plots are therefore identical, except that they differ by a constant (a vertical shift on the graph).

In summary, as i increases

- we move from left to right in the Pareto plot,
- the size of the observations increases,
- the threshold t_i increases (since we chose $t_i = x_{i-1,n}$),
- the distribution of the multiplicative excesses tends to a $Pa(1/\gamma)$ distribution (as we saw in this section),
- the slope of the line tends to γ (by definition of the Pareto plot) and
- the Pareto plot becomes more and more linear (by definition of the Pareto plot).

The fact that the limiting distribution of the multiplicative excesses is known, can be exploited to obtain an estimator of the EVI, as will be seen in the following section.

3.1.3 The Hill estimator

We have seen that the limiting distribution of multiplicative excesses is Pareto with parameter $1/\gamma, \gamma > 0$.

The most natural way to estimate the EVI, given n observations from a heavy-tailed distribution and a threshold t (or number of excesses k), is to fit the Pareto distribution to the multiplicative excesses $z_1, z_2, \ldots, z_k$ by means of maximum likelihood estimation. It is easy to show that this yields

$$\hat{\gamma} = \frac{1}{k} \sum_{i=1}^{k} \log z_i.$$

This leads us to the definition of the most famous estimator of the EVI, namely the Hill estimator (Hill, 1975).

Definition 3.1.3 The Hill estimator

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent, identically distributed random variables with common Pareto type distribution function F concentrated on $[0, \infty)$. Let k $(2 \leq k \leq n-1)$ denote the number of excesses, and let $Z_i = X_{n-i+1,n}/X_{n-k,n}$, $i = 1, 2, \ldots, k$, be the multiplicative excesses.

The Hill estimator of the (positive) EVI, denoted by $H_{k,n}$, is defined as

$$H_{k,n} = \frac{1}{k}\sum_{i=1}^{k} \log Z_i.$$

It is easy to show that the following is an equivalent definition for the Hill estimator:

$$H_{k,n} = \tfrac{1}{k}\textstyle\sum_{i=1}^{k} i\log\left(\frac{X_{n-i+1,n}}{X_{n-i,n}}\right).$$

This formula is computationally more convenient.

We illustrate some properties of the Hill estimator by obtaining the Hill estimate of the EVI at different thresholds, using 200 observations simulated from a *Burr*$(\beta = 1, \gamma = 0.5, \rho = -0.7)$ distribution. Consider the following resulting graph:

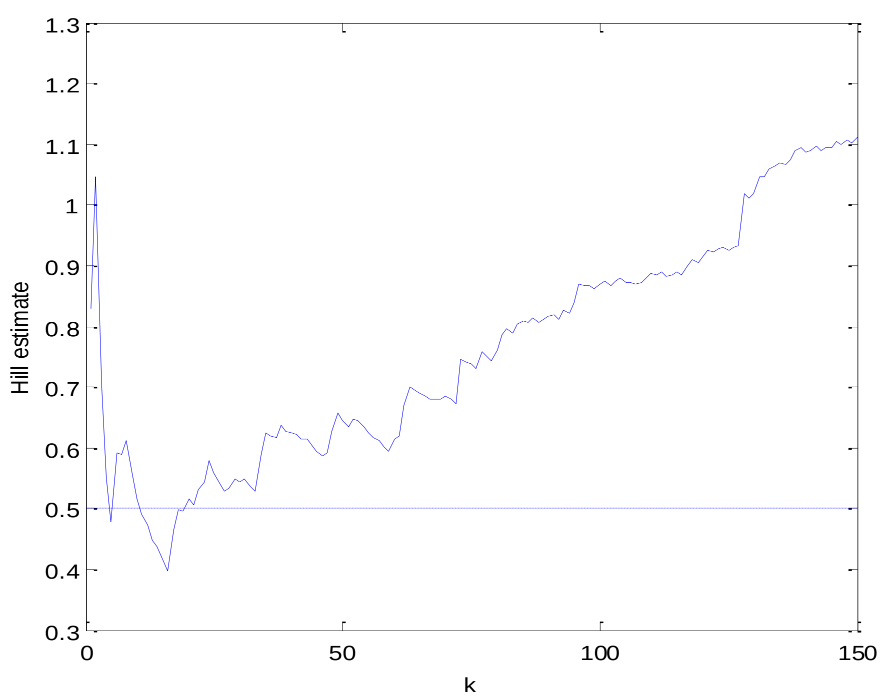

Figure 3.1.3 Hill plot for 200 observations from a Burr sample.

The horizontal dashed line indicates the true value of the EVI, namely 0.5.

The following are general properties of the Hill estimator, which are also evident from this example:

- For small k the estimates have low bias, but large variance. This is because the threshold is large enough for the first order approximation (Pareto distribution) to be valid, but the number of observations k on which the estimate is based, is small.
- For large k we have the reverse of this situation. The variance is small, since we have a large number of excesses, but since the threshold is small, the Pareto approximation is unsatisfactory, leading to a large bias.

This *variance-bias trade-off* occurs frequently in EVT when a choice of threshold has to be made. Methods of choosing the threshold for the Hill estimator will be covered in Chapter 4.

3.1.4 The perturbed Pareto distribution

In Section 3.1.3 we have considered *first order estimation*, since the Hill estimator was derived from the assumption of first order regular variation of the tail function of the underlying distribution.

The Hill plot of the Burr sample (Figure 3.1.3) showed that the Hill estimator suffers from severe bias. This is a general drawback of the Hill estimator and not unique to the sample we considered. The bias is significant even for a small value of k (number of excesses). The reason is that the Pareto approximation (and first order regular variation) holds when the threshold t is large. The fit becomes inadequate the moment the threshold becomes small enough to include more than just the very few largest observations.

We will now attempt to reduce this bias by fitting a distribution which assumes second order regular variation of the tail function of the underlying distribution. Second order variation also requires that $t \to \infty$, but the threshold may be smaller for the fit to be satisfactory. Hence, we can include more excesses when fitting the distribution.

Second order regular variation is a generalization of first order regular variation, as can be seen from Definitions 2.4.4.1 and 2.4.4.2. The progression from first to second order regular variation is similar to what happens when an extra term is added in a Taylor expansion in order for the resulting polynomial to give a better approximation of a function.

For second order estimation we have the same setting as in Section 3.1.2:

Let X be a random variable with distribution function F, where F is a heavy-tailed distribution concentrated on $[0, \infty)$. Let Z be the multiplicative excess X/t given $X \geq t$ for a positive threshold t. Let $\overline{F}_t$ be the tail function of Z.

We have also seen in Section 3.1.2 that $\overline{F}_t(z) = \overline{F}(tz)/\overline{F}(t)$.

Instead of the first order regular variation property of heavy-tailed distributions, we now use the fact that all heavy-tailed distributions of interest to us are second order regularly varying.

From the definition of second order regular variation (Definition 2.4.4.2) the following holds for the case $\rho < 0$:

$$\lim_{t\to\infty} \left(\frac{\overline{F}(tz)}{\overline{F}(t)} - z^{-1/\gamma}\right)(A(t))^{-1} = dz^{-1/\gamma}\frac{z^{\rho/\gamma}-1}{\rho/\gamma}$$

for all $z > 0$, with $-\infty < d < \infty$.

Note the following:

- We only consider the case where $\rho < 0$, since we are using the distribution which results from this assumption for parameter estimation. If we encounter the situation where $\rho = 0$ for the underlying distribution, this will be indicated by the estimate of ρ being close to zero.
- The result holds for $z > 0$ and therefore also for $z \geq 1$ (since z is a multiplicative excess) in our case.

The definition above implies that for large t

$$\overline{F}_t(z) = \overline{F}(zt)/\overline{F}(t) \approx z^{-1/\gamma} + \frac{\gamma d}{\rho} z^{-1/\gamma}(z^{\rho/\gamma} - 1)A(t).$$

Setting $c = \gamma dA(t)/\rho$, we have

$$\overline{F}_t(z) \approx (1-c)\, z^{-1/\gamma} + \ cz^{-(1-\rho)/\gamma}.$$

The survival function of the multiplicative excesses tends to $(1-c)z^{-1/\gamma} + cz^{-(1-\rho)/\gamma}$ as the threshold becomes large, with resulting density function $\gamma^{-1}z^{-1/\gamma-1}(1-c+c(1-\rho)z^{\rho/\gamma})$. This limiting distribution is called the *perturbed Pareto distribution (PPD).*

In order for the density function to be nonnegative for all $z \geq 1$, it is a straightforward algebraic exercise to show that we require $\rho^{-1} \leq c \leq 1$.

We now state the definition of the PPD formally:

Definition 3.1.4 The perturbed Pareto distribution (PPD)

A random variable X is said to be distributed perturbed Pareto with parameters γ, ρ and c, *denoted* $X \sim PPD(\gamma, \rho, c)$, *when*

$$\bar{F}(x) = (1-c)x^{-1/\gamma} + cx^{-(1-\rho)/\gamma} \textit{ and}$$

$$f(x) = \gamma^{-1}x^{-1/\gamma-1}(1-c+c(1-\rho)x^{\rho/\gamma}),$$

where $x \geq 1, \gamma > 0$, $\rho < 0$ *and* $\rho^{-1} \leq c \leq 1$.

One can see that the Pareto distribution is a special case of the PPD when $c = 0$.

The PPD is also defined in Beirlant et al. (2004), but with a slightly different parameterization. The survival function is given by

$$\bar{F}_1(x) = (1-c)x^{-1/\gamma} + cx^{-1/\gamma-\tau}.$$

Feuerverger and Hall (1999) state the PPD in its full generality as a mixture of Pareto distributions, with survival function

$$\bar{F}_2(x) = C_1 x^{\alpha_1} + C_2 x^{\alpha_2}.$$

Beirlant, Joossens and Segers (2009) propose the extended Pareto distribution (EPD), with survival function

$$\bar{F}_3(x) = \{x(1+\delta-\delta x^{\tau})\}^{-1/\gamma}.$$

This can be rewritten as

$$\bar{F}_3(x) = x^{-1/\gamma}(1+\delta)^{-1/\gamma}\left\{1 - \frac{\delta}{1+\delta}x^{\tau}\right\}^{-1/\gamma}.$$

Since $\tau < 0$, we have for large values of x

$$\bar{F}_3(x) \approx x^{-1/\gamma}(1+\delta)^{-1/\gamma}\left\{1 + \left(\frac{1}{\gamma}\right)\frac{\delta}{1+\delta}x^{\tau}\right\} = (1+\delta)^{-1/\gamma}x^{-1/\gamma} + \frac{\delta(1+\delta)^{-1/\gamma-1}}{\gamma}x^{-1/\gamma+\tau},$$

which corresponds to the PPD model with survival function $\bar{F}_2(x)$, as defined by Feuerverger and Hall (1999).

In this thesis the parameterization of the PPD which will be used will be as specified by Definition 3.1.4.

3.1.5 A second order generalization of the Hill estimator

In Sections 3.1.2 and 3.1.3 we assumed first order regular variation which resulted in the Pareto as limiting distribution of the multiplicative excesses. The EVI is a parameter of the Pareto and we estimated the EVI using maximum likelihood estimation, resulting in the Hill estimator.

In Section 3.1.4 we assumed second order regular variation which resulted in the PPD as limiting distribution of the multiplicative excesses. The EVI is a parameter of the PPD and we can estimate the EVI using MLE, resulting in a second order generalization of the Hill estimator.

We will now investigate whether or not this *second order estimator* improves on the Hill estimator by reducing the bias problem. Using the same data set as the one used in Section 3.1.3 (200 observations from a *Burr*$(\beta = 1, \gamma = 0.5, \rho = -0.7)$ distribution), the following graph was obtained by determining MLEs of γ as parameter of the PPD:

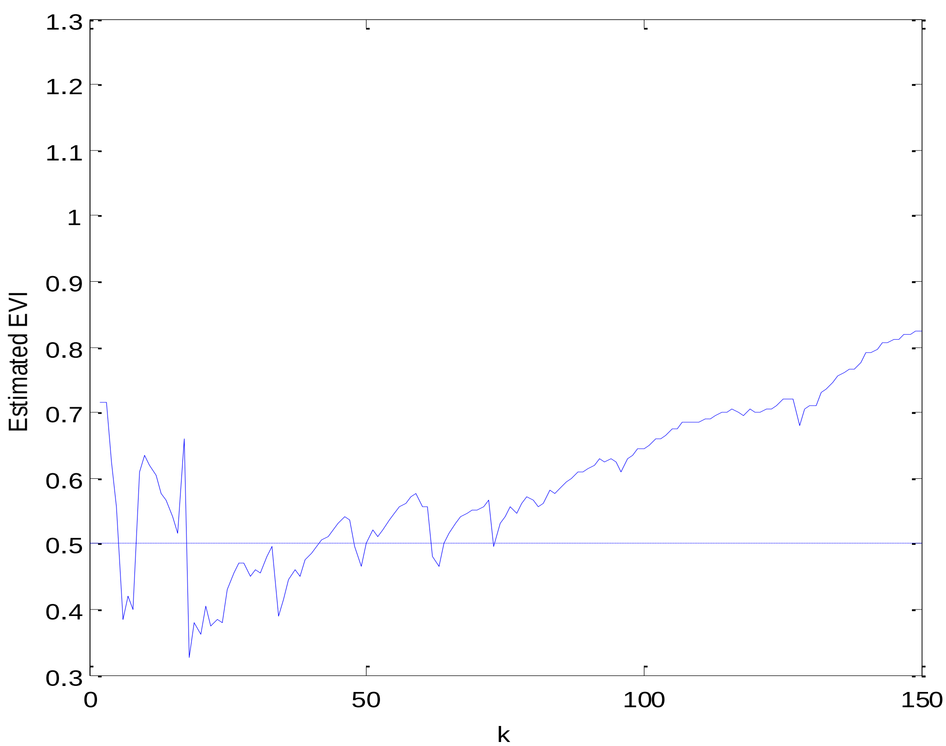

Figure 3.1.5 MLEs of the EVI for the Burr sample by fitting the PPD.

As can be seen from the graph (and also confirmed in subsequent chapters), fitting the PPD instead of the Pareto substantially improves estimation. A larger number of excesses can be used when fitting the PPD, since for large values of k the estimates still have low bias. The choice of threshold is also not as critical, since for a wide range of values the estimates stay relatively close to the true value of the EVI.

To obtain the above graph the MLEs of the parameters γ and c of the PPD were determined. The parameter ρ was not estimated, but kept constant at a value of –1. (See Section 4.2 for more detail on MLE of PPD parameters.)

Using MLE and fixing ρ at –1 is only one possible method of estimating the EVI as parameter of the PPD. A variety of estimation methods exist. In particular, the following aspects need to be taken into account:

- **Estimation of the second order parameter**

It is a well known fact that it is difficult to estimate the second order parameter ρ accurately. One can see this when considering for example the likelihood as a function of ρ for given values of γ and c. The result is usually a very flat likelihood function. Small changes in the parameters γ and c cause the value of ρ, at which a maximum is reached, to change dramatically.

Some authors suggest that ρ should not be estimated at all, but that ρ should be fixed at some value, usually -1. See for instance Beirlant et al. (2004) and Dierckx (2000).

Apart from keeping ρ fixed at -1, we also investigate two other alternatives. The first is to estimate ρ, but to restrict its range to $-2 \leq \rho < 0$. This choice of range will be justified in Section 4.9.3. The second alternative is to estimate ρ seperately by a method suggested in the literature, and then to estimate γ and c, given this value of ρ.

- **Choice of threshold**

As was seen in Figure 3.1.5, the estimation of the EVI by fitting the PPD is quite robust to changes in the threshold. However, if one has to test the performance of estimators by means of simulation as is done subsequently, a choice of threshold has to be made. Some methods of threshold selection will be discussed in Section 3.4.

- **Bayesian estimation vs. MLE**

Only Bayesian estimation and MLE will be considered. Other alternatives, for example the method of moments, will not be investigated. Since the PPD estimator which we will propose as our benchmark estimator is based on Bayesian estimation of the EVI, we devote the next section in its entirety to the subject of Bayesian estimation of the PPD parameters.

3.2 Bayesian estimation of PPD parameters

This section will describe how to obtain Bayesian estimates of the PPD parameters. We use Gibbs sampling to this end, the details of which will be described in this section.

Since this section has a quite complicated hierarchical structure, we present the following scheme to show where topics fit in in relation to each other:

3.2.1 Overview of basic Bayesian theory.

3.2.2 Choice of prior and resulting posterior (of the PPD).

3.2.3 Gibbs sampling (of PPD parameters).

Specific subsections deal with the following subjects:

- Choosing initial values for the parameters.
- Choosing the number of Gibbs repetitions (sizes of the Gibbs samples).
- The rejection method.
- Applying the restriction $1/\rho \leq c \leq 1$.
- Approximations to speed up the Gibbs sampler.

3.2.4 Using Gibbs samples for the purpose of making inferences.

Specific subsections deal with the following subjects:

- Loss functions leading to estimates being either the mean, median or mode of the draws.
- The half-sample mode.
- Highest posterior density (hpd) region.

3.2.5 Quantile estimation.

3.2.6 Estimating exceedance probability.

3.2.1 Overview of basic Bayesian theory

Let $\underline{x} = (x_1, x_2, \ldots, x_n)$ be a vector of observed data of a random variable X assumed to be distributed according to a distribution with density function $f(x|\underline{\theta})$, where $\underline{\theta}$ is a vector of unkown parameters in the parameter space Ω.

We assume that we have *a priori* knowledge regarding the distribution of the parameter vector $\underline{\theta}$. This is formalized by specifying a prior density function $\pi(\underline{\theta})$. The choice of prior will be discussed in Section 3.2.2. Let $f(\underline{x}|\underline{\theta})$ denote the likelihood function of $\underline{\theta}$:

$$f(\underline{x}|\underline{\theta}) = \prod_{i=1}^{n} f(x_i|\underline{\theta}).$$

According to Bayes' theorem the distribution of $\underline{\theta}|\underline{x}$, called the posterior distribution of $\underline{\theta}$, is given by

$$\pi(\underline{\theta}|\underline{x}) = \frac{\pi(\underline{\theta})f(\underline{x}|\underline{\theta})}{\int_{\Omega}\pi(\underline{\theta})f(\underline{x}|\underline{\theta})d\underline{\theta}} \propto \pi(\underline{\theta})f(\underline{x}|\underline{\theta}).$$

The marginal posterior distribution of parameter θ_i in the parameter vector $\underline{\theta}$ is obtained by intergrating out the remaining parameters. If $\underline{\theta}$ is p dimensional, we have

$$\pi_i(\theta_i|\underline{x}) = \frac{\int_{\theta_1}\dots\int_{\theta_{i-1}}\int_{\theta_{i+1}}\dots\int_{\theta_p}\pi(\underline{\theta})f(\underline{x}|\underline{\theta})d\theta_p\dots d\theta_{i+1}d\theta_{i-1}\dots d\theta_1}{\int_{\Omega}\pi(\underline{\theta})f(\underline{x}|\underline{\theta})d\underline{\theta}}.$$

It is very rarely possible to obtain any of these integrals analytically. Markov Chain Monte Carlo (MCMC) methods can be employed to estimate this marginal posterior distribution. One of these methods, namely Gibbs sampling, will be discussed in Section 3.2.3 for the case of the PPD.

Assuming quadratic (square error) loss, the Bayesian estimate of θ_i is the expected value of its marginal posterior distribution $\pi_i(\theta|\underline{x})$:

$$\hat{\theta}_i = \int \theta\pi_i(\theta|\underline{x})d\theta.$$

Assuming absolute error loss, the Bayesian estimate of θ_i is the median of the marginal posterior distribution; assuming 0-1 loss, the Bayesian estimate of θ_i is the mode of the marginal posterior distribution. More detail on loss functions and how to estimate the mean, median and mode of the marginal posteriors will be given in Section 3.2.4.

For the purpose of inference one constructs an interval which is analogous to a confidence interval. The *highest posterior density* (hpd) *region* is such an interval. The construction and interpretation of the hpd region will be covered in Section 3.2.4.

3.2.2 Choice of prior and resulting posterior

Two main categories of priors exist. The first is the subjective prior. The way in which the prior is selected reflects the subjective opinion of the analyst concerning the distribution of the parameter vector $\underline{\theta}$, prior to observing the data. Usually an expert in the field under investigation is consulted and the prior density is constructed according to the experience of the expert based on similar studies in the past.

Since subjectivity can vary from one expert to another and since it is often difficult to justify the choice of a specific prior, objective priors have been introduced. These priors assume no subjective prior information for a given situation and are derived from the assumed probability density function of the data $f(x|\underline{\theta})$. Various objective priors exist, for example Jeffreys' prior and the reference prior.

In this thesis we will consider Zellner's maximal data information (MDI) prior. The MDI prior is an objective prior which provides maximal average data information on the vector of parameters $\underline{\theta}$. This prior is often easy to derive, and is also used in many studies, especially in the field of econometrics.

The MDI is defined as follows (Zellner, 1971 and Beirlant et al., 2004):

Definition 3.2.2 The maximal data information (MDI) prior

If random variable X has density function $f(x|\underline{\theta})$, the MDI prior of the parameter vector $\underline{\theta}$ is defined as

$$\pi(\underline{\theta}) \propto \exp\left(E\left(\log f(X|\underline{\theta})\right)\right).$$

For the Pareto distribution it is easy to show that this yields $\pi(\gamma) \propto (1/\gamma)e^{-\gamma}$.

For the PPD the MDI cannot be obtained in closed form. However, the PPD is a second order generalization of the Pareto distribution. We will therefore use

$$\pi(\gamma, \rho, c) \propto \frac{1}{\gamma} e^{-\gamma}$$

as prior for the PPD, which is an approximation of the MDI prior for the PPD. Despite the simplification of moving from a three parameter to a one parameter prior, the resulting prior still works well in practice, as we will see later on.

The resulting posterior, given observations $z_1, z_2, \ldots, z_k$ assumed to be from a PPD, is

$$\pi(\gamma, \rho, c|z_1, z_2, \ldots, z_k) \propto \frac{1}{\gamma} e^{-\gamma} \prod_{i=1}^{k} \gamma^{-1} z_i^{-1/\gamma-1}(1 - c + c(1-\rho) z_i^{\rho/\gamma}).$$

3.2.3 Gibbs sampling

The MCMC method we will employ in this thesis is Gibbs sampling, the details of which will now be discussed for the estimation of the PPD parameters. Gibbs sampling is well-known and frequently employed for obtaining Bayesian estimates of parameters.

The setting is the following: Given observations (multiplicative excesses) $z_1, z_2, \ldots, z_k$, one wishes to fit a PPD distribution to these observations by determining the Bayesian estimates of the PPD parameters γ, ρ and c.

The Gibbs sampling procedure will now be given, followed by a detailed discussion of each step:

1. Start with initial estimates of the parameter vector: $\gamma^{(0)}, \rho^{(0)}$ and $c^{(0)}$.

2. For a large number of repetitions $i = 1$ to N:

2.1 Generate $c^{(i)}$, which is one simulated value (one draw) of c from the conditional posterior of c, given $\gamma^{(i-1)}$ and $\rho^{(i-1)}$:

$$\pi\left(c \middle| \gamma^{(i-1)}, \rho^{(i-1)}, \underline{z}\right) \propto \prod_{j=1}^{k} (1 - c + c(1 - \rho^{(i-1)}) z_j^{\rho^{(i-1)}/\gamma^{(i-1)}}).$$

Note that one can omit the parts of the conditional posterior which are constant with respect to c, namely $\frac{1}{\gamma^{(i-1)}} e^{-\gamma^{(i-1)}}$ and $\prod_{j=1}^{k} (\gamma^{(i-1)})^{-1} z_j^{-1/\gamma^{(i-1)}-1}$. The same idea applies to Step 2.3. Also note that we apply the restriction $1/\rho^{(i-1)} \leq c \leq 1$. See Definition 3.1.4.

2.2 Generate $\gamma^{(i)}$, which is one simulated value (one draw) of γ from the conditional posterior of γ, given $\rho^{(i-1)}$ and $c^{(i)}$:

$$\pi(\gamma | \rho^{(i-1)}, c^{(i)}, \underline{z}) \propto \frac{1}{\gamma} e^{-\gamma} \prod_{j=1}^{k} \gamma^{-1} z_j^{-1/\gamma - 1} (1 - c^{(i)} + c^{(i)}(1 - \rho^{(i-1)}) z_j^{\rho^{(i-1)}/\gamma}).$$

2.3 Generate $\rho^{(i)}$, which is one simulated value (one draw) of ρ from the conditional posterior of ρ, given $\gamma^{(i)}$ and $c^{(i)}$:

$$\pi(\rho | \gamma^{(i)}, c^{(i)}, \underline{z}) \propto \prod_{j=1}^{k} (1 - c^{(i)} + c^{(i)}(1 - \rho) z_j^{\rho/\gamma^{(i)}}).$$

The restriction $-2 \leq \rho < 0$ is applied. The justification for this range is given in Section 4.9.3.

3. The resulting vectors of draws are $\underline{\gamma} = \left(\gamma^{(1)}, \gamma^{(2)}, \dots, \gamma^{(N)}\right)$, $\underline{\rho} = \left(\rho^{(1)}, \rho^{(2)}, \dots, \rho^{(N)}\right)$ and $\underline{c} = \left(c^{(1)}, c^{(2)}, \dots, c^{(N)}\right)$. These draws can be regarded as simulated values from the marginal posterior distributions of γ, ρ and c, respectively. How to estimate the parameters from these drawn values will be discussed in Section 3.2.4.

We now describe each of the steps above in more detail.

Step 1

Since our first draw is a value of c, we actually do not need to choose $c^{(0)}$. We choose to start with a draw of c, since it is not clear what an initial estimate of c should be.

As initial value of γ, we choose $\gamma^{(0)}$ as the Hill estimate.

As initial value of ρ, we choose $\rho^{(0)}$ as -1, which is the midpoint of the interval $[-2,0)$ on which we restrict ρ in Step 2.3.

Step 2: Choosing N

We will now discuss the choice of the number of repetitions N.

Due to time contraints, especially when doing a simulation study, it is not always possible to draw large posterior samples. In this thesis we restrict the number of draws to 1000 when doing simulation studies. If possible, from a computation time perspective, N should be chosen considerably larger, namely $N = 10\,000$, or even $N = 100\,000$. In this thesis the number of repetitions N used will always be stated explicitly.

Another choice which has to be made when doing Gibbs sampling, is the number of burn-in draws. Since one starts with arbitrary (and maybe inaccurate) initial values of the parameters of the distribution, it is suggested that the first part of the marginal posterior draws should be disregarded in order to compensate for this. The initial draws that are left out from the final posterior samples, are referred to as the burn-in stage of Gibbs sampling. We will use 10 burn-in draws when doing simulation studies (followed by 1000 draws which are retained).

Step 2.2: The rejection method

We will now illustrate how Step 2.2 is performed. Steps 2.1 and 2.3 are performed in a similar fashion. The method used in this thesis to simulate one value from the given conditional posterior density, is the *rejection method*, which will be described here (Rice, 2006).

Suppose we need to simulate values from a given density function $f(x)$. For the purpose of illustration, we represent the density function graphically:

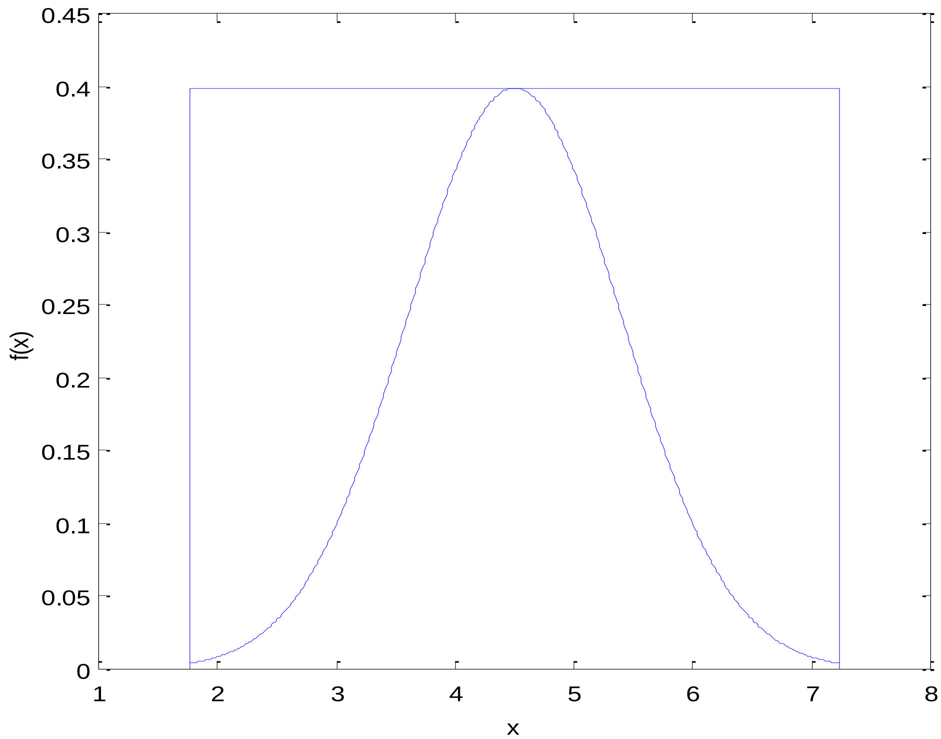

Figure 3.2.3.1 Graphical representation of density to simulate from.

In order to apply the rejection method, we need to construct the rectangle shown in the graph. In other words, we need to know the maximum of the density function, and the interval in which the density is "significant". In the above graph we defined the "significant" interval as all values of x for which $f(x) > 0.01\max(f(x))$. The significance factor of 0.01 is also the value used in the MATLAB programs. The resultant interval in the above case is $[x_{\min} = 1.77, x_{\max} = 7.23]$, and $\max(f(x)) = 0.3989$.

For every value that has to be simulted from the density $f(x)$, the following steps are performed:

a) Simulate a value x^* from a $U(x_{\min}, x_{\max})$ distribution.

b) Independent of x^*, simulate a value y^* from a $U(0, \max(f(x)))$ distribution.

c) If $y^* \leq f(x^*)$, take x^* as the simulated value. Otherwise reject x^* and redo from Step a).

On a more technical point, one does not have to know the maximum precisely, but any value larger than the maximum. This value must however be close to the maximum of $f(x)$, otherwise the proportion of rejected values will be large, rendering the algorithm inefficient. Also, ideally one should choose the significance factor as small as possible. However, if the density is skewed, a value like 0.0001 will cause a very large section of the tail to be included in the rectangle. To find the rectangle will require a large number of calculations and again the proportion of rejected values will be large, leading to inefficiency of the algorithm.

An important advantage of the rejection method is that the normalizing constant of the density need not be known. For instance, if we graph $af(x)$ instead of $f(x)$, where $a = 1/\max(f(x))$, we obtain the following graph:

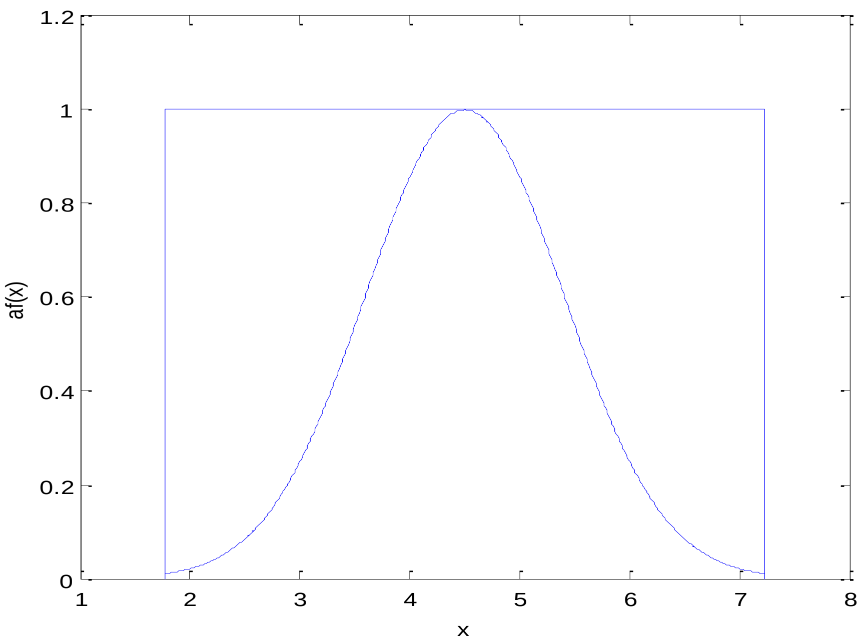

Figure 3.2.3.2 Unnormalized density from which to simulate observations.

Following Steps a) to c) above will result in exactly the same simulated values as before.

We now turn to the rejection method as applied to our specific problem.

We wish to simulate a value $\gamma^{(i)}$ from the conditional posterior

$$\pi(\gamma|\rho^{(i-1)}, c^{(i)}, \underline{z}) \propto \frac{1}{\gamma} e^{-\gamma} \prod_{j=1}^{k} \gamma^{-1} z_j^{-1/\gamma-1} (1 - c^{(i)} + c^{(i)}(1-\rho^{(i-1)}) z_j^{\rho^{(i-1)}/\gamma}).$$

For the sake of simplicty, let

$$f(\gamma) = \frac{1}{\gamma} e^{-\gamma} \prod_{j=1}^{k} \gamma^{-1} z_j^{-1/\gamma-1} (1 - c + c(1-\rho) z_j^{\rho/\gamma}).$$

Note that $f(\gamma)$ is not a density function, since it does not integrate to 1.

Since the normalizing constant is unknown, it is always advisable to first compute $\log(f(\gamma))$, then obtain another function, say $g(\gamma)$, by scaling $\log(f(\gamma))$ appropriately and then to calculate $\exp(g(\gamma))$. This is done because of numerical considerations. Even though, mathematically, working with $f(\gamma)$ and $\exp(g(\gamma))$ would yield the same result, working with $f(\gamma)$ creates some numerical problems. When the sample size k is large, one can see that for only a very specific region of values of γ we find that $f(\gamma)$ is not either very large or very close to zero. Calculation of $f(\gamma)$ yields numerically either zero or infinity if γ is outside this region.

We solve this problem by using

$$\log\big(f(\gamma)\big) = -(k+1)\log(\gamma) - \gamma - \left(1 + \frac{1}{\gamma}\right)\sum_{i=1}^{k} \log z_i + \sum_{i=1}^{k} \log\big(1 - c + c(1-\rho) z_i^{\rho/\gamma}\big).$$

The procedure is as follows:

a) Choose a range of values of γ, say $\underline{\gamma} = (0.01, 0.02, \ldots, 2)$.
b) Calculate $\log f\left(\underline{\gamma}\right) = (\log f(0.01), \log f(0.02), \ldots, \log f(2))$.
c) Calculate $g\left(\underline{\gamma}\right) = \log f(\underline{\gamma}) - \max(\log f(\underline{\gamma}))$.
d) Calculate $\exp(g(\underline{\gamma}))$. The result is a vector of "density values" at points in $\underline{\gamma}$, rescaled so that the maximum value is 1.
e) Retain those values of $\exp(g(\underline{\gamma}))$ which are greater than 0.01 and the corresponding values of $\underline{\gamma}$.
f) Use the result in e) and apply the rejection method to simulate a single observation from $\pi(\gamma|\rho, c, \underline{z})$.

In Step f) above, it is much faster to divide the range of γ as obtained in Step e) into discrete intervals. For example, if $\exp(g(\underline{\gamma}))$ is greater than 0.01 in the points $(0.32, 0.33, \ldots, 0.87)$, we simulate x^* from the points $(0.32, 0.33, \ldots, 0.87)$, where the same probability is assigned to each of the points. Independent of x^*, we then simulate a value y^* from a $U(0,1)$ distribution (since $\max\left(\exp(g(\underline{\gamma}))\right) = 1$).

If we follow this approach, we must take care that the increments in Step a) are small enough. However, using very small increments are too time consuming. The algorithm as it is given above is quite inefficient as far as computation time is concerned.

The procedure we applied to improve efficiency, made use of two iterations rather than one. We choose a "rough" range of values for γ, divide this range into 20 intervals and perform Steps b) to e) above. We then take the minimum and maximum of the values of γ obtained in Step e) and regard these values as the lower and upper limit of the new updated range of γ, respectively. We then divide this updated range into 20 intervals, and perform Steps b) to e) again.

To futher increase efficiency, we base our initial “rough” range from which we start, on the updated range of the previous Gibbs repetition, i.e. the updated range of the previous draw of γ.

See also the subsection entitled *Approximations to speed up the Gibbs sampler* (after Step 2.3 below).

Step 2.3

This step falls away if the value of ρ is given. In Section 3.1.5 we discussed the estimation of the second order parameter ρ and stated that we will investigate the following alternatives:

- estimating ρ, but restricting its range to $-2 \leq \rho < 0$,
- keeping ρ fixed at -1, and
- estimating ρ separately and then, given this value of ρ, estimating γ and c.

In this section we have assumed up to now that we are dealing with the first alternative. (A justification for this range of ρ will be given in Section 4.9.3.) If we investigate the second and third alternatives, Step 2.3 falls away completely and in all formulas ρ will be treated as a given constant.

The second issue that requires some explanation, is the implementation of the restriction $-2 \leq \rho < 0$. One notes that the restriction $1/\rho \leq c \leq 1$ was applied in Step 2.1. One may ask whether this should not be an additional restriction in Step 2.3, again to ensure that $(1-c)z^{-1/\gamma} + cz^{-(1-\rho)/\gamma}$ is a distribution function.

In order to answer this question, let us see what happens if we limit both ρ and c on each draw in this way. In other words we wish to see what happens if we apply the restrictions $\rho \leq 1/c$ (if $c > 0$) or $\rho \geq 1/c$ (if $c < 0$) and $\rho < 0$ to ρ, and the restriction $1/\rho \leq c \leq 1$ to c.

The simulated value of c is almost always negative, which means the restriction $\rho \geq 1/c$ restricts ρ to values close to 0. With ρ close to 0, the restriction $1/\rho \leq c \leq 1$ leaves c free to assume very large negative values. For such c, $1/c \approx 0$. Then $\rho \geq 1/c$ implies that $\rho \approx 0$. We see that the simulated value of ρ converges to 0 almost immediately. Hence, the method of restricting both ρ and c in this way, does not work.

We therefore apply the restriction $1/\rho \leq c \leq 1$ only to c, since we have already restricted ρ to the range $-2 \leq \rho < 0$.

A partial additional restriction is applied, however. If a situation occurs in Step 2.3 where the PPD density function is negative, the density function value in that point is set to zero. This results in the likelihood, and consequently the conditional posterior of ρ, to be zero for that value of ρ. Therefore such a value of ρ cannot be drawn when applying the rejection method.

The method of restricting ρ and c in the way we have just described, works very well in practice, as will be seen from the simulation results in this and subsequent chapters.

As far as programming the Gibbs sampler is concerned, we restricted ρ to $\rho \leq -0.01$ for numerical reasons. This restriction is necessary when, for instance, we need to calculate the valid range of c, which has a lower limit of $1/\rho$. The restriction $\rho \leq -0.01$ is applied in all MATLAB programs.

Approximations to speed up the Gibbs sampler

The Gibbs sampler described above requires a huge amount of computation time. In this subsection we describe a technique developed in order to shorten computation time by approximating the conditional posterior.

As mentioned above, the quantity

$$-(k+1)\log(\gamma) - \gamma - (1+\tfrac{1}{\gamma})\sum_{i=1}^{k} \log z_i + \sum_{i=1}^{k} \log\left(1 - c + c(1-\rho) z_i^{\rho/\gamma}\right)$$

needs to be calculated when applying the Gibbs sampler. The part of the formula which is computationally intensive is

$$\sum_{i=1}^{k} \log\left(1 - c + c(1-\rho) z_i^{\rho/\gamma}\right).$$

We accelerate the Gibbs sampler by approximating this quantity when $k \geq 200$.

Let

$$h(z_i) = \log\left(1 - c + c(1-\rho) z_i^{\rho/\gamma}\right).$$

As an example of how the approximation is done, consider a sample of size $n = 10\,000$ from a *Burr*$(\beta = 1, \gamma = 0.5, \rho = -1)$ distribution. Suppose we choose $k = 2500$. We now plot $h(z_i)$ for $i = 1, 2, \ldots, 2500$, with $\gamma = 0.7$, $\rho = -0.1$ and $c = -2$:

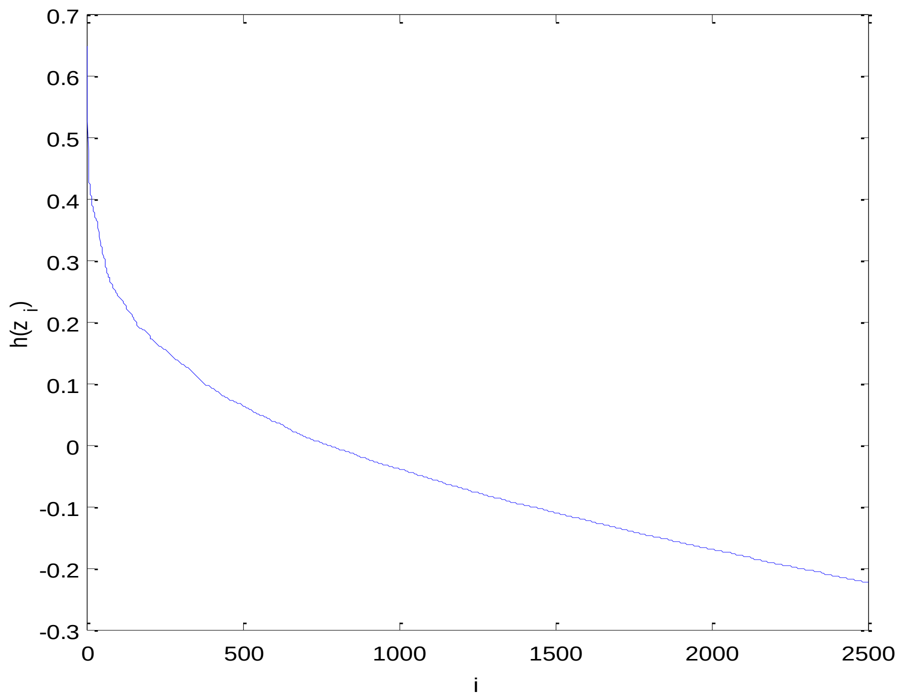

Figure 3.2.3.3 Plot of $h(z_i)$ for $i = 1, 2, \ldots, 2500$.

We will partition the x-axis and calculate $h(z_i)$ for only a few values of i, approximating the rest of the values by means of interpolation.

The shape of the graph in Figure 3.2.3.3 is typical of a graph of $h(z_i)$, in the sense that the slope changes more rapidly for small values of i. To understand why this is the case, consider the multiplicative excesses $z_1 \geq z_2 \geq \ldots \geq z_k$. Here z_1 is the excess associated with the largest observation, z_2 the excess associated with the second largest, etc., as shown in Section 3.1.1. Since the underlying distribution is heavy-tailed, the difference between z_1 and z_2 will generally be much greater than the difference between z_2 and z_3, etc. The change in $h(z_i)$ will therefore be more rapid for small values of i.

If we calculate $h(z_i)$ at (say) 100 values of i, we obtain more accurate results if the values of i are closer to each other for small i. The following graph shows the values of i (indicated by the vertical lines) at which we would typically want to calculate $h(z_i)$:

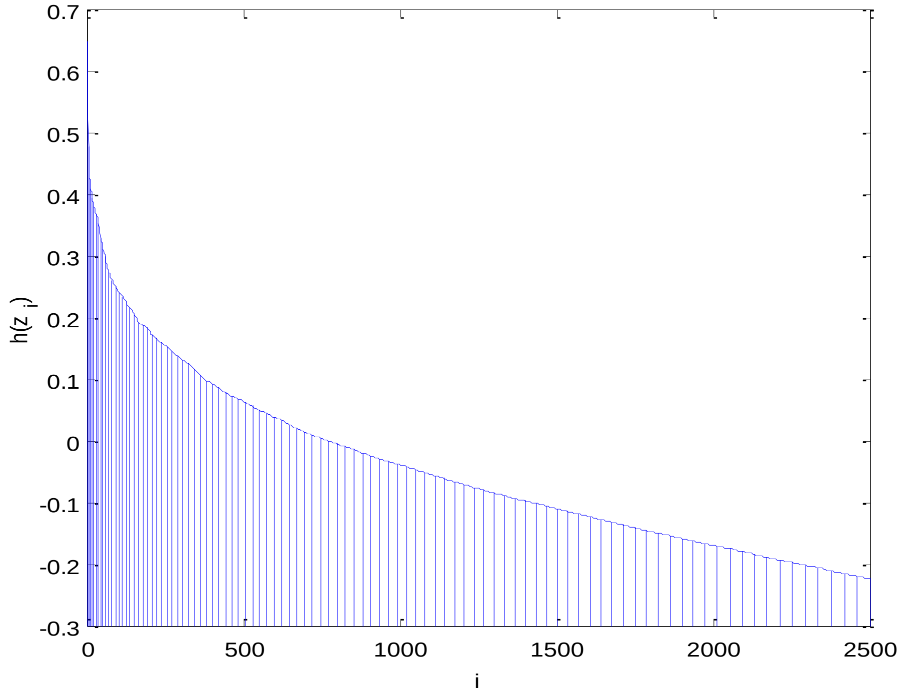

Figure 3.2.3.4 Points at which $h(z_i)$ is calculated.

The procedure we applied to choose the values of i at which to calculate $h(z_i)$, is as follows:

1. Let $\underline{j} = (1, 2, \ldots, 100)$, if one chooses to calculate $h(z_i)$ at 100 values of i.
2. Let $\log(\underline{j}) = (\log(1), \log(2), \ldots, \log(100))$.
3. We normalize this vector to range from 0 to 1. Let $\underline{p} = \log(\underline{j})/\log(100)$.
4. If we now let $\underline{i}^* = (k^{p_1}, k^{p_2}, \ldots k^{p_{100}})$, we have a vector which ranges from 1 to k.
5. Rounding these values to the nearest integer gives us the values of i at which we calculate $h(z_i)$. We denote the resulting vector by $\underline{i}$.

For $k = 2500$, this method yields $\underline{i} = (1, 3, 6, 11, 15, 21, \ldots, 2416, 2458, 2500)$, as indicated by the vertical lines in Figure 3.2.3.4.

Now that we have the values of $h(z_i)$ at the values of i in $\underline{i}$, we need to approximate $\sum_{i=1}^{k} h(z_i)$. As mentioned previously, the approximation is done by interpolation. The way to approximate $h(z_3) + h(z_4) + h(z_5) + h(z_6)$ for the case where $k = 2500$, is described here as an example. Here $h(z_3)$ and $h(z_6)$ are known. If we now assume an approximately straight line between $h(z_3)$ and $h(z_6)$,

$$h(z_3) + h(z_4) + h(z_5) + h(z_6) \approx 4[(h(z_3) + h(z_6))/2].$$

This procedure is followed to approximate the sum between every two values of $h(z_i)$ that are known.

For our sample above, the true value of $\sum_{i=1}^{2500} h(z_i)$ is -114.0375 and is approximated as -113.8564, which corresponds to an error of 0.16%.

When running the Gibbs sampler with 10 burn-in draws and 1000 draws after that for a sample of size $n = 20\,000$ ($k = 10\,000$), the approximation method yields the answer in approximately 1.1% of the time it takes without the approximation.

We have seen that the approximation works, in the sense that it speeds up the sampler significantly and the approximation error is small enough for the effect of the approximation on the Gibbs sampling to be negligible.

3.2.4 Using Gibbs samples for the purpose of making inferences

In Section 3.2.1 we mentioned that one generally assumes one of three loss functions: square error loss, absolute error loss or 0-1 loss. Under these loss functions, the resulting Bayesian estimator of the parameter is the expected value, the median and the mode of its marginal posterior distribution, respectively.

We will not go into the definition of a loss function, what considerations need to be made when choosing a loss function, how estimators are derived by minimizing the posterior risk, etc. These are topics covered in any standard text on decision theory.

Rather, we simply note that the Bayesian estimator of a parameter is generally defined in one of three ways, namely the expected value, the median or the mode of its marginal posterior distribution. The decision as to which one to use (when estimating the EVI by fitting a PPD) will be made on the basis of the relative performances of these estimators, as judged by examining the simulation results.

The question thus remains how to estimate the mean, median and mode of the marginal posterior distributions.

In Section 3.2.3 we have seen that after performing Gibbs sampling we end up with draws which can be regarded as simulated values from the marginal posterior distributions of the parameters. For the PPD these resulting vectors were $\underline{\gamma} = \left(\gamma^{(1)}, \gamma^{(2)}, \dots, \gamma^{(N)}\right)$, $\underline{\rho} = \left(\rho^{(1)}, \rho^{(2)}, \dots, \rho^{(N)}\right)$ and $\underline{c} = \left(c^{(1)}, c^{(2)}, \dots, c^{(N)}\right)$.

In order to estimate the mean of the marginal posterior distributions, we simply calculate the arithmetic mean of the marginal posterior samples (draws), yielding the following Bayesian estimates of the parameters:

$$\hat{\gamma} = \frac{1}{N}\sum_{i=1}^{N} \gamma^{(i)},\ \hat{\rho} = \frac{1}{N}\sum_{i=1}^{N} \rho^{(i)} \text{ and } \hat{c} = \frac{1}{N}\sum_{i=1}^{N} c^{(i)}.$$

In order to estimate the median of the marginal posterior distributions, we simply calculate the median of the marginal posterior samples (draws).

It is not straight forward to estimate the mode of a distribution, given a sample from that distribution. Several methods exist in the literature. The method we suggest is called half-sample mode, which we will now discuss.

The half-sample mode

We used the half-sample mode (HSM) to estimate the mode of a marginal posterior distribution. Bickel and Frühwirth (2006) considered various robust estimators of the mode and concluded that the HSM performs well under a wide range of conditions.

The calculation of the HSM involves finding the 50% highest posterior density (hpd) region of the observations, which is the shortest interval containing at least 50% of the observations. (For more detail on the hpd region, see the subsection entitled *Highest posterior density (hpd) region* below.) One then retains only the observations in the 50% hpd region and repeats the process, each time halving the number of observations, till fewer than four observations are left. Then, if there is only one observation left, the estimate of the mode is that observation. If two observations are left, the estimate of the mode is the mean of those two observations. If three observations are left, say $x_1 \leq x_2 \leq x_3$, the estimate of the mode is the mean of the two observations which are closest together. If $x_3 - x_2 = x_2 - x_1$, the estimate of the mode is x_2.

The algorithm works fine if all observations are unique, but if there is a large number of observations with the same value, the recursive procedure enters an infinite loop. As an example, consider the following six observations:

0.6	0.61	0.61	0.61	0.61	0.62

In the next step the 50% hpd region should contain at least three observations. The resulting interval is $[0.61, 0.61]$, which yields the following observations:

0.61	0.61	0.61	0.61

We now have four observations. In the next step the 50% hpd region should contain at least two observations. The resulting interval is again $[0.61, 0.61]$, which again contains four observations.

In order to overcome this problem, an insignificant amount of random noise is added to the observations, in order to ensure uniqueness. Suppose we have observations $x_1 \leq x_2 \leq \cdots \leq x_n$.

Let $m_i = \left|\frac{x_i}{20000}\right|$ denote the absolute value of the maximum amount of noise which will be added to observation x_i. If there are zero values in the data set, say $x_j = 0$, set m_j equal to the smallest nonzero value in the set $\{m_1, m_2, \ldots, m_n\}$. We generate n random numbers between -1 and 1, denoted by $r_1, r_2, \ldots, r_n$. The new set of observations are $y_1 < y_2 < \ldots < y_n$, where $y_i = x_i + r_i m_i$. Since a factor of $1/20000$ is used, x_i and y_i are equal to four significant numbers for all $i = 1,2,\ldots,n$. The HSM algorithm is then applied to $y_1 < y_2 < \ldots < y_n$ instead of $x_1 \leq x_2 \leq \ldots \leq x_n$ in order to estimate the mode.

Gibbs sampling terminology

Gibbs sampling terminology can become quite cumbersome. In the rest of the thesis we will make use of the following terminology for the sake of brevity and readability:

draws of $\sim$: The phrase *draws of* c, for instance, will mean the same as c *sample* as defined above.

mean/median/mode estimate of $\sim$: If we use the term *mode estimate of* ρ, for instance, it will refer to the estimate of ρ obtained when calculating the mode of the sample of draws from the marginal posterior distribution of ρ, as obtained by Gibbs sampling. Similarly for *mean estimate* and *median estimate*. Also note that *mode estimate* always impies that the half-sample mode was used in the calculation.

simulation study: The design and results of the Chapter 3 simulation study are presented in Section 3.3. In some sections preceding Section 3.3 we make use of some of the results of the simulation study in order, for instance, to narrow an argument down to what was found to be relevant. In such instances one can refer to Section 3.3 for more detail, but it will not be necessary to read Section 3.3 in order to understand the argument. Section 3.3 will also provide justification for considering certain distributions, sample sizes, thresholds, etc. In short, for brevity “simulation study” will be used instead of “simulation study (see Section 3.3 for more detail and justification for the choice of …)”.

Highest posterior density (hpd) region

For a specified probability $1 - \alpha$, this is the region of values that contains $100(1 - \alpha)\%$ of the probability of the marginal posterior density and also has the characteristic that the posterior density within this region is never lower than on the outside.

The definition of the hpd region can be explained using the following graph:

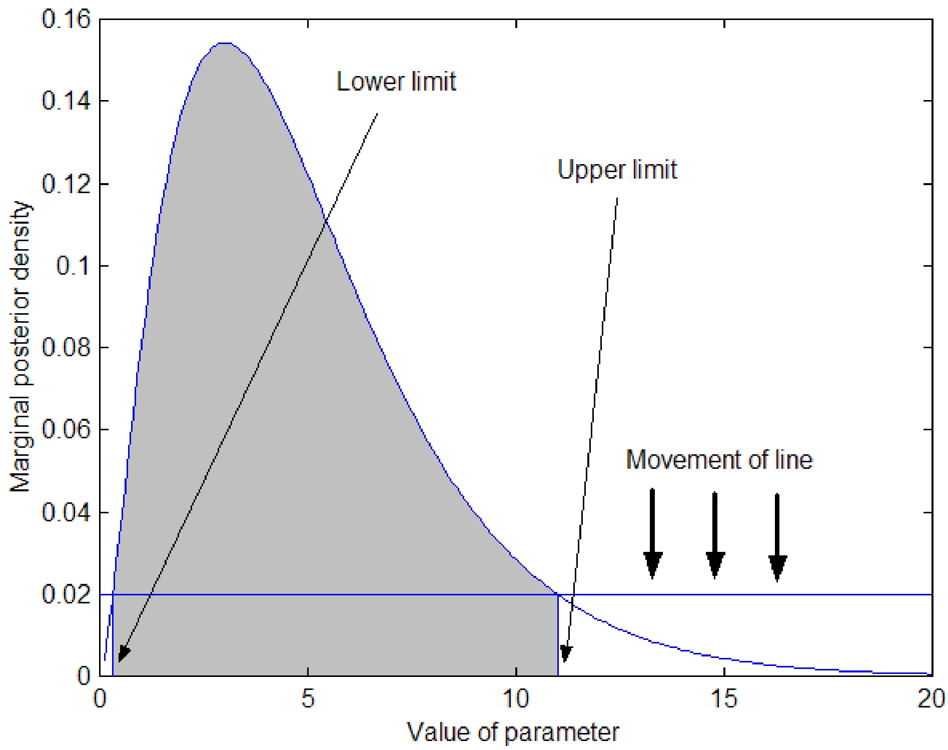

Figure 3.2.4.1 Definition of the hpd region.

Figure 3.2.4.1 shows the graphical interpretation of the definition of the hpd region. Suppose we need to obtain the 95% hpd region. We first plot the marginal posterior density function. We then draw a horizontal line through the point were this density is a maximum and move the line down until the shaded area is 95%. The values on the *x*-axis corresponding to the points were the line intersects the density function are the lower and upper limits of the hpd region, respectively.

One can also define the hpd region as the shortest interval containing at least $100(1-\alpha)\%$ of observations. Applying this alternative definition, the calculation of this interval is straight-forward.

In the remaining part of our discussion on the hpd region, we will focus on a problem which is sometimes encountered, namely that a parameter estimate can fall outside its hpd region.

As an example of a case where an estimate falls outside an hpd region, consider a sample of 1000 draws of c, where ρ was estimated as -0.01 before employing Gibbs sampling, and is considered a given constant. Since $1/\rho \leq c \leq 1$, we have $-100 \leq c \leq 1$. The following is a histogram of the draws of c:

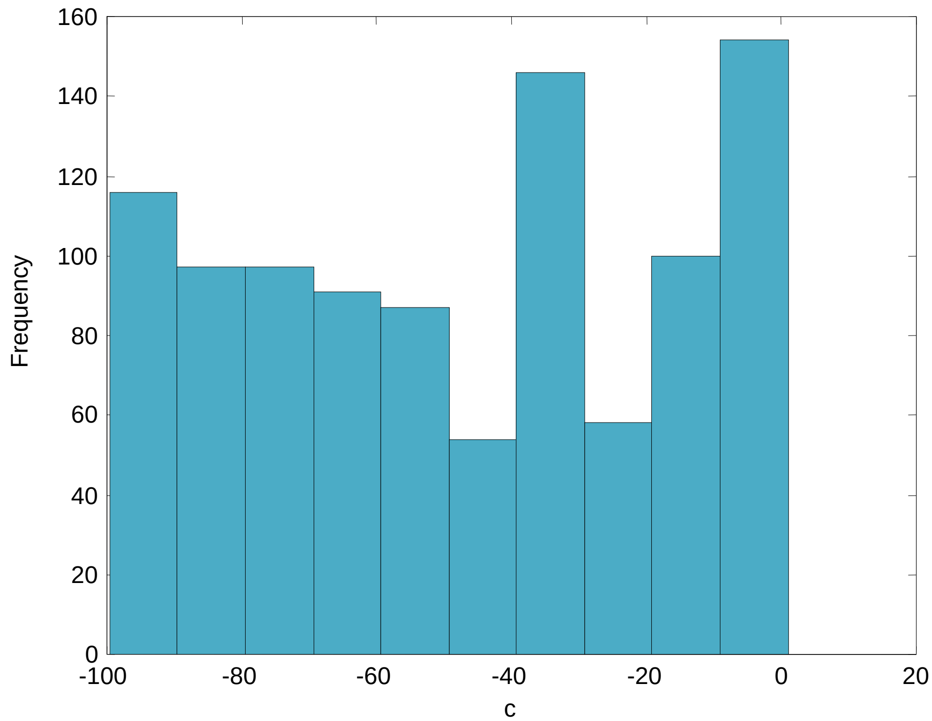

Figure 3.2.4.2 Gibbs sample of c, given $\rho = -0.01$.

The (half-sample) mode of this sample is 1 and the 90% hpd region is $(-94.57, -4.03)$.

If we decide on the mode estimate of c as our estimate, the conclusion will be that our estimate of c is 1, and that there is a 90% probability that the true value of c is in the interval $(-94.57, -4.03)$.

We now consider the cause of this problem, where and how frequently it is encountered, and what steps (if any) need to be taken.

This situation is caused by the shape of the distribution, particularly by the fact that the mode is at one of the boundaries of the range of the parameter. Note that the true mode may indeed be at one of the boundaries, which means that the variation in the histogram is not the result of the sample being too small. Therefore, as is also the case in the example above, the problem cannot be solved by simply drawing a larger Gibbs sample.

Skewness of the distribution may also cause an estimate to fall outside its hpd region. In some instances, especially in the case of c, the distribution is heavily skewed. In such cases the mean is influenced by the tail of the distribution to such an extent that the mean estimate falls outside the hpd region.

Fortunately, the problem of estimates outside their hpd regions poses little or no problems as far as our proposed PPD estimator is concerned. As we shall see later in this chapter, the PPD estimator which we propose as benchmark estimator of the EVI (γ), has the following properties:

1. The PPD estimate is always the median or mode estimate of γ.
2. The parameter ρ is always estimated by making use of Gibbs sampling. In other words, we do not choose $\rho = -1$ or estimate ρ by another method before conducting Gibbs sampling.

Note that we choose the estimates of ρ and c to always correspond to our choice of estimate of γ. If, for example, we use the median estimate of γ with $k = 100$ (number of excesses) as our estimate of γ, then our estimate of ρ is the median estimate of ρ, as yielded by the same Gibbs sampling process. Similarly for c. This approach is the most straightforward approach, and has been applied throughout.

This approach implies that mean estimates falling outside their hpd regions present no problem, since we use only the median and mode esimates of γ, and therefore only the median and mode estimates of ρ and c (Property 1 above).

The median of a sample is always inside the 90%, 95% and 99% hpd regions, because of the definition of the median. This means that the only problems we could encounter, will be when the mode estimates fall outside their hpd regions.

Fortunately, not one of the mode estimates of γ in the entire simulation study was ever outside its 90%, 95% or 99% hpd regions. For the parameters ρ and c the extent of the problem is summarized in the following table, which gives the percentage of times the mode estimates fell outside their respective hpd regions.

		Sample size								
	Outside	**50**	**100**	**200**	**500**	**1000**	**2000**	**5000**	**10 000**	**20 000**
$\hat{\rho}$	**90% hpd**	44.0%	42.1%	37.8%	30.3%	24.6%	19.4%	13.9%	10.7%	8.20%
$\hat{\rho}$	**95% hpd**	43.6%	42.0%	37.7%	30.2%	24.5%	19.3%	13.9%	10.7%	8.20%
$\hat{\rho}$	**99% hpd**	42.7%	41.5%	37.5%	30.1%	24.4%	19.2%	13.8%	10.6%	8.15%
$\hat{c}$	**90% hpd**	4.47%	3.80%	2.74%	1.63%	1.02%	0.64%	0.49%	0.38%	0.27%
$\hat{c}$	**95% hpd**	4.47%	3.79%	2.74%	1.63%	1.01%	0.64%	0.49%	0.38%	0.27%
$\hat{c}$	**99% hpd**	4.47%	3.79%	2.74%	1.63%	1.01%	0.64%	0.49%	0.38%	0.27%

Table 3.2.4 Percentage of mode estimates outside the hpd regions

Since we restrict the value of ρ to $-2 \leq \rho < 0$, we suggest that one should not use the results from the Gibbs sampling for inference concerning ρ. A better option would be to use estimators of ρ suggested in the literature, one of which will be discussed in Section 4.1. Since we are not going to use the hpd region of ρ, it of no consequence to us whether or not the estimate of ρ falls inside the hpd region.

Also, since c is a nuisance parameter, we are not really interested in inference concerning it.

As far as quantile estimation is concerned, we found that, across the enitre simulation study, no mode or median quantile estimates were outside their respective hpd regions.

In conclusion, we see that the problem of estimates falling outside the hpd regions has, for our purposes, no practical implications.

The process of quantile estimation will be discussed in the next section.

3.2.5 Quantile estimation

In this section we will see how to estimate extreme quantiles by making use of Gibbs sampling.

Let z_p be the quantile of the $PPD(\gamma, \rho, c)$ distribution, which has upper probability p. Then

$$p = (1-c)z_p^{-\frac{1}{\gamma}} + cz_p^{-\frac{1-\rho}{\gamma}}. \qquad \mathbf{(3.1)}$$

For given values of the parameters γ, ρ and c, we can solve for z_p. This is done numerically.

We can, therefore, estimate a PPD quantile by estimating the parameters of the PPD, substituting them in (3.1), and solving for z_p. The problem with this approach is that we cannot do inference by constructing an interval for z_p. This necessitates the method of quantile estimation which will now be presented. This method is a standard method of quantile estimation when using Gibbs sampling.

The first thing to note about quantiles, is that they can be viewed as parameters. We can see this in the case of the PPD by examining (3.1) and noting that we can solve for any parameter if z_p and the two other parameters are given. Hence, the PPD can be reparameterized in terms of z_p and two of the three other parameters γ, ρ and c. The fact that quantiles can be interpreted as parameters is the reason why we treated quantiles in the same manner as we treated parameters in the previous subsection on the hpd region.

Gibbs sampling enables us to obtain draws of a quantile. In other words, Gibbs sampling yields a vector of draws which can be regarded as a sample from the marginal posterior distribution of the quantile, which is the same as what we obtained for the parameters γ, ρ and c. We will now explain how this is done.

At every repetition of the Gibbs sampler, we have a new set of parameters γ, ρ and c, and we can calculate the value of the quantile z_p for that set of parameters. These calculated values of the quantiles can be regarded as draws from the marginal posterior distribution of the quantile. We can use these draws to estimate quantiles and to make inferences by constructing hpd regions, following the exact same methodology as we did for the parameters γ, ρ and c.

For the PPD Gibbs sampler, another aspect requires attention. As mentioned when we discussed Step 2.3 of the Gibbs sampler in Section 3.2.3, the restriction $1/\rho \le c \le 1$ is not applied when drawing a value of ρ. Therefore we cannot let Step 2.4 be the step in which we draw a value of the quantile, since $(1-c)z^{-1/\gamma} + cz^{-(1-\rho)/\gamma}$ might not be a survival function (in the sense that the corresponding density function can be negative for some values of z). We have to draw the value of the quantile after Step 2.1 or after Step 2.2. In the MATLAB program we draw it after c is drawn in Step 2.1.

Up to now we have described how to estimate a PPD quantile. We are actually not interested in PPD quantiles, but in quantiles of the underlying distribution from which the original data were generated, i.e. the distribution of the original data, not the distribution of the excesses.

The setting is the same as we had for the Gibbs sampler. Let X be a random variable which has a heavy-tailed distribution. We have n observations at our disposal. Ordered, they are denoted by $x_{1,n} \le x_{2,n} \le \cdots \le x_{n,n}$. The number of multiplicative excesses is k (a given quantity at this stage), and we take the threshold to be $t = x_{n-k,n}$.

We require the value (quantile) x such that, for a given probability p_1, $P(X > x) = p_1$. Since we are interested in extreme quantiles, we can assume that $x > t$ and that p_1 is small. The solution is as follows:

$$P(X > x) = p_1$$

$$\frac{P(X/t > x/t)}{P(X > t)} = \frac{p_1}{P(X > t)}$$

Since $x > t$, $\frac{P(X/t>x/t)}{P(X>t)} = P\left(\frac{X}{t} > \frac{x}{t} \middle| X > t\right)$. Also $P(X > t) \approx \frac{k}{n}$.

We now have

$$P\left(\frac{X}{t} > \frac{x}{t} \middle| X > t\right) \approx \frac{np_1}{k}.$$

Let $Z = X/t$, $z = x/t$ and $p_2 = np_1/k$. Then

$$P(Z > z) \approx p_2.$$

Note that t and p_2 are known quantities. If we assume the threshold t is large enough, the distribution of the multiplicative excess Z is approximately $PPD(\gamma, \rho, c)$. Given γ, ρ and c, we can solve for (the approximate value of) z, using equation (3.1) as described above. Then

$$x \approx tz.$$

In summary, if we have the setting mentioned above and require x such that $P(X > x) = p_1$, we do the following:

1. Let $p_2 = np_1/k$.
2. Choose k, calculate the k excesses and let $t = x_{n-k,n}$.
3. Use the Gibbs sampler as described in Section 3.2.3 to fit the PPD to the k excesses by obtaining draws of γ, ρ and c, but insert the following steps after Step 2.1:
    1. Calculate z, which is the quantile of the $PPD\left(\gamma^{(i-1)}, \rho^{(i-1)}, c^{(i)}\right)$ with upper probability p_2.
    2. Let $x^{(i)} = tz$.

For purposes of estimation and inference we use the vector of draws $\underline{x} = \left(x^{(1)}, x^{(2)}, \dots, x^{(N)}\right)$ in the same manner as we used the draws of γ, ρ and c.

Some simulation results of quantile estimation will be shown in Section 3.4.5.

3.2.6 Estimating exceedance probability

The calculation of an exceedance probability, given a certain value (quantile), involves the inverse of the process described in the previous section on quantile estimation.

The setting is the same as we had in the previous section, and again have $p_1 = P(X > x)$, but instead of p_1 being given and an estimate of x required, we are given x and need to estimate p_1. Here x is an extreme quantile and usually $x > x_{n,n}$. (If x is not an extreme quantile, we do not need EVT. In such a case we can estimate the probability in the same way as we did for Question a) in Section 2.1.)

Following the same reasoning and defining variables and quantities in the same way as we did in Section 3.2.5, we have

$$p_1 \approx \frac{kP(Z > z)}{n}.$$

Again we assume the threshold t to be large enough. The distribution of the multiplicative excess Z is therefore approximately $PPD(\gamma, \rho, c)$. We estimate the parameters γ, ρ and c by applying Gibbs sampling. As a result, an estimate of p_1 can be calculated.

For clarity, we now restate the problem and the proposed procedure to estimate the exceedance probability:

Let X be random variable which has a heavy-tailed distribution. We have n observations at our disposal. Ordered, they are denoted $x_{1,n} \leq x_{2,n} \leq \cdots \leq x_{n,n}$. We require an estimate of $p_1 = P(X > x)$ for some given $x > x_{n,n}$.

The procedure for estimating p_1 is:

1. Choose k, calculate the k excesses and let $t = x_{n-k,n}$. The choice of k will be discussed in Section 3.4.5.
2. Use the Gibbs sampler as described in Section 3.2.3 to fit the PPD to the k excesses by obtaining draws of γ, ρ and c. This yields parameter estimates $\hat{\gamma}$, $\hat{\rho}$ and $\hat{c}$. Whether these estimates should be mean, median or mode estimates, will be discussed in Section 3.4.4.
3. Let $z = x/t$, and calculate
$$\hat{p}_2 = (1-\hat{c})z^{-1/\hat{\gamma}} + \hat{c}z^{-(1-\hat{\rho})/\hat{\gamma}} \approx P(Z > z).$$
4. Then
$$\hat{p}_1 = \frac{k\hat{p}_2}{n}.$$

We now have a method of obtaining a point estimate of the exceedance probability, which suffices for the purpose of this text, even though one can apply Gibbs sampling in a similar fashion to that of Section 3.2.5, if a hpd region is required.

For the remaining part of this section we discuss a problem which is sometimes encountered when obtaining a point estimate of the exceedance probability.

The problem pertains to the fact that we require $1/\rho \leq c \leq 1$ in order for

$$(1-c)z^{-1/\gamma} + cz^{-(1-\rho)/\gamma}$$

to be a survival function.

In Step 2.1 of the Gibbs sampler (Section 3.2.3) the value of c is restricted in this manner, but only with respect to the previous draw of ρ. This implies that, when calculating the estimates $\hat{\rho}$ and $\hat{c}$, these estimates need not necessarily satisfy $1/\hat{\rho} \leq \hat{c}$. Here $\hat{\rho}$ and $\hat{c}$ denote any estimates of ρ and c, i.e. the mean, median or mode of the draws.

If $\hat{\rho}$ and $\hat{c}$ do not satisfy $1/\hat{\rho} \leq \hat{c}$, there are three alternatives:

1. Consider the distribution function of the resulting PPD merely as a function approximation, even though this function is not a distribution function. This is not recommended because of the fact that, when an exceedance probability is calculated, one might come up with a negative answer.
2. Set $\hat{\rho} = 1/\hat{c}$. A similar argument against this option can be presented as the one presented in Section 3.2.3. We stated there that an additional restriction on ρ should not be applied, since we have already restricted ρ to values between -2 and 0.
   Furthermore, note that an additional restriction on $\hat{\rho}$ might adversely affect results. The reason for this can best be explained by the fact that $\hat{c}$ is negative (almost without exception), and that the requirement $\hat{\rho} \geq 1/\hat{c}$ forces $\hat{\rho}$ closer to zero when $\hat{c} < 0$. For ρ close to zero

$(1-c)z^{-1/\gamma} + cz^{-(1-\rho)/\gamma} \approx z^{-1/\gamma}$. The fitted distribution is approximately a Pareto distribution. Therefore the purpose and advantage of fitting a second order model is defeated to a great extent.

3. Set $\hat{c} = 1/\hat{\rho}$. This is the alternative we suggest. It also works very well in practice. We do not intend to draw inferences on the parameter c. Thus, adjusting our estimate of c does not pose a problem for us in that respect.

This problem is of course not unique to exceedance probability estimation, but may emerge whenever one intends to use the fitted PPD model

$$(1-\hat{c})z^{-1/\hat{\gamma}} + \hat{c}z^{-(1-\hat{\rho})/\hat{\gamma}},$$

where estimates $\hat{\gamma}$, $\hat{\rho}$ and $\hat{c}$ were obtained by Gibbs sampling.

The extent of the problem is fortunately not that great. For mean estimates of ρ and c the problem occurs quite frequently, and adjustments made to the value of $\hat{c}$ are significant. Fortunately we will never use these estimators, since the EVI estimator we propose is not a mean estimator, as we will see in Section 3.4.4. Across the entire simulation study the problem never occurred for median estimates of ρ and c. For the mode estimates the problem occurred at only 0.25% of the samples, and it might be safe to assume that in those cases the adjustments made to the value $\hat{c}$ were not too great.

3.3 Simulation study

In order for us to compare the performance of different estimators, we will make use of simulation. Section 3.3.1 outlines the methodology and design of the simulation study. The results of the simulation study are presented in Section 3.3.2.

3.3.1 Simulation study methodology and design

The methodology of the simulation study presented here is standard. Specifically, when examining the performance of an EVI estimator, a simulation study consists of the following general steps:

1. Choose a sample size n. We will test the performance of estimators across a wide range of sample sizes. The sample sizes we consider are shown in the subsection *Sample sizes and number of repetitions* below.
2. Choose a distribution with known true underlying EVI γ. We consider samples from a wide variety of heavy-tailed distributions. See the subsection *Distributions* below.
3. Choose a threshold or, equivalently, the number of order statistics k. All estimators we consider are based on the k largest order statistics. In the case of the PPD estimator, the k largest order statistics are divided by the threshold ($t = x_{n-k,n}$) to obtain k multiplicative excesses. In this section we start off by fixing the value of k, i.e. k is expressed as a

proportion of the sample size n before any estimation is done. See the subsection *Fixed k* below. We will also develop an adaptive method of selecting k in Section 3.4.

4. Generate a sample of size n from the distribution in Step 2.
5. Given k (Step 3), use the EVI estimator under consideration to obtain $\hat{\gamma}$.
6. Calculate the square error of estimation $(\gamma - \hat{\gamma})^2$.
7. Repeat Steps 4 through 6 a large number of times N, referred to as *repetitions* subsequently, and not to be confused with the number of Gibbs repetitions. We already fixed the number of Gibbs repetitions at 1000 for all estimation done as part of the simulation study. From here onwards we will not refer to Gibbs repetitions, unless it is stated explicitly as *Gibbs* repetitions. The number of repetitions used in the simulation study, which is a function of the sample size n, is given in the subsection *Sample sizes and number of repetitions*.
8. Calculate the mean of the square errors obtained in Step 6, yielding the mean square error (MSE). Justification for the use of the MSE as measure of performance is given in the subsection *MSE as measure of performance* below.

We now turn to the details of the simulation study design. This design will be used throughout Chapters 3 and 4.

Distributions

We will simulate observations from 18 distributions belonging to the Fréchet, Burr, *t* and loggamma families of distributions, since for them expressions for the first and second order parameters are known, as shown in Table 2.4.4 in Section 2.4.4. We restrict the Burr to $\beta = 1$ and loggamma to $\alpha = 2$ in order to limit the number of distributions to a number that is feasible for a simulation study. For the loggamma distribution $\alpha = 2$ is chosen, since $\alpha = 1$ yields the Pareto distribution. The simulation study will not include observations simulated from the limiting distributions like the Pareto distribution, the GPD, the PPD and the GEV distribution, since virtually any estimator would yield good results if data were from these distributions.

Having decided on which families of distributions to simulate from, the next step is to decide on a range of values for γ and ρ.

The main application of the results of this thesis will be in the context of insurance. Most actuaries assume the EVI of insurance claim size distributions to be less than one. This is because the expected value of the underlying distribution is infinite if the EVI is greater than one. Since we are only concerned with the $\gamma > 0$ case, we will choose as range of the EVI $0 < \gamma \leq 1$.

The second order parameter ρ is known to be difficult to estimate. It also does not have a clear and intuitive interpretation as we have in the case of the EVI, where the EVI can be interpreted as the limiting slope of a Pareto plot. We know that $\rho \leq 0$, but what the lower bound for simulation should be, is unclear. The range we choose to simulate from is $-2 \leq \rho \leq 0$. A justification for this range is given in Section 4.9.3.

The following table shows the 18 distributions from which samples were generated:

No.	**Distribution**	$\boldsymbol{\gamma}$	$\boldsymbol{\rho}$
1	$Burr(\beta = 1, \gamma = 0.1, \rho = -2)$	0.1	-2
2	$Burr(\beta = 1, \gamma = 0.5, \rho = -2)$	0.5	-2
3	$Burr(\beta = 1, \gamma = 1, \rho = -2)$	1	-2
4	$Burr(\beta = 1, \gamma = 0.1, \rho = -1)$	0.1	-1
5	$Burr(\beta = 1, \gamma = 0.5, \rho = -1)$	0.5	-1
6	$Burr(\beta = 1, \gamma = 1, \rho = -1)$	1	-1
7	$Burr(\beta = 1, \gamma = 0.1, \rho = -0.5)$	0.1	-0.5
8	$Burr(\beta = 1, \gamma = 0.5, \rho = -0.5)$	0.5	-0.5
9	$Burr(\beta = 1, \gamma = 1, \rho = -0.5)$	1	-0.5
10	*Fréchet*(10)	0.1	-1
11	*Fréchet*(2)	0.5	-1
12	*Fréchet*(1)	1	-1
13	$\lvert t_{10} \rvert$	0.1	-0.2
14	$\lvert t_2 \rvert$	0.5	-1
15	$\lvert t_1 \rvert$	1	-2
16	$\log\Gamma(\lambda = 10, \alpha = 2)$	0.1	0
17	$\log\Gamma(\lambda = 2, \alpha = 2)$	0.5	0
18	$\log\Gamma(\lambda = 1, \alpha = 2)$	1	0

Table 3.3.1.1 Simulation study distributions.

Note that this list will be revised in Section 3.3.2.

Sample sizes and number of repetitions

The following table indicates which sample sizes will be considered for the simulation study, together with the number of samples of each size that will be generated:

Sample size	Repetitions
50	1000
100	500
200	500
500	200
1000	200
2000	200
5000	100
10 000	100
20 000	100

Table 3.3.1.2 Simulation study sample sizes and number of repetitions

The number of repetitions must be larger for smaller samples, since the variation is larger. The same samples from the distributions will be used across all techniques, reducing the number of repetitions required to obtain reliable comparisons considerably. For example, the same 1000 samples of size $n = 50$ from a *Burr*$(\beta = 1, \gamma = 0.1, \rho = -1)$ distribution will be used for all estimation techniques. The repetitions for samples of size $n = 500$ onward seem small, but they were tested and found to be sufficient to see clear patterns and yield statistically significant differences between methods of estimation.

Fixed k as method of threshold selection

We know from the theoretical aspects mentioned in Section 3.1.4 and from what we have seen in Figure 3.1.5, that fitting the PPD yields more *stable* estimates of the EVI than the Hill estimator. This means that the estimates obtained by fitting the PPD stay relatively close to the true value of the EVI for a wider range of values of k. As a further example, refer to Figure 3.4.2.2 which indicates a range of values of k which can be regarded as a region over which the estimates are stable.

Even though the choice of threshold is not as critical as in the case of the Hill estimator, we still need to investigate what the effect of choice of threshold is on the accuracy of estimation. The most straightforward way to examine this effect, is to consider a range of values of k, expressing k as a proportion of the sample size n. We fix the value of k at one of these values in the range, before any estimation is done.

Our approach is to divide the range of k into equal parts by considering as possible values of k, $k = 5\%, 10\%, \dots, 95\%$ of n, rounded to the nearest integer.

MSE as measure of performance

The mean square error (MSE) is a well-known and frequently employed measure to assess performance when conducting a simulation study. In this section we provide further justification for its use in our specific context.

We need to bear in mind that the EVI is only a means to an end. The ultimate purpose of an extreme value analysis is inference concerning large quantiles. In order to get a feel for the problem of performance measurement, we need to take into account what effect error in the estimation of the EVI has on the error in the estimation of large quantiles.

Consider the formulas we used in Section 2.1 to estimate a large quantile. We calculated $y_2 = y_1 + 0.5183(x_2 - x_1) = 10.3691$, yielding a quantile estimate of $e^{10.3691} = 31\,859.957$.

Rewriting the formula in general, we obtain

$$\exp(y_1 + \hat{\gamma}(x_2 - x_1)),$$

where $\hat{\gamma}$ is the estimated EVI. Recall that (x_1, y_1) is the anchor point and (x_2, y_2) the point which is obtained by extrapolating beyond the data.

If the error that was made when estimating γ with $\hat{\gamma}$ is denoted by δ, we have that our quantile estimate is $\exp(y_1 + (\gamma + \delta)(x_2 - x_1))$, and the true quantile is $\exp(y_1 + \gamma(x_2 - x_1))$.

The relative error of the estimate is therefore $\frac{\exp(y_1+(\gamma+\delta)(x_2-x_1))}{\exp(y_1+\gamma(x_2-x_1))} = \exp(\delta(x_2 - x_1))$. It is clear that changes in $\hat{\gamma}$ have an exponential impact on accuracy of large quantile estimation. The penalty on the error should therefore be at least quadratic. For this reason we will use the MSE rather than other measures like the mean/median absolute error for measuring estimation accuracy.

In order to compare estimation accuracy across distributions with different values of the true underlying EVI, one might be tempted to use the root mean square error and divide it by the true EVI, to account for this difference. We can see from the formula for the error factor of the estimated quantile, which is $\exp(\delta(x_2 - x_1))$, that this does not make sense. We see this by noting that it is only the difference δ which changes from one method of estimation to the next. For a given setting, the values of x_1 and x_2 are fixed. The value of x_1 depends only on the threshold and x_2 only on the exceedance probability of the quantile that needs to be estimated. The error factor is therefore only a function of the difference between the true EVI and the estimated EVI, and is independent of the true value of the EVI itself.

Aside from the importance of quantile inference, there is another crucial reason why we base our justification of the MSE on the effect of EVI estimation on quantile estimation. It is a matter of practical impossibility to derive, program and test by simulation, the performance of all estimators with respect to quantile estimation, because of the large number of estimation methods we consider in this thesis.

It is obviously true in general that “better” EVI estimation and “better” extreme quantile estimation are closely linked. Since we compare estimators only on the basis of EVI estimation (in this thesis), it is even more important to us that this link is strong.

3.3.2 Simulation study results

In this section we report the simulation study results when estimating the EVI in the way we described above. This entails the following:

- The PPD is fitted to k multiplicative excesses, yielding an estimate of the PPD parameter γ (EVI).
- We employ Bayesian estimation of the PPD parameters γ, ρ and c, using Gibbs sampling.
- Whether the mean, median or mode of the γ draws was used, will always be stated explicitly.

When reporting results, the table entries are always 100×MSE. This is done simply for readability and ease of comparison.

For $n = 50$, the median estimator of the EVI yielded the following results:

					k					
Distr.	**5%**	**10%**	**15%**	**20%**	**25%**	**30%**	**35%**	**40%**	**45%**	**50%**
1	0.280	0.171	0.118	0.099	0.080	0.073	0.065	0.059	0.055	**0.050**
	(0.015)	(0.008)	(0.006)	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
2	4.733	3.538	2.648	2.346	1.979	1.775	1.591	1.458	1.305	**1.193**
	(0.166)	(0.126)	(0.092)	(0.079)	(0.068)	(0.060)	(0.054)	(0.051)	(0.047)	(0.044)
3	18.354	13.835	10.972	9.877	8.611	7.802	6.849	6.293	5.543	**5.172**
	(0.575)	(0.449)	(0.367)	(0.317)	(0.282)	(0.259)	(0.232)	(0.211)	(0.188)	(0.181)
4	0.269	0.164	0.106	0.086	0.070	0.064	0.059	0.058	**0.058**	0.059
	(0.014)	(0.007)	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
5	4.822	3.229	2.246	1.866	1.507	1.339	1.161	1.103	**1.062**	1.114
	(0.186)	(0.135)	(0.090)	(0.080)	(0.065)	(0.060)	(0.055)	(0.053)	(0.052)	(0.056)
6	16.484	12.507	8.567	7.307	5.864	5.105	4.428	3.918	3.798	**3.478**
	(0.548)	(0.441)	(0.316)	(0.284)	(0.233)	(0.209)	(0.194)	(0.190)	(0.191)	(0.182)
7	0.392	0.260	0.194	**0.185**	0.196	0.216	0.253	0.281	0.345	0.419
	(0.023)	(0.015)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.011)	(0.012)
8	4.699	3.685	3.031	**2.980**	3.253	3.523	4.646	5.225	7.032	8.282
	(0.196)	(0.182)	(0.151)	(0.150)	(0.145)	(0.155)	(0.177)	(0.189)	(0.224)	(0.236)
9	15.476	10.001	**7.831**	8.005	8.436	9.949	12.534	16.365	20.729	25.896
	(0.638)	(0.421)	(0.379)	(0.372)	(0.393)	(0.452)	(0.497)	(0.631)	(0.671)	(0.812)
10	0.246	0.155	0.100	0.084	0.070	0.062	0.055	0.051	0.046	**0.043**
	(0.014)	(0.007)	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
11	4.983	3.323	2.386	2.040	1.704	1.539	1.332	1.194	1.071	**0.993**
	(0.187)	(0.122)	(0.087)	(0.075)	(0.065)	(0.058)	(0.052)	(0.046)	(0.043)	(0.041)
12	19.023	14.074	10.286	9.007	7.453	6.747	5.858	5.315	4.617	**4.039**
	(0.618)	(0.465)	(0.352)	(0.314)	(0.270)	(0.249)	(0.221)	(0.201)	(0.172)	(0.153)
13	**1.669**	1.928	2.459	2.920	3.848	4.575	5.729	6.779	8.710	10.161
	(0.079)	(0.077)	(0.071)	(0.076)	(0.083)	(0.089)	(0.096)	(0.109)	(0.139)	(0.150)
14	4.621	3.104	2.182	1.831	1.461	1.336	**1.292**	1.370	1.671	2.105
	(0.196)	(0.136)	(0.094)	(0.080)	(0.071)	(0.065)	(0.062)	(0.063)	(0.077)	(0.094)
15	19.153	14.737	11.413	9.807	8.164	7.492	6.422	5.870	4.967	**4.479**
	(0.599)	(0.488)	(0.380)	(0.323)	(0.269)	(0.248)	(0.217)	(0.199)	(0.176)	(0.160)
16	0.337	0.206	0.139	0.116	0.097	0.090	0.087	0.085	0.082	**0.081**
	(0.021)	(0.011)	(0.007)	(0.006)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
17	4.858	3.304	2.299	1.987	1.715	1.558	1.537	1.490	**1.463**	1.490
	(0.208)	(0.147)	(0.101)	(0.093)	(0.083)	(0.075)	(0.078)	(0.075)	(0.071)	(0.073)
18	16.425	11.462	7.443	6.513	5.354	4.925	4.573	4.338	4.228	**3.966**
	(0.588)	(0.447)	(0.300)	(0.268)	(0.237)	(0.222)	(0.212)	(0.212)	(0.219)	(0.203)
Mean	7.601	5.538	4.134	3.725	3.326	**3.232**	3.248	3.403	3.710	4.057
	(0.085)	(0.065)	(0.050)	(0.045)	(0.041)	(0.041)	(0.041)	(0.045)	(0.047)	(0.053)

Table 3.3.2.1 100×MSE of median estimates for samples of size n = 50.

The table above is a standard format which will be used quite frequently, and represents the following:

- Each column represents a different method of estimating the EVI. In this case, as in most other cases, the method of estimation differs only with respect to the choice of the number of excesses k as a % of n.
- Distributions 1 to 18 refer to the distributions listed in Table 3.3.1.1. The list of distributions will change subsequently. See the subsection *Revised list of distributions* below.
- Across rows, the minimum 100×MSE is printed in bold.
- The average 100×MSE across all distributions is given in the second last row.
- Values in brackets always represent the standard error of the value immediately above them.

As an example, let us explain how we arrived at the number 2.980 (in bold in the k = 20% column and the row corresponding to Distribution 8) and its standard error 0.150 (in brackets below 2.980):

1. For n = 50 the number of repetitions is 1000 (see Table 3.3.1.2). We therefore generated a 1000 samples of size 50 from Distribution 8, the *Burr*$(\beta = 1, \gamma = 0.5, \rho = -0.5)$ distribution.
2. For each sample we estimated the EVI, using k = 10, which is 20% of n. Each estimate is a Bayesian estimate obtained by fitting the PPD to 10 excesses by means of obtaining Gibbs samples of γ, ρ and c, and calculating the median of the γ draws. Denote the estimated EVI of sample i by $\hat{\gamma}_i$, where $i = 1,2, \dots ,1000$.
3. We then calculated 100×MSE = $100\left(\frac{1}{1000}\sum_{i=1}^{1000}(0.5 - \hat{\gamma}_i)^2\right)$ = 2.980.
4. The standard error is calculated as $100\sqrt{s^2/1000} = 0.150$, where $s^2 = \frac{1}{999}\sum_{i=1}^{1000}(\hat{\gamma}_i - \bar{\gamma})^2$ and $\bar{\gamma} = \frac{1}{1000}\sum_{i=1}^{1000}\hat{\gamma}_i$.

The mean in the second last row is simply obtained by calculating the average of the MSEs of the 18 distributions. For example, for the k = 20% column

$$(0.099 + 2.346 + \cdots + 6.513)/18 = 3.725.$$

The overall standard error for a specific column is calculated as follows. Let s_j denote the standard error of Distribution j, as it appears in the table in brackets. The overall standard error is

$$\sqrt{\frac{\sum_{j=1}^{18} s_j^2}{18^2}}.$$

For the k = 20% column, for example, the overall standard error is

$$\sqrt{\frac{0.005^2 + 0.079^2 + \cdots + 0.268^2}{18^2}} = 0.045.$$

Removal of the t_{10} distribution

On closer inspection of Table 3.3.2.1 one notices that, as the size of the true underlying value of γ increases, so does the value of the MSE. For example, consider Distributions 1 to 3 at k = 20%:

No.	Distribution	γ	100×MSE
1	$Burr(\beta = 1, \gamma = 0.1, \rho = -2)$	0.1	0.099
2	$Burr(\beta = 1, \gamma = 0.5, \rho = -2)$	0.5	2.346
3	$Burr(\beta = 1, \gamma = 1, \rho = -2)$	1	9.877

Table 3.3.2.2 Typical effect of the true EVI on the MSE.

It is therefore a good idea to take this effect into account when assessing the performance of an estimator. The easiest way to do this, is to consider the results for $\gamma = 0.1$ separately from the result for $\gamma = 0.5$, etc.

When separating the results in this way, we see that the *t* distribution presents a problem. Consider, for example, the $k = 30\%$ column in Table 3.3.2.1, which is where the overall minimum MSE is reached:

No.	Distribution	γ	100×MSE
13	$\lvert t_{10} \rvert$	0.1	4.575
14	$\lvert t_2 \rvert$	0.5	1.336
15	$\lvert t_1 \rvert$	1	7.492

Table 3.3.2.3 100×MSE for the *t* distribution at k = 30%.

Here the MSE for $\gamma = 0.1$ is greater than the MSE when $\gamma = 0.5$. This is the case with $k = 15\%$ to $k = 50\%$, as can be seen in Table 3.3.2.1.

Another problem comes to the fore when we consider all distributions where $\gamma = 0.1$. The next table shows the results for $k = 30\%$:

No.	Distribution	γ	100×MSE
1	$Burr(\beta = 1, \gamma = 0.1, \rho = -2)$	0.1	0.073
4	$Burr(\beta = 1, \gamma = 0.1, \rho = -1)$	0.1	0.064
7	$Burr(\beta = 1, \gamma = 0.1, \rho = -0.5)$	0.1	0.216
10	*Fréchet*(10)	0.1	0.062
13	$\lvert t_{10} \rvert$	0.1	4.575
16	$\log\Gamma(\lambda = 10, \alpha = 2)$	0.1	0.090

Table 3.3.2.4 100×MSE for distributions with $\gamma = 0.1$ at k = 30%.

We see that the MSE corresponding to the t distribution is more than 20 times that of the second largest MSE. The result is that, for $\gamma = 0.1$, the choice of optimal threshold is completely dominated by the t distribution. The simulations performed on the other four distributions have virtually no effect.

For $\gamma = 0.1$, the influence of the t distribution would not be that great if the optimal k was close to the optimal k of the other distributions. This is not the case. As can be seen from Table 3.3.2.1, k should be very small to obtain the lowest MSE, and even that MSE is very high (1.669).

For the t_{10} distribution, $\rho = -0.2$. The problem of being unable to accurately estimate the EVI for observations from a t distribution with ρ close to zero, is not uncommon. In fact, we found this to be the case with several estimators considered here and in Chapter 4.

We will subsequently omit the t_{10} distribution entirely from the simulation study.

Revised list of distributions

After removing the t_{10} distribution, and sorting by the EVI, we obtain the revised list of distributions below:

No.	Distribution	γ	ρ
1	$Burr(\beta = 1, \gamma = 0.1, \rho = -2)$	0.1	-2
2	$Burr(\beta = 1, \gamma = 0.1, \rho = -1)$	0.1	-1
3	$Burr(\beta = 1, \gamma = 0.1, \rho = -0.5)$	0.1	-0.5
4	$Fréchet(10)$	0.1	-1
5	$\log\Gamma(\lambda = 10, \alpha = 2)$	0.1	0
6	$Burr(\beta = 1, \gamma = 0.5, \rho = -2)$	0.5	-2
7	$Burr(\beta = 1, \gamma = 0.5, \rho = -1)$	0.5	-1
8	$Burr(\beta = 1, \gamma = 0.5, \rho = -0.5)$	0.5	-0.5
9	$Fréchet(2)$	0.5	-1
10	$\lvert t_2 \rvert$	0.5	-1
11	$\log\Gamma(\lambda = 2, \alpha = 2)$	0.5	0
12	$Burr(\beta = 1, \gamma = 1, \rho = -2)$	1	-2
13	$Burr(\beta = 1, \gamma = 1, \rho = -1)$	1	-1
14	$Burr(\beta = 1, \gamma = 1, \rho = -0.5)$	1	-0.5
15	$Fréchet(1)$	1	-1
16	$\lvert t_1 \rvert$	1	-2
17	$\log\Gamma(\lambda = 1, \alpha = 2)$	1	0

Table 3.3.2.5 Revised list of distributions.

The numbers of the distributions as they appear in Table 3.3.2.5 will be used subsequently in all tables presenting simulation results.

Simulation results

For $n = 50$, the results of the median estimator of the EVI are shown in the following table:

					k					
Distr.	**5%**	**10%**	**15%**	**20%**	**25%**	**30%**	**35%**	**40%**	**45%**	**50%**
1	0.280	0.171	0.118	0.099	0.080	0.073	0.065	0.059	0.055	**0.050**
	(0.015)	(0.008)	(0.006)	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
2	0.269	0.164	0.106	0.086	0.070	0.064	0.059	0.058	**0.058**	0.059
	(0.014)	(0.007)	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
3	0.392	0.260	0.194	**0.185**	0.196	0.216	0.253	0.281	0.345	0.419
	(0.023)	(0.015)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.011)	(0.012)
4	0.246	0.155	0.100	0.084	0.070	0.062	0.055	0.051	0.046	**0.043**
	(0.014)	(0.007)	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
5	0.337	0.206	0.139	0.116	0.097	0.090	0.087	0.085	0.082	**0.081**
	(0.021)	(0.011)	(0.007)	(0.006)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
γ **= 0.1**	0.305	0.191	0.131	0.114	0.103	**0.101**	0.104	0.107	0.117	0.131
	(0.008)	(0.005)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
6	4.733	3.538	2.648	2.346	1.979	1.775	1.591	1.458	1.305	**1.193**
	(0.166)	(0.126)	(0.092)	(0.079)	(0.068)	(0.060)	(0.054)	(0.051)	(0.047)	(0.044)
7	4.822	3.229	2.246	1.866	1.507	1.339	1.161	1.103	**1.062**	1.114
	(0.186)	(0.135)	(0.090)	(0.080)	(0.065)	(0.060)	(0.055)	(0.053)	(0.052)	(0.056)
8	4.699	3.685	3.031	**2.980**	3.253	3.523	4.646	5.225	7.032	8.282
	(0.196)	(0.182)	(0.151)	(0.150)	(0.145)	(0.155)	(0.177)	(0.189)	(0.224)	(0.236)
9	4.983	3.323	2.386	2.040	1.704	1.539	1.332	1.194	1.071	**0.993**
	(0.187)	(0.122)	(0.087)	(0.075)	(0.065)	(0.058)	(0.052)	(0.046)	(0.043)	(0.041)
10	4.621	3.104	2.182	1.831	1.461	1.336	**1.292**	1.370	1.671	2.105
	(0.196)	(0.136)	(0.094)	(0.080)	(0.071)	(0.065)	(0.062)	(0.063)	(0.077)	(0.094)
11	4.858	3.304	2.299	1.987	1.715	1.558	1.537	1.490	**1.463**	1.490
	(0.208)	(0.147)	(0.101)	(0.093)	(0.083)	(0.075)	(0.078)	(0.075)	(0.071)	(0.073)
γ **= 0.5**	4.786	3.364	2.465	2.175	1.937	**1.845**	1.926	1.973	2.267	2.530
	(0.078)	(0.058)	(0.043)	(0.039)	(0.036)	(0.035)	(0.037)	(0.038)	(0.043)	(0.046)
12	18.354	13.835	10.972	9.877	8.611	7.802	6.849	6.293	5.543	**5.172**
	(0.575)	(0.449)	(0.367)	(0.317)	(0.282)	(0.259)	(0.232)	(0.211)	(0.188)	(0.181)
13	16.484	12.507	8.567	7.307	5.864	5.105	4.428	3.918	3.798	**3.478**
	(0.548)	(0.441)	(0.316)	(0.284)	(0.233)	(0.209)	(0.194)	(0.190)	(0.191)	(0.182)
14	15.476	10.001	**7.831**	8.005	8.436	9.949	12.534	16.365	20.729	25.896
	(0.638)	(0.421)	(0.379)	(0.372)	(0.393)	(0.452)	(0.497)	(0.631)	(0.671)	(0.812)
15	19.023	14.074	10.286	9.007	7.453	6.747	5.858	5.315	4.617	**4.039**
	(0.618)	(0.465)	(0.352)	(0.314)	(0.270)	(0.249)	(0.221)	(0.201)	(0.172)	(0.153)
16	19.153	14.737	11.413	9.807	8.164	7.492	6.422	5.870	4.967	**4.479**
	(0.599)	(0.488)	(0.380)	(0.323)	(0.269)	(0.248)	(0.217)	(0.199)	(0.176)	(0.160)
17	16.425	11.462	7.443	6.513	5.354	4.925	4.573	4.338	4.228	**3.966**
	(0.588)	(0.447)	(0.300)	(0.268)	(0.237)	(0.222)	(0.212)	(0.212)	(0.219)	(0.203)
γ **= 1**	17.486	12.769	9.419	8.419	7.314	7.003	**6.777**	7.016	7.314	7.838
	(0.243)	(0.185)	(0.143)	(0.129)	(0.117)	(0.117)	(0.115)	(0.130)	(0.132)	(0.151)
Mean	7.950	5.750	4.233	3.773	3.295	3.153	**3.102**	3.204	3.416	3.698
	(0.090)	(0.068)	(0.053)	(0.047)	(0.043)	(0.043)	(0.043)	(0.048)	(0.049)	(0.056)

Table 3.3.2.6 100×MSE of median estimates for samples of size n = 50.

The distribution numbers now correspond to the distributions listed in Table 3.3.2.5. Also included now are results per *EVI group*. For example, the average 100×MSE for distributions with $\gamma = 1$ at $k = 35\%$, was 6.777.

To show all results in this fashion will lead to too many large tables and an unnecessary amount of detail. The median and mode estimators alone will comprise $2 \times 9 = 18$ tables of this size, since for each of the median and the mode, we have nine sample sizes.

Therefore, we will summarize the results of Table 3.3.2.6 in the following way:

		Median		
γ = 0.1	***k***	**25%**	**30%**	**35%**
	100×MSE	0.103	**0.101**	0.104
		0.690	(0.002)	1.195
γ = 0.5	***k***	**25%**	**30%**	**35%**
	100×MSE	1.937	**1.845**	1.926
		2.618	(0.035)	2.328
γ = 1	***k***	**30%**	**35%**	**40%**
	100×MSE	7.003	**6.777**	7.016
		1.960	(0.115)	2.073
Mean	***k***	**30%**	**35%**	**40%**
	100×MSE	3.153	**3.102**	3.204
		1.173	(0.043)	2.379

Table 3.3.2.7 Simulation summary table for n = 50.

This format leaves out detail on individual distributions, and gives only very succint information on the EVI group means and overall mean.

For example, let us consider the results for $\gamma = 1$. These results are the combined results for the $\gamma = 1$ group of distributions, as can be seen from Table 3.3.2.6. Here the lowest average 100×MSE value was 6.777 (at k = 35%). The closest average 100×MSE values to the minimum were 7.003 (at k = 30%) and 7.016 (at k = 40%).

The standard error corresponding to the minimum of 6.777, is 0.115. The values 7.003 and 7.016 are 1.959 and 2.072 standard errors from 6.777, respectively. For example,

$$\frac{7.003 - 6.777}{0.115} = 1.959.$$

(Note that results as they appear in the tables are accurate to three decimal places. Calculating these values by hand might yield different results due to rounding errors.)

We now present all the simulation results in this reduced format. We will show results pertaining to the mean, median and mode estimators. Even though the mean estimate never yielded the lowest MSE, not even for any EVI group, we include the minimum MSE of the mean estimates for the sake of completeness.

We always include (except for the mean estimate) at least the two MSEs closest to the minimum MSE. We also attempt to show all MSEs within two standard errors of the minimum MSE, but we limit the number of MSEs shown, in order to be able to present the results of the median, mode and mean estimators in a single table. Because of lack of space, we sometimes omit the second column (of headings) of Table 3.3.2.7 above. See for example Table 3.3.2.9.

Recall that, for the results shown here, the PPD was fitted to k multiplicative excesses, yielding an estimate of the PPD parameter γ (EVI). We employed Bayesian estimation of the PPD parameters γ, ρ and c, using Gibbs sampling.

The following results were obtained:

			Median				**Mode**				**Mean**
γ **= 0.1**	***k***		**25%**	**30%**	**35%**		**35%**	**40%**	**45%**		**30%**
	100×MSE		0.103	**0.101**	0.104		0.081	**0.080**	0.085		**0.165**
			0.690	(0.002)	1.195		0.580	(0.002)	3.021		(0.004)
γ **= 0.5**	***k***		**25%**	**30%**	**35%**		**35%**	**40%**	**45%**		**30%**
	100×MSE		1.937	**1.845**	1.926		1.850	**1.768**	1.909		**2.384**
			2.618	(0.035)	2.328		2.634	(0.031)	4.538		(0.047)
γ **= 1**	***k***		**30%**	**35%**	**40%**		**40%**	**45%**	**50%**		**30%**
	100×MSE		7.003	**6.777**	7.016		7.915	**7.632**	7.733		**7.006**
			1.960	(0.115)	2.073		2.270	(0.125)	0.807		(0.130)
Mean	***k***		**30%**	**35%**	**40%**		**40%**	**45%**	**50%**		**30%**
	100×MSE		3.153	**3.102**	3.204		3.441	**3.392**	3.478		**3.363**
			1.173	(0.043)	2.379		1.062	(0.046)	1.859		(0.049)

Table 3.3.2.8 Simulation results for $n = 50$.

For samples of size n = 50, the estimator which yielded the lowest average MSE (3.102/100), is the median estimator, with k = 18 (≈ 35% of n). We can make similar statements with respect to other sample sizes. We now present tables similar to Table 3.3.2.8 for all the other sample sizes. At the end of this series of tables we will provide further interpretation of the results.

		Median					Mode				Mean
γ **= 0.1**		**20%**	**25%**	**30%**	**35%**		**30%**	**35%**	**40%**		**25%**
		0.069	**0.065**	0.067	0.069		0.060	**0.058**	0.060		**0.101**
		1.576	(0.002)	0.781	1.733		1.126	(0.002)	0.901		(0.003)
γ **= 0.5**		**25%**	**30%**	**35%**			**35%**	**40%**	**45%**		**25%**
		1.308	**1.282**	1.327			1.282	**1.265**	1.322		**1.596**
		0.786	(0.034)	1.344			0.542	(0.032)	1.808		(0.046)
γ **= 1**		**25%**	**30%**	**35%**			**35%**	**40%**	**45%**		**30%**
		5.277	**4.968**	4.989			5.860	**5.542**	5.644		**5.098**
		2.707	(0.114)	0.190			2.645	(0.120)	0.843		(0.128)
Mean		**25%**	**30%**	**35%**	**40%**		**35%**	**40%**	**45%**		**25%**
		2.343	**2.225**	2.250	2.298		2.538	**2.420**	2.477		**2.397**
		2.807	(0.042)	0.574	1.724		2.685	(0.044)	1.289		(0.047)

Table 3.3.2.9 Simulation results for $n = 100$.

		Median					Mode				Mean
γ **= 0.1**		**20%**	**25%**	**30%**	**35%**		**30%**	**35%**	**40%**		**25%**
		0.057	**0.054**	0.054	0.056		0.046	**0.045**	0.046		**0.082**
		1.582	(0.002)	0.276	1.360		0.522	(0.001)	0.972		(0.003)
γ **= 0.5**		**25%**	**30%**	**35%**	**40%**		**35%**	**40%**	**45%**		**25%**
		1.014	**1.003**	1.004	1.047		0.949	**0.943**	0.958		**1.282**
		0.432	(0.027)	0.044	1.645		0.272	(0.023)	0.643		(0.037)
γ **= 1**		**30%**	**35%**	**40%**			**35%**	**40%**	**45%**		**30%**
		3.855	**3.843**	3.953			4.244	**4.157**	4.237		**4.167**
		0.128	(0.098)	1.120			0.927	(0.093)	0.849		(0.119)
Mean		**30%**	**35%**	**40%**			**35%**	**40%**	**45%**		**30%**
		1.731	**1.727**	1.782			1.846	**1.814**	1.848		**1.955**
		0.096	(0.036)	1.514			0.954	(0.034)	0.999		(0.044)

Table 3.3.2.10 Simulation results for $n = 200$.

		Median					Mode					Mean
γ = 0.1		**20%**	**25%**	**30%**	**35%**		**25%**	**30%**	**35%**	**40%**		**45%**
		0.035	0.034	0.035	**0.033**		**0.030**	0.030	0.030	0.030		**0.047**
		1.143	0.583	0.898	(0.002)		(0.002)	0.163	0.422	0.535		(0.002)
γ = 0.5		**30%**	**35%**	**40%**			**30%**	**35%**	**40%**			**35%**
		0.675	**0.666**	0.684			0.660	**0.637**	0.644			**0.846**
		0.347	(0.026)	0.721			1.003	(0.023)	0.284			(0.036)
γ = 1		**20%**	**25%**	**30%**	**35%**		**30%**	**35%**	**40%**			**30%**
		2.684	2.604	2.525	**2.522**		2.811	**2.635**	2.843			**2.863**
		1.740	0.877	0.023	(0.093)		1.991	(0.089)	2.351			(0.123)
Mean		**25%**	**30%**	**35%**	**40%**		**30%**	**35%**	**40%**			**30%**
		1.188	1.139	**1.135**	1.211		1.234	**1.164**	1.240			**1.333**
		1.567	0.128	(0.034)	2.222		2.177	(0.032)	2.347			(0.046)

Table 3.3.2.11 Simulation results for $n = 500$.

			Median					Mode				Mean
γ = 0.1	***k***		**20%**	**25%**	**30%**	**35%**		**20%**	**25%**	**30%**		**35%**
	100×MSE		0.025	**0.024**	0.024	0.025		0.022	0.022	**0.021**		**0.033**
			0.796	(0.001)	0.015	0.811		1.101	0.946	(0.001)		(0.002)
γ = 0.5	***k***		**25%**	**30%**	**35%**			**25%**	**30%**	**35%**		**30%**
	100×MSE		0.504	**0.471**	0.503			0.529	**0.488**	0.514		**0.594**
			1.880	(0.018)	1.850			2.478	(0.017)	1.579		(0.026)
γ = 1	***k***		**20%**	**25%**	**30%**	**35%**		**25%**	**30%**	**35%**		**30%**
	100×MSE		2.128	2.026	**2.016**	2.133		2.143	**2.121**	2.237		**2.297**
			1.444	0.135	(0.078)	1.500		0.290	(0.076)	1.530		(0.095)
Mean	***k***		**25%**	**30%**	**35%**			**25%**	**30%**	**35%**		**30%**
	100×MSE		0.900	**0.885**	0.938			0.950	**0.927**	0.978		**1.031**
			0.543	(0.028)	1.878			0.831	(0.027)	1.855		(0.035)

Table 3.3.2.12 Simulation results for $n = 1000$.

		Median				**Mode**					**Mean**
γ = 0.1		**25%**	**30%**	**35%**		**25%**	**30%**	**35%**			**40%**
		0.0174	0.0177	0.0195		**0.0166**	0.0175	0.0190			**0.0232**
		(0.0009)	0.3167	2.2988		(0.0007)	1.2188	3.4697			(0.0009)
γ = 0.5		**20%**	**25%**	**30%**		**20%**	**25%**	**30%**	**35%**		**25%**
		0.3617	**0.3431**	0.3479		0.3984	**0.3729**	0.3763	0.3949		**0.4158**
		1.5010	(0.0124)	0.3868		1.8785	(0.0136)	0.2557	1.6206		(0.0165)
γ = 1		**20%**	**25%**	**30%**		**20%**	**25%**	**30%**	**35%**		**30%**
		1.5326	1.5477	1.5592		1.7080	**1.7061**	1.7445	1.8020		**1.7365**
		(0.0631)	0.2390	0.4221		0.0318	(0.0598)	0.6425	1.6042		(0.0741)
Mean		**20%**	**25%**	**30%**		**20%**	**25%**	**30%**	**35%**		**30%**
		0.6743	**0.6725**	0.6783		0.7497	**0.7386**	0.7537	0.7809		**0.7680**
		0.0915	(0.0201)	0.2910		0.5135	(0.0216)	0.6946	1.9555		(0.0269)

Table 3.3.2.13 Simulation results for $n = 2000$.

		Median					**Mode**				**Mean**
γ = 0.1		**15%**	**20%**	**25%**	**30%**		**15%**	**20%**	**25%**		**25%**
		0.0121	0.0139	0.0122	0.0133		**0.0118**	0.0154	0.0129		**0.0151**
		(0.0009)	2.0834	0.1425	1.4171		(0.0008)	4.5770	1.4352		(0.0011)
γ = 0.5		**20%**	**25%**	**30%**			**15%**	**20%**	**25%**		**25%**
		0.2636	**0.2326**	0.2920			0.3133	0.2992	**0.2763**		**0.2556**
		2.7635	(0.0112)	5.2869			2.6946	1.6732	(0.0137)		(0.0130)
γ = 1		**15%**	**20%**	**25%**			**15%**	**20%**	**25%**		**20%**
		1.0662	**1.0434**	1.0815			1.2664	**1.1847**	1.2243		**1.1462**
		0.4324	(0.0527)	0.7237			1.3113	(0.0624)	0.6358		(0.0638)
Mean		**15%**	**20%**	**25%**			**15%**	**20%**	**25%**		**25%**
		0.4753	**0.4654**	0.4674			0.5610	**0.5283**	0.5334		**0.5018**
		0.5158	(0.0192)	0.1048			1.4459	(0.0226)	0.2274		(0.0212)

Table 3.3.2.14 Simulation results for $n = 5000$.

		Median					**Mode**					**Mean**
γ **= 0.1**		**10%**	**15%**	**20%**			**15%**	**20%**	**25%**			**30%**
		0.0108	**0.0087**	0.0094			0.0099	**0.0089**	0.0096			**0.0107**
		4.0319	(0.0005)	1.3851			1.8429	(0.0005)	1.3922			(0.0005)
γ **= 0.5**		**10%**	**15%**	**20%**	**25%**		**10%**	**15%**	**20%**	**25%**		**25%**
		0.2018	0.2130	**0.1930**	0.2110		0.2303	0.2497	**0.2192**	0.2448		**0.2170**
		0.7686	1.7506	(0.0114)	1.5744		0.9597	2.6356	(0.0116)	2.2098		(0.0149)
γ **= 1**		**15%**	**20%**	**25%**			**20%**	**25%**	**30%**			**25%**
		0.8228	0.9098	**0.7502**			1.0709	**0.8420**	1.0430			**0.8545**
		2.0566	4.5212	(0.0353)			5.1189	(0.0447)	4.4950			(0.0777)
Mean		**20%**	**25%**	**30%**			**20%**	**25%**	**30%**			**25%**
		0.3920	**0.3423**	0.4203			0.4579	**0.3864**	0.4524			**0.3814**
		3.7174	(0.0134)	5.8356			4.2201	(0.0169)	3.8934			(0.0279)

Table 3.3.2.15 Simulation results for $n = 10\,000$.

		Median				**Mode**				**Mean**
γ **= 0.1**		**20%**	**25%**	**30%**		**25%**	**30%**			**25%**
		0.0183	**0.0074**	0.0087		**0.0080**	0.0089			**0.0075**
		31.1838	(0.0003)	3.9464		(0.0004)	2.0888			(0.0003)
γ **= 0.5**		**15%**	**20%**	**25%**		**15%**	**20%**	**25%**		**20%**
		0.1526	0.1578	0.1702		**0.1661**	0.1842	0.1911		**0.1571**
		(0.0088)	0.5974	2.0041		(0.0090)	2.0207	2.7868		(0.0080)
γ **= 1**		**10%**	**15%**	**20%**		**10%**	**15%**	**20%**		**15%**
		0.5759	0.5978	0.6303		0.6852	**0.6758**	0.6779		**0.6546**
		(0.0314)	0.6958	1.7326		0.2449	(0.0385)	0.0560		(0.0501)
Mean		**10%**	**15%**	**20%**		**15%**	**20%**			**15%**
		0.2875	**0.2670**	0.2835		**0.3001**	0.3109			**0.2922**
		1.6640	(0.0123)	1.3411		(0.0140)	0.7718			(0.0181)

Table 3.3.2.16 Simulation results for $n = 20\,000$.

The results from Tables 3.3.2.8 to 3.3.2.16 are summarized in the following table:

Sample size (n)	**Optimal estimator**	**Optimal k (% of n)**	**Optimal k (number of excesses)**	**100×MSE**
50	Median	35%	18	3.102
100	Median	30%	30	2.225
200	Median	35%	70	1.727
500	Median	35%	175	1.135
1000	Median	30%	300	0.885
2000	Median	25%	500	0.676
5000	Median	20%	1000	0.465
10 000	Median	25%	2500	0.342
20 000	Median	15%	3000	0.267

Table 3.3.2.17 Summary of simulation results for all sample sizes.

Table 3.3.2.17 reports only the method of estimation which yielded the lowest overall MSE. Some features of these results should be noted:

1. The first and most obvious feature is that the average MSE decreases as the sample size increases. This is to be expected, since there is more information (a larger data set) available to estimate the EVI.
2. As the sample size increases, so does the optimal number of excesses. This is also to be expected, since more observations are available which can be regarded as being in the tail of the distribution.
3. As the sample size increases, k decreases as percentage of n, in general. This feature also makes sense. If we have a larger sample size, we can afford to consider a smaller proportion of observations as being in the tail. By doing this we decrease the bias (since the threshold is larger and therefore the limiting distribution fits better), and we still have enough observations to keep the variance in check.

Additional remarks

- Two conditions which are frequently stated when proving limiting results of threshold models, are that one should have $k \to \infty$ and $k/n \to 0$. Features 2 and 3 illustrate in a practical way that, in our sense of optimality, these conditions are adhered to.
- As we have already seen, there is a large number of difficult technical issues to be taken into account when estimating γ in the manner we propose in this chapter. However, the idea of obtaining a Bayesian estimate of the PPD parameter γ by Gibbs sampling does not in the least qualify as a concept that breaks new gound. Given the conceptual simplicity of this estimation method, the simulation results we obtained are quite remarkable, as will be seen in Chapter 4, when the results are compared to those obtained for a wide range of other estimators in the literature.

- We have shown in this section that the mean estimator of γ was always outperformed by either the median estimator or the mode estimator. This was true for all EVI groups and all overall MSEs, across all sample sizes. Therefore, the only estimators we needed to consider as a benchmark estimator, were estimators constructed as a combination of the median and mode estimators, as presented here.
- The detail of the results in terms of the EVI groups will not be discussed here. This will be discussed in the next section, where they will be used to develop a threshold selection technique.

3.4 Threshold selection

In Section 3.3 we showed simulation results obtained by fixing the number of excesses k to a proportion of the sample size. Our approach was to divide the range of k into equal parts by considering as possible values of k, $k = 5\%, 10\%, \dots, 95\%$ of n. We have seen that for $n = 500$, for instance, $k = 35\%$ of n yielded the lowest average MSE across all distributions.

We now introduce methods of *adaptive* threshold selection. This means that, instead of fixing k for a given sample size before any estimation is done, we adapt k for a given sample, based on estimates obtained from the sample.

In most cases an adaptive threshold selection technique returns an optimal k value, at which the estimate of the EVI is optimal, in a certain predefined sense. This approach is not the one we follow here. Instead we attempt to find a range or region of k values where the EVI estimates are stable. The estimate of the EVI is then defined as the mean of the EVI estimates over that region.

A great advantage of this method is that, since the estimate is the mean of several estimates, the variance is reduced considerably, leading to good results.

3.4.1 Quantifying the stability of estimates

Let us recall our motivation for considering second order estimates. We have seen that, by fitting the PPD (second order estimation) instead of the Pareto distribution (first order estimation), our estimates are more “stable”.

Figure 3.1.5 illustrated this idea. Figure 3.1.5 showed the MLEs of the EVI when fitting the PPD, using the set of 200 observations from a *Burr*$(\beta = 1, \gamma = 0.5, \rho = -0.7)$ distribution. We now construct a graph which is similar to Figure 3.1.5, using the same data set, but with the following differences:

- We computed the mode estimates of the EVI instead of the MLEs. (The mode estimator is defined in the same way as in the previous section. See also *Gibbs sampling terminology* in Section 3.2.4.)

- In Figure 3.4.1.1 $k = 10, 20, \dots, 150$ corresponding to $k = 5\%, 10\%, \dots, 75\%$ of n. (In Figure 3.1.5 we used smaller increments: $k = 2,3, \dots ,150$.)
- We have removed the line representing the true value (0.5) of the EVI.
- The least squares regression line (dashed) has been added.

The following graph is obtained:

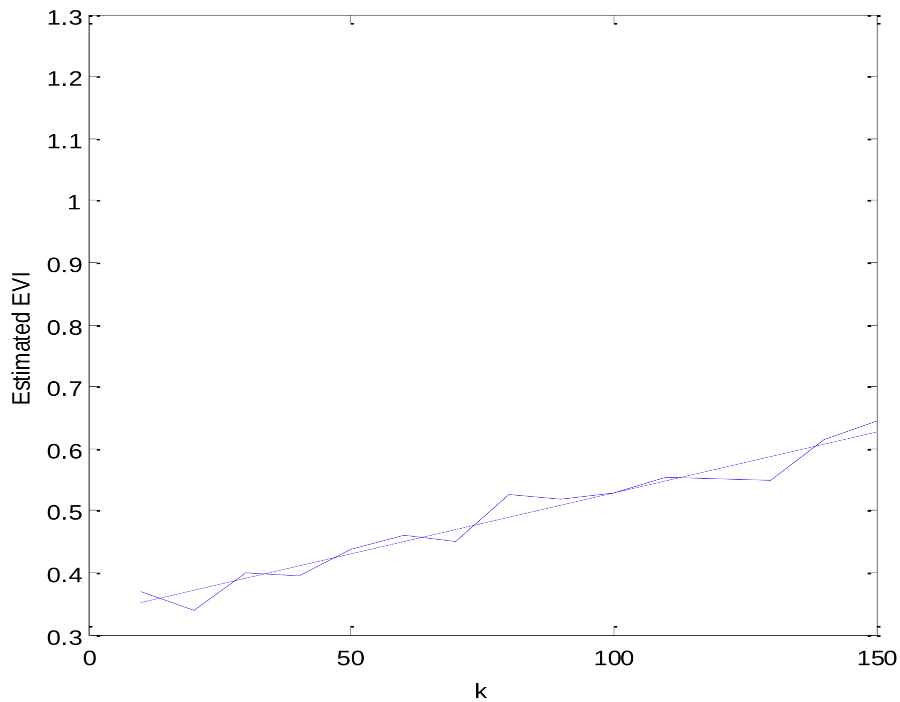

Figure 3.4.1.1 Mode estimates of the EVI for the Burr sample.

We stated in Section 3.1.5 that a larger number of excesses can be used when fitting the PPD, since for large values of k the estimates still have low bias, and also that the choice of threshold is not as critical, since for a wide range of values the estimates stay relatively close to the true value of the EVI.

Even though these statements are true when one compares Figure 3.1.5 or Figure 3.4.1.1 to the Hill plot (Figure 3.1.3), it is clear from Figure 3.4.1.1 that the choice of threshold is still anything but obvious, and the choice of threshold does affect the resulting EVI estimate quite substantially.

We will now quantify the stability of the estimates, so that the resulting measure can be used to obtain an objective choice of threshold.

For the sake of clarity and brevity, denote the estimated EVI values by $\{y_i\}$ and the corresponding intervals of k by $x = 1, 2, \dots, 15$.

We propose as measure of instability

$$\vartheta^2 = \sigma^2 + b^2,$$

where σ^2 is the variance of the y values and b the slope of the least squares regression line (of y on the indices x).

By a perfectly stable graph (in the context of estimates at different thresholds), we mean that $y_i = a$, for all i and some constant a. Graphs may deviate from perfect stability in two ways. Firstly, there may be random fluctuations, and secondly, the slope of the regression line through the points may not be zero.

The random fluctations are quantified by the variance of the y values. The slope b of the regression line can be viewed as a bias (systematic deviation from the perfectly stable graph).

The idea of the measure ϑ^2 is similar to that of an MSE, namely that it is, in the sense described above, the sum of the variance and the square of the bias.

Some additional remarks are in order:

- As mentioned before, in Figure 3.4.1.1 we do not show the line which indicated the true EVI (at 0.5). The reason for this is to stress that we are now only interested in defining *stability*. If we obtain, for example, $y_i = 0.2$, for all i, the estimates will also be regarded as perfectly stable.
- ϑ^2 is location invariant with respect to the x values, in the sense that the proportion of ϑ^2 contributed by σ^2 remains the same if a constant is added to all the x values.
- ϑ^2 is scale invariant with respect to the y values, in the sense that the proportion of ϑ^2 contributed by σ^2 remains the same if all the y values are multiplied by a constant.
- ϑ^2 is unfortunately not scale invariant (in the above sense) with respect to the x values. We found that it works well in practice if one always chooses $x = 1, 2, \ldots, m$, where m is the number of y values considered. This is the approach that was followed throughout.
- It may also be suggested that there is a double penalty for systematic deviation, in the sense that a straight line with slope b should have $\sigma^2 = 0$. In other words σ^2 should be defined, not as the variance of the y values, but as the variance of the residuals after fitting the regression line. This option was explored, but does not work well in practice. Consider the two sets of estimates in Figure 3.4.1.2 below.

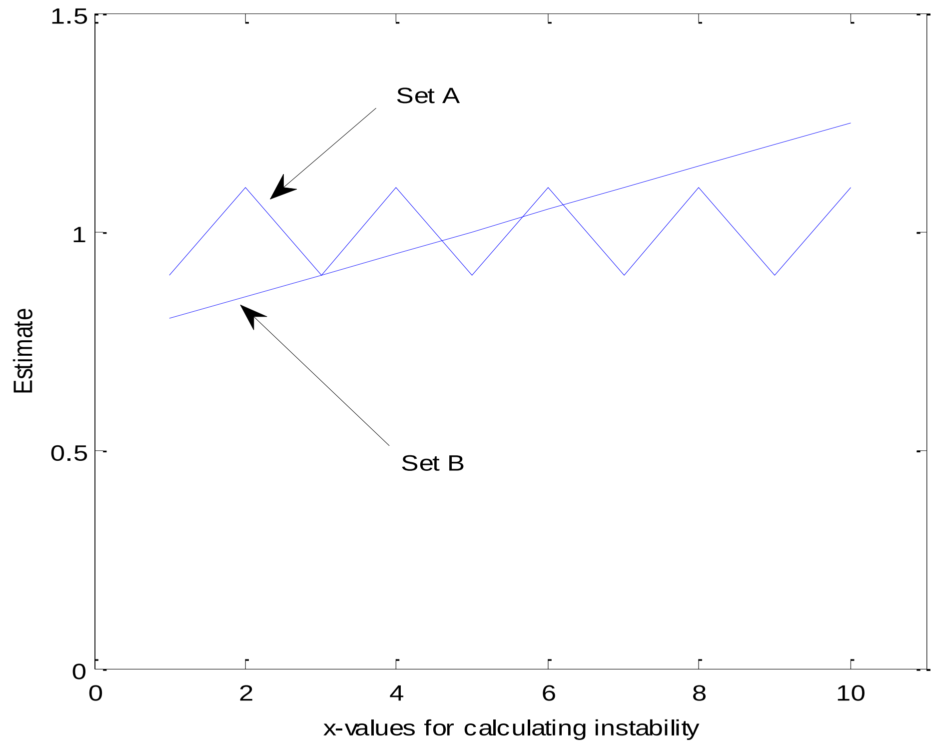

Figure 3.4.1.2 Two contrasting sets of estimates.

If asked which of these two sets of estimates is more stable for the purpose of estimation, Set A would be the obvious choice, since it is quite clear that one can safely choose the value of the final estimate around 1, with the inherent variance of the estimates causing fluctuations around that value. For Set B it is very difficult to decide what would be regarded as the final estimate, since the value of the estimate is completely dependent on the choice of the threshold (value of x).

The proposed method returns as instability estimates 0.0111 for Set A, and 0.0254 for Set B. Therefore Set A is considered better in the sense that it has lower instability. When applying the method of calculating the variance of the residuals rather than that of the y values, the instability estimates are 0.0108 for Set A, and 0.0025 for Set B, which wrongly indicates Set B as preferable. Mathematically, the problem is that the contribution of b^2 as a proportion of ϑ^2 is too large in this case. Furthermore, the contribution of b^2 as a proportion of ϑ^2 now also varies as a function of the scale of the y-axis, since the scale invariance property has been compromised.

- As a final remark note that in the context of EVI estimation simple linear regression may not be optimal. An example in Section 3.1.3 illustrated how the variance of the Hill estimates decreases as k increases. This is true in general for all EVI estimators. In such a setting where σ^2 is a function of x, applying weighted regression may be more appropriate and could lead to better results. This refinement needs to be considered in future research.

3.4.2 Methods for finding a stable region

Based on the definition of our measure of instability ϑ^2 in Section 3.4.1, we now consider three methods of finding a stable region of EVI estimates.

Note that we have calculated the EVI estimates at $k = 5\%, 10\%, \dots, 95\%$ of n, for all sample sizes. There are therefore 19 values of k which are considered, denoted by $k_1, k_2, \dots, k_{19}$. For the sake of brevity, an interval of four (say) consecutive k values, for example an interval including k_4, k_5, k_6, and k_7, will be referred to as a *region of length* four.

Also note that we now extend Figure 3.4.1.1 to include values of k up to $k = 190 = 95\%$ of n.

The crude grid of 19 values of k ($k = 5\%, 10\%, \dots, 95\%$ of n) can obviously be refined. In this thesis such a refinement was not investigated, except for very large samples. In Section 3.4.4 the necessity of refining the grid for samples of sizes $n = 20\,000$, $n = 50\,000$, and $n = 100\,000$ will be argued. For these sample sizes the grid of k values will be chosen as $k = 1\%, 2\%, \dots$, 30% of n, leading to improved results.

Method 1: Fixed region length

The first method entails fixing the length of the region. Suppose we fix the region length at 12. We then calculate ϑ^2 for the region $k_1, \dots, k_{12}$, the region $k_2, \dots, k_{13}$, up to the region $k_8, \dots, k_{19}$. The region for which ϑ^2 is a minimum, is regarded as the optimal region.

The following table shows the calculation of ϑ^2 for a fixed region length of 12, applied to the mode estimates of the Burr sample of Figure 3.4.1.1:

	Region:	1 to 12		2 to 13		3 to 14		4 to 15		5 to 16		6 to 17	
Index	**EVI**	***x***	***y***	***x***	***y***	***x***	***y***	***x***	***y***	***x***	***y***	***x***	***y***
1	0.3558	1	0.3558										
2	0.3656	2	0.3656	1	0.3656								
3	0.3823	3	0.3823	2	0.3823	1	0.3823						
4	0.4530	4	0.4530	3	0.4530	2	0.4530	1	0.4530				
5	0.4105	5	0.4105	4	0.4105	3	0.4105	2	0.4105	1	0.4105		
6	0.4683	6	0.4683	5	0.4683	4	0.4683	3	0.4683	2	0.4683	1	0.4683
7	0.4355	7	0.4355	6	0.4355	5	0.4355	4	0.4355	3	0.4355	2	0.4355
8	0.4841	8	0.4841	7	0.4841	6	0.4841	5	0.4841	4	0.4841	3	0.4841
9	0.5153	9	0.5153	8	0.5153	7	0.5153	6	0.5153	5	0.5153	4	0.5153
10	0.5676	10	0.5676	9	0.5676	8	0.5676	7	0.5676	6	0.5676	5	0.5676
11	0.5708	11	0.5708	10	0.5708	9	0.5708	8	0.5708	7	0.5708	6	0.5708
12	0.5657	12	0.5657	11	0.5657	10	0.5657	9	0.5657	8	0.5657	7	0.5657
13	0.5972			12	0.5972	11	0.5972	10	0.5972	9	0.5972	8	0.5972
14	0.6236					12	0.6236	11	0.6236	10	0.6236	9	0.6236
15	0.6379							12	0.6379	11	0.6379	10	0.6379
16	0.8480									12	0.8480	11	0.8480
17	0.8254											12	0.8254
18	0.9246												
19	1.0151												
	$\boldsymbol{b}$		0.0208		0.021		0.0209		0.0203		0.0293		0.0327
	$\boldsymbol{\sigma^2}$		0.0061		0.0062		0.0062		0.0059		0.0135		0.0166
	$\boldsymbol{\vartheta^2}$		0.0066		0.0066		0.0066		**0.0063**		0.0144		0.0176
	$\boldsymbol{\hat{\gamma}}$		0.4645		0.4846		0.5061		**0.5274**		0.5604		0.5949

Table 3.4.2 Finding the optimal region with fixed length of 12.

The first column shows the index i of k_i, for $k = k_1, \dots, k_{19} = 5\%, 10\%, \dots, 95\%$ of n. The second column shows the estimated EVI corresponding to that value of k. The first row shows how the fixed region shifts from $k_1, \dots, k_{12}$ to the region $k_2, \dots, k_{13}$, and up to the region $k_8, \dots, k_{19}$. We omitted the last two regions due to a lack of space.

From column three onwards, we show the x and y coordinates which we used to calculate the slope b. The last four rows show the slope, the variance of the y values σ^2, the measure of instability $\vartheta^2 = \sigma^2 + b^2$, and the estimated EVI, which is the mean of the y values.

The numbers in bold correspond to the region $k_4, \dots, k_{15}$. This region yielded the lowest instability (0.0063). The corresponding EVI estimate is 0.5274.

Graphically, we can represent the result of this table as follows:

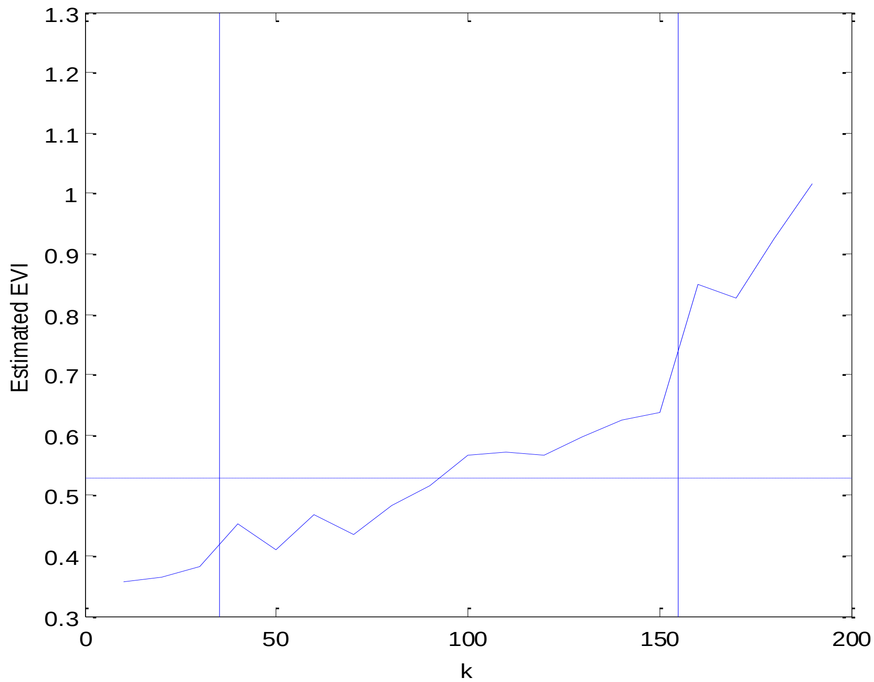

Figure 3.4.2.1 Fixed stable region length for the Burr sample.

The vertical lines represent the region which yielded the lowest value of ϑ^2. (The region was from $k = 40$ to $k = 150$. We subtracted half an interval from the lower boundary and added half an interval to the upper boundary, to show all the EVI estimates included in the region.) The dashed line shows the final estimated EVI, which was 0.5274. (Calculated as the mean of the estimates from $k = 40$ to $k = 150$.)

Method 2: Reducing the region until instability increases

This method calculates ϑ^2over the enitre region $k_1, \ldots, k_{19}$. Either the first or the last value in the region is left out, depending on which reduction in the region decreases ϑ^2 by the greatest amount. This process is repeated until neither reduction of the region decreases the instability ϑ^2.

When applying this method to our example, we obtain the following:

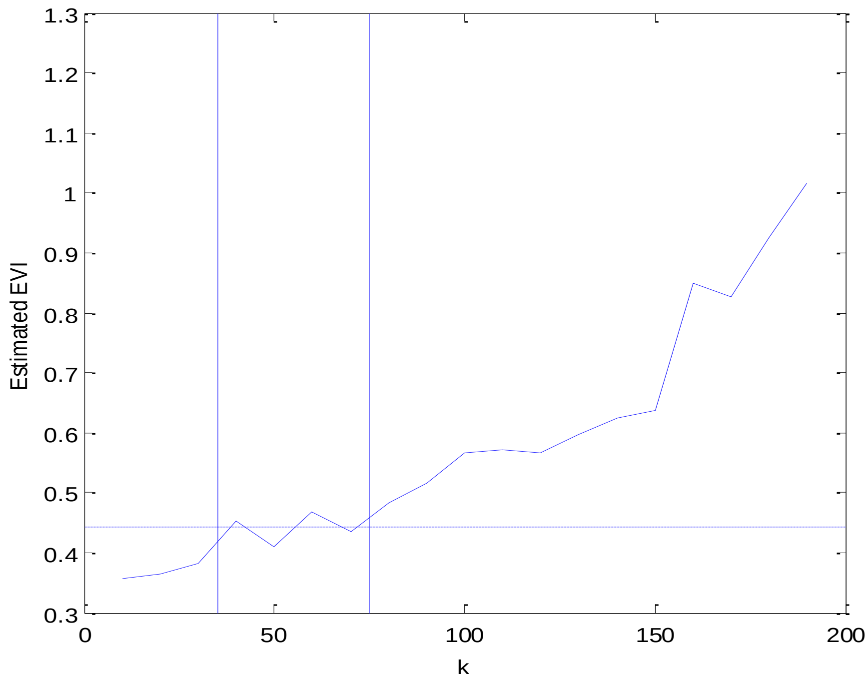

Figure 3.4.2.2 Reducing the region for the Burr sample.

Here the final region is from $k = 40$ to $k = 70$, yielding an EVI estimate of 0.4418. This method seems to find a stable region which corresponds closely to what one would select graphically, but we will see from the simulation results that a method which selects a wider region works better for small sample sizes (such as $n = 200$).

Method 3: Fixing the upper limit and reducing the region from the left

This method is somewhat counter-intuitive, but seems to work best for large sample sizes (for $n = 20\ 000$ as we will see from simulation results).

We fix the upper limit of the region, say at 10 (k_{10}), and then leave out the first value in the region until ϑ^2 does not decrease any more.

For a fixed upper limit of 10, we get the following for the Burr sample:

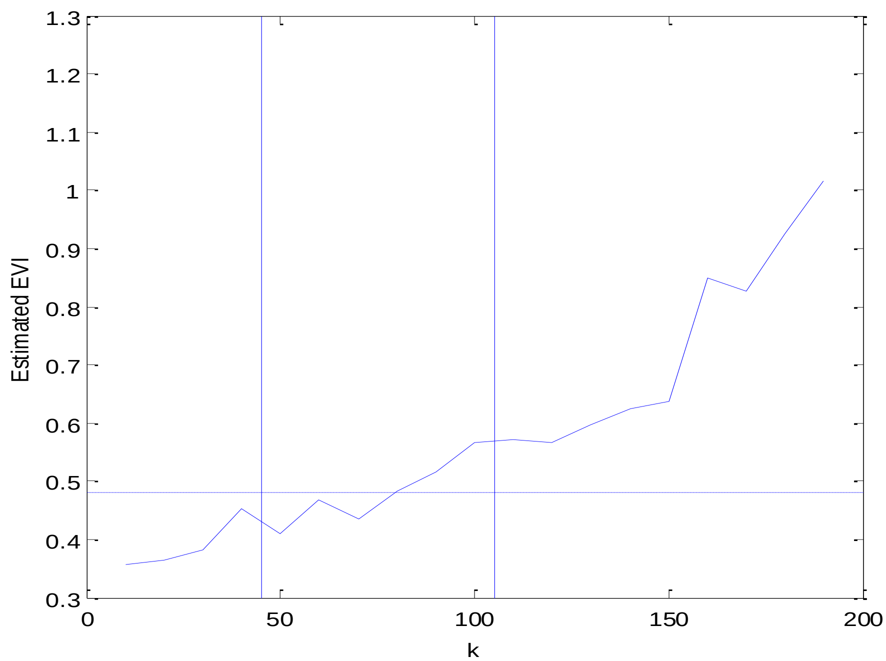

Figure 3.4.2.3 Reducing the region with fixed upper limit at 10 ($k = 100$).

We find the optimal lower limit to be at $k = 50$. The resulting estimate of the EVI is 0.4802.

3.4.3 Simulation results

For samples of size $n = 50$ we obtain the following table, followed by a detailed description of the contents:

Method:		**Fixed**	**1**	**1**		**2**	**2**	**3**	**3**
			Median	**Mode**		**Median**	**Mode**	**Median**	**Mode**
γ **= 0.1**	**Par**	**40%**	**12**	**13**	**14**			**8**	**10**
	MSE	**0.0803**	**0.1124**	0.0760	**0.0759**	**0.1169**	**0.0902**	**0.0982**	**0.0751**
		Mode	28.6923	0.6923	0.6154	32.1538	11.6154	17.7692	(0.0013)
γ **= 0.5**	**Par**	**40%**	**12**	**13**	**14**			**8**	**11**
	MSE	**1.7682**	**2.0792**	**1.6296**	1.6332	**2.3074**	**2.0015**	**1.8103**	**1.6546**
		Mode	14.8383	(0.0303)	0.1188	22.3696	12.2739	5.9637	0.8251
γ **= 1**	**Par**	**35%**	**13**	**13**	**14**			**10**	**12**
	MSE	**6.7773**	**6.8575**	6.8799	6.8374	**7.9987**	**8.1963**	**6.6706**	**7.3615**
		Median	1.7006	1.9045	1.5177	12.0846	13.8826	(0.1099)	6.2866
Mean	**Par**	**35%**	**12**	**13**	**14**			**10**	**12**
	MSE	**3.1024**	**3.1920**	3.0257	**3.0120**	**3.6718**	**3.6257**	**3.0347**	**3.2088**
		Median	4.3584	0.3324	(0.0413)	15.9758	14.8596	0.5496	4.7651

Table 3.4.3.1 Threshold methods for samples of size $n = 50$.

In this section we present the simulation results in the format similar to that of the above table.

The features of the table are the following:

- We limit the results to the minimum of what is relevant for the purpose of finding an optimal estimation method.
- In the first two rows we show which threshold technique was used. “Fixed” refers to the results obtained from the fixed threshold methods applied in Section 3.3. The numbers 1, 2 and 3 correspond to the method numbering in Section 3.4.2. The second row indicates which estimator was used.
- The column under the heading “Fixed” contains three entries per EVI group, namely the fixed proportion of sample size yielding the lowest MSE, the corresponding 100×MSE below that, and then the estimator which yielded the lowest MSE, namely the mode or median estimator. Note that we show here only the single best results from Section 3.3, namely the minimum 100×MSE obtained over all estimators (mean, median and mode) and over all fixed thresholds ($k = 5\%, 10\%, \dots, 95\%$ of n). We do not include distance in standard error from the minimum MSE for this column, since this column is never in contention for the optimal method.
- From column four onwards there are three entries for each EVI group, namely the parameter (Par), 100×MSE and distance from the minimum MSE in standard errors.
- For Method 1 the parameter is the fixed region length. Method 2 has no parameter. For Method 3 the parameter is the fixed upper limit, for example a value of 10 will indicate that k_{10} is the fixed upper limit.
- The 100×MSE rows (MSE) display in underlined bold the overall lowest 100×MSE, i.e. over all methods (fixed theshold and Methods 1, 2 and 3), over all estimators (mean, median and mode estimators) and over all choices of parameter. For example, for $\gamma = 0.1$ the minimum of all values in the fourth row of the table, was 0.0751, indicated in bold, underlined text. This means that the technique which yielded the lowest MSE of all methods considered, was using the mode estimator and applying Method 3, with a fixed upper limit of k_{10}. Bold only indicates the minimum for that specific method (shown in the column heading) and all other results are shown in standard format. For example, for $\gamma = 0.1$ a region length of 14 yielded the lowest 100×MSE for the mode estimator and Method 1 (columns five and six). This value was 0.0759, indicated in bold. For the mode estimator and Method 1, a region length of 13 yielded a 100×MSE of 0.0760, which is not in bold, since it is not the minimum for the mode estimator and Method 1.
- The rows below the MSE rows, show the standard error of the overall best method in brackets. All other entries are the distances of the MSE in standard errors from the minimum MSE. For instance

$$\frac{0.0902 - 0.0751}{0.0013} = 11.6154.$$

We now have to draw conclusions from these results. We have to decide on a method of estimation which we will regard as our benchmark method for estimating the EVI for samples of size $n = 50$. We always attempt to make the rule as simple as possible. In other words, we do not specify three different rules, one for $\gamma \approx 0.1$, one for $\gamma \approx 0.5$ and one for $\gamma \approx 1$. Instead we choose a single rule, if possible. (This is not always possible. See the results for $n = 1000$, for instance.)

For $n = 50$ we choose as our benchmark estimator the mode estimator, Method 1 and a region length of 14. This estimator yields the lowest overall MSE over all distributions (as can be seen from the second last row in the table). Furthermore, this method yields an MSE within two standard errors from the minimum MSE, for all EVI groups, namely 0.6154 standard errors for $\gamma = 0.1$, 0.1188 standard errors for $\gamma = 0.5$, and 1.5177 standard errors for $\gamma = 1$. Also, as will be seen subsequently in Table 3.4.4.1, this choice is consistent with respect to the optimal region length for samples of size $n = 100$, $n = 200$ and $n = 500$.

For $n = 100$, we have the following results:

Method:		**Fixed**	**1**	**1**		**2**	**2**	**3**	**3**
			Median	**Mode**		**Median**	**Mode**	**Median**	**Mode**
γ **= 0.1**	**Par**	**35%**	**12**	**13**	**14**			**7**	**10**
	MSE	**0.0580**	**0.0719**	**<u>0.0529</u>**	0.0536	**0.0763**	**0.0608**	**0.0637**	**0.0533**
		Mode	12.6667	(0.0015)	0.4112	15.6000	5.2667	7.2000	0.2667
γ **= 0.5**	**Par**	**40%**	**13**	**13**	**14**			**8**	**11**
	MSE	**1.2650**	**1.3632**	1.1306	**<u>1.1258</u>**	**1.5437**	**1.3696**	**1.2538**	**1.1738**
		Mode	8.7601	0.1742	(0.0271)	15.4207	8.9963	4.7232	1.7712
γ **= 1**	**Par**	**30%**	**12**	**13**	**14**			**9**	**12**
	MSE	**4.9676**	**<u>4.8579</u>**	4.9618	4.9188	**5.6578**	**5.8676**	**4.9292**	**5.4398**
		Median	0.1189	0.8738	0.5122	6.7275	8.4920	0.5997	4.8940
Mean	**Par**	**30%**	**12**	**13**	**14**			**9**	**12**
	MSE	**2.2254**	**2.2190**	2.1658	**<u>2.1491</u>**	**2.5641**	**2.5722**	**2.2035**	**2.3552**
		Median	1.8640	0.4450	(0.0375)	11.0667	11.2827	1.4507	5.4960

Table 3.4.3.2 Threshold methods for samples of size $n = 100$.

Here the two methods which perform best, are mode estimators, Method 1, with region lengths of 13 and 14, respectively. Even though a length of 14 marginally outperforms one of 13, we select as our benchmark estimator a region length of 13 for the sake of consistency with respect to the optimal region length for samples of size $n = 50$, $n = 200$ and $n = 500$, as indicated in Table 3.4.4.1.

For $n = 200$, we have the following results:

Method:		**Fixed**	**1**	**1**		**2**	**2**	**3**	**3**
			Median	**Mode**		**Median**	**Mode**	**Median**	**Mode**
γ **= 0.1**	**Par**	**35%**	**11**	**12**	**13**			**7**	**10**
	MSE	**0.0451**	**0.0531**	**0.0399**	0.0405	**0.0545**	**0.0426**	**0.0522**	**0.042**
		Mode	0.0016	(0.0011)	0.5699	13.2727	2.4545	11.1818	1.9091
γ **= 0.5**	**Par**	**40%**	**11**	**12**	**13**			**9**	**10**
	MSE	**0.9428**	**0.9621**	**0.8134**	0.8152	**1.0267**	**0.9344**	**0.9485**	**0.8682**
		Mode	0.0269	(0.0202)	0.0915	10.5594	5.9901	6.6881	2.7129
γ **= 1**	**Par**	**35%**	**11**	**12**	**13**			**9**	**11**
	MSE	**3.8429**	**3.5795**	**3.6134**	3.6792	**3.8284**	**3.9483**	**3.7152**	**4.0098**
		Median	(0.1002)	0.3383	0.9950	2.4840	3.6806	1.3543	4.2944
Mean	**Par**	**35%**	**11**	**12**	**13**			**9**	**11**
	MSE	**1.7270**	**1.6185**	**1.5741**	1.5982	**1.7296**	**1.7358**	**1.6617**	**1.7354**
		Median	0.0366	(0.0306)	0.7864	5.0817	5.2843	2.8627	5.2712

Table 3.4.3.3 Threshold methods for samples of size $n = 200$.

From the above table it is clear that the optimal method is the mode estimator, Method 1, with region length of 12.

For samples of size $n = 500$:

Method:		**Fixed**	**1**	**1**		**2**	**2**	**3**	**3**
			Median	**Mode**		**Median**	**Mode**	**Median**	**Mode**
γ **= 0.1**	**Par**	**25%**	**8**	**10**	**11**			**8**	**8**
	MSE	**0.0296**	**0.0276**	**0.0246**	0.0251	**0.0278**	**0.0254**	**0.0312**	**0.0262**
		Mode	3.0000	(0.0010)	0.4890	3.2000	0.8000	6.6000	1.6000
γ **= 0.5**	**Par**	**35%**	**9**	**10**	**11**			**9**	**9**
	MSE	**0.6370**	**0.5596**	0.5491	**0.5414**	**0.609**	**0.5826**	**0.6407**	**0.5874**
		Mode	0.9286	0.3935	(0.0196)	3.4490	2.1020	5.0663	2.3469
γ **= 1**	**Par**	**35%**	**6**	**10**	**11**			**8**	**9**
	MSE	**2.5224**	**2.3095**	2.3882	**2.3324**	**2.3205**	**2.3843**	**2.3918**	**2.5662**
		Median	(0.0952)	0.8267	0.2405	0.1155	0.7857	0.8645	2.6964
Mean	**Par**	**35%**	**6**	**10**	**11**			**8**	**9**
	MSE	**1.1349**	**1.0331**	1.0440	**1.0217**	**1.0421**	**1.0546**	**1.0825**	**1.1208**
		Median	0.4057	0.7919	(0.0281)	0.7260	1.1708	2.1637	3.5267

Table 3.4.3.4 Threshold methods for samples of size $n = 500$.

For samples of size $n = 500$, we choose as optimal method the mode estimator, Method 1, with region length of 11.

For samples of size $n = 1000$:

Method:		**Fixed**	**1**	**1**	**2**	**2**	**3**	**3**
			Median	**Mode**	**Median**	**Mode**	**Median**	**Mode**
γ **= 0.1**	**Par**	**30%**	**7**	**10**			**8**	**7**
	MSE	**0.0212**	**0.0194**	**0.0190**	**0.0197**	**0.0187**	**0.0221**	**0.0192**
		Mode	0.7778	0.3333	1.1111	(0.0009)	3.7778	0.5556
γ **= 0.5**	**Par**	**30%**	**9**	**10**			**8**	**8**
	MSE	**0.4709**	**0.4361**	**0.4211**	**0.4211**	**0.4232**	**0.4528**	**0.4490**
		Median	0.9036	0.0000	(0.0166)	0.1265	1.9096	1.6807
γ **= 1**	**Par**	**30%**	**8**	**10**			**7**	**8**
	MSE	**2.0159**	**1.8216**	**1.8772**	**1.7479**	**1.8869**	**1.9389**	**1.9888**
		Median	1.1516	2.0203	(0.0640)	2.1719	2.9844	3.7641
Mean	**Par**	**30%**	**9**	**10**			**7**	**8**
	MSE	**0.8847**	**0.8037**	**0.8167**	**0.7713**	**0.8208**	**0.8528**	**0.8661**
		Median	1.3846	1.9402	(0.0234)	2.1154	3.4829	4.0513

Table 3.4.3.5 Threshold methods for samples of size $n = 1000$.

For sample size $n = 1000$, we use a more complex, two stage method to estimate the EVI. We can see that Method 2 is the method of choice, but for small values of the EVI (in the vicinity of 0.1) the mode estimator performs better, and for larger EVIs, the median estimator performs better.

Up to this point we have in fact clearly seen this tendency, namely that the median estimator performs better for larger EVI. Note that the overall best estimator for the $\gamma = 1$ group is always the median estimator. This is true for all sample sizes, as will be seen in the results to follow. Also, the larger the sample size, the earlier this transition occurs (at a smaller value of EVI).

There always seems to be a point, in terms of the true underlying EVI, when the optimal estimator switches from one based on the mode estimator to one based on the median estimator. For samples of size $n = 1000$ this seems to be somewhere between $\gamma = 0.1$ and $\gamma = 0.5$. We want to make use of this knowledge in our construction of our benchmark estimator.

We therefore define our estimator of the EVI for samples of size $n = 1000$ as follows:

1. Estimate the EVI by applying Method 2 to the median estimator.
2. If the estimated EVI is above 0.3, the estimate remains unchanged. Otherwise, estimate the EVI by applying Method 2 to the mode estimator.

For samples of size $n = 2000$:

Method:		**Fixed**	**1**	**1**	**2**	**2**	**3**	**3**
			Median	**Mode**	**Median**	**Mode**	**Median**	**Mode**
γ = 0.1	**Par**	**25%**	**8**	**10**			**7**	**7**
	MSE	**0.0166**	**0.0147**	**0.0146**	**0.0144**	**<u>0.0141</u>**	**0.016**	**0.0144**
		Mode	1.0000	0.8333	0.5000	(0.0006)	3.1667	0.5000
γ = 0.5	**Par**	**25%**	**9**	**11**			**7**	**7**
	MSE	**0.3431**	**0.3036**	**0.3061**	**<u>0.3018</u>**	**0.3027**	**0.3158**	**0.3168**
		Median	0.1440	0.3440	(0.0125)	0.0720	1.1200	1.2000
γ = 1	**Par**	**20%**	**8**	**8**			**7**	**7**
	MSE	**1.5326**	**1.3869**	**1.4845**	**<u>1.3613</u>**	**1.4186**	**1.4222**	**1.4924**
		Median	0.4895	2.3556	(0.0523)	1.0956	1.1644	2.5067
Mean	**Par**	**25%**	**8**	**9**			**7**	**7**
	MSE	**0.6725**	**0.6038**	**0.6428**	**<u>0.5912</u>**	**0.6117**	**0.6181**	**0.6428**
		Median	0.6632	2.7158	(0.0190)	1.0789	1.4158	2.7158

Table 3.4.3.6 Threshold methods for samples of size $n = 2000$.

We define our estimator of the EVI for samples of size $n = 2000$ in the same way as we did for samples of size $n = 1000$:

1. Estimate the EVI by applying Method 2 to the median estimator.
2. If the estimated EVI is above 0.3, the estimate remains unchanged. Otherwise, estimate the EVI by applying Method 2 to the mode estimator.

For samples of size $n = 5000$:

Method:		**Fixed**	**1**	**1**	**2**	**2**	**3**	**3**
			Median	**Mode**	**Median**	**Mode**	**Median**	**Mode**
γ = 0.1	**Par**	**15%**	**6**	**9**			**5**	**6**
	MSE	**0.0118**	**0.0116**	**0.0112**	**0.0106**	**<u>0.0104</u>**	**0.011**	**0.0107**
		Mode	2.0000	1.3333	0.3333	(0.0006)	1.0000	0.5000
γ = 0.5	**Par**	**25%**	**7**	**8**	**0**	**0**	**6**	**7**
	MSE	**0.2326**	**0.2273**	**0.2339**	**<u>0.2164</u>**	**0.2211**	**0.2312**	**0.2361**
		Median	1.0583	1.6990	(0.0103)	0.4563	1.4369	1.9126
γ = 1	**Par**	**20%**	**8**	**9**	**0**	**0**	**5**	**6**
	MSE	**1.0434**	**1.0072**	**0.9967**	**<u>0.9289</u>**	**0.9409**	**0.9617**	**1.0181**
		Median	1.7836	1.5444	(0.0439)	0.2733	0.7472	2.0319
Mean	**Par**	**20%**	**7**	**9**	**0**	**0**	**5**	**6**
	MSE	**0.4654**	**0.4398**	**0.4384**	**<u>0.4074</u>**	**0.4132**	**0.4259**	**0.4478**
		Median	2.0377	1.9497	(0.0159)	0.3648	1.1635	2.5409

Table 3.4.3.7 Threshold methods for samples of size $n = 5000$.

We define our estimator of the EVI for samples of size $n = 5000$ in the same way as we did for the previous two sample sizes:

1. Estimate the EVI by applying Method 2 to the median estimator.
2. If the estimated EVI is above 0.3, the estimate remains unchanged. Otherwise, estimate the EVI by applying Method 2 to the mode estimator.

For samples of size $n = 10\,000$:

Method:		**Fixed**	**1**	**1**	**2**	**2**	**3**	**3**
			Median	**Mode**	**Median**	**Mode**	**Median**	**Mode**
γ **= 0.1**	**Par**	**15%**	**7**	**6**			**5**	**6**
	MSE	**0.0087**	**0.0084**	**0.0084**	**0.0076**	**<u>0.0075</u>**	**0.0078**	**0.0076**
		Median	3.0000	3.0000	0.3333	(0.0003)	1.0000	0.3333
γ **= 0.5**	**Par**	**20%**	**8**	**8**			**5**	**6**
	MSE	**0.1930**	**0.1913**	**0.1931**	**<u>0.1715</u>**	**0.1776**	**0.1728**	**0.1777**
		Median	2.4750	2.7000	(0.0080)	0.7625	0.1625	0.7750
γ **= 1**	**Par**	**25%**	**7**	**8**			**5**	**5**
	MSE	**0.7502**	**0.7189**	**0.7014**	**<u>0.6668</u>**	**0.6929**	**0.6892**	**0.7142**
		Median	1.7082	1.1344	(0.0305)	0.8557	0.7344	1.5541
Mean	**Par**	**25%**	**7**	**8**			**5**	**6**
	MSE	**0.3423**	**0.3243**	**0.3182**	**<u>0.2981</u>**	**0.3095**	**0.3065**	**0.3175**
		Median	2.3604	1.8108	(0.0111)	1.0270	0.7568	1.7477

Table 3.4.3.8 Threshold methods for samples of size $n = 10\,000$.

We define our estimator of the EVI for samples of size $n = 10\,000$ in the same way as we did for the previous three sample sizes:

1. Estimate the EVI by applying Method 2 to the median estimator.
2. If the estimated EVI is above 0.3, the estimate remains unchanged. Otherwise, estimate the EVI by applying Method 2 to the mode estimator.

For samples of size $n = 20\,000$:

Method:		Fixed	1	1	2	2	3		3
			Median	**Mode**	**Median**	**Mode**	**Median**		**Mode**
γ **= 0.1**	**Par**	**25%**	**6**	**6**			**3**	**4**	**6**
	MSE	**0.0074**	**0.0072**	**0.0074**	**0.0064**	**<u>0.0062</u>**	**0.0064**	**0.0076**	**0.0069**
		Median	2.5000	3.0000	0.5000	(0.0004)	0.5000	3.5000	1.7500
γ **= 0.5**	**Par**	**15%**	**4**	**5**			**4**	**5**	**6**
	MSE	**0.1526**	**0.1620**	**0.1627**	**0.1410**	**0.1458**	**0.1388**	**<u>0.1359</u>**	**0.1450**
		Median	3.7826	3.8841	0.7391	1.4348	0.4203	(0.0069)	1.3188
γ **= 1**	**Par**	**10%**	**7**	**7**			**4**	**5**	**4**
	MSE	**0.5759**	**0.6206**	**0.6416**	**0.5637**	**0.5762**	**<u>0.5145</u>**	**0.5257**	**0.5406**
		Median	4.4025	5.2739	2.0415	2.5602	(0.0241)	0.4647	1.0830
Mean	**Par**	**15%**	**7**	**7**			**4**	**5**	**5**
	MSE	**0.2670**	**0.2789**	**0.2885**	**0.2506**	**0.2567**	**<u>0.2328</u>**	**0.2356**	**0.2476**
		Median	5.1222	6.1889	1.9778	2.6556	(0.0090)	0.3111	1.6444

Table 3.4.3.9 Threshold methods for samples of size $n = 20\,000$.

We define our estimator of the EVI for samples of size $n = 20\,000$ as follows:

1. Estimate the EVI by applying Method 3 to the median estimator, with fixed upper limit of 4.
2. If the estimated EVI is above 0.3, the estimate remains unchanged. Otherwise, estimate the EVI by applying Method 2 to the mode estimator.

Let us disregard the EVI group $\gamma = 0.1$ for a moment. The shift from Method 2 to Method 3 (in moving from $n = 10\ 000$ to $n = 20\ 000$) suggests in a sense that the sample size becomes so large that any estimates with a threshold beyond $k = 20\%$ of n, need not be considered at all.

While it is a pity we cannot apply the same method of estimation for all samples of size $n = 1000$ or greater, this fixed upper limit of k_4 does limit estimation time. Our estimation method implies that we only have to determine estimates for the first four intervals of k, unless of course the resulting estimate is so small that we have to apply Method 2.

3.4.4 The benchmark estimator

We now summarize the results obtained in the previous section, defining our benchmark estimator, which we will call the *PPD estimator*.

The construction of the PPD estimator entails the following:

1. Given a sample of size n, calculate k multiplicative excesses at 19 values of k, namely $k = 5\%, 10\%, \dots, 95\%$ of n.
2. At each of these values, fit the PPD to the multiplicative excesses. Do this by determining the Bayesian estimates of γ, ρ and c, using Gibbs sampling.

3. This yields two sets of 19 estimates of the EVI γ. One set contains the mode estimates, which we obtain by calculating the mode of the Gibbs sample of the draws from the marginal posterior distribution of γ. (Refer to Section 3.2.) The other set contains the median estimates, obtained by calculating the median of the Gibbs draws.
4. To one of these sets (depending on the sample size), we apply one of the three methods (Methods 1 to 3 described in Section 3.4.2) to obtain a reduced set of estimates (estimates inside a stable region). The mean of this reduced set is then the proposed benchmark estimate of the EVI.

The procedure to be followed for different sample sizes will now be given.

For samples of sizes $n = 50$, 100, 200 and 500, we apply Method 1 to the set of mode estimates, with region length specified in the following table:

Sample size	50	100	200	500
Region length	14	13	12	11

Table 3.4.4.1 Region length for samples of size $n = 500$ or less.

For samples of sizes $n = 1000$, 2000, 5000 and $10\ 000$, we apply Method 2 to the set of median estimates. If the estimated EVI is above 0.3, the estimate remains unchanged. Otherwise, estimate the EVI by applying Method 2 to the set of mode estimates.

For samples of size $n = 20\ 000$, apply Method 3 to the set of median estimates, with a fixed upper limit of 4. If the estimated EVI is above 0.3, the estimate remains unchanged. Otherwise, estimate the EVI by applying Method 2 to the set of mode estimates.

The results obtained when using the PPD estimator, are shown in the tables below. We also include the results from Section 3.3, namely the best results when using a fixed threshold. Also included are the root mean square errors (RMSEs) of the PPD estimator, to ease comparison with results in other papers. As before, a value in brackets is the standard error of the value directly above it.

For the fixed threshold results we do not include the fixed proportion of sample size yielding the lowest 100×MSE and also not the estimator (mode or median) which yielded the lowest 100×MSE, due to lack of space. This information can be obtained from Tables 3.4.3.1 to 3.4.3.9. For instance, in Table 3.4.4.2 below, all results for the EVI group $\gamma = 0.1$ (Distributions 1 to 5 and the subtotal) for $n = 50$ (fifth column) were obtained from the mode estimator, fixing the threshold at $k = 40\%$ of n.

				n = 50			n = 100		
Distr.		γ	ρ	**Fixed**	**PPD**	**RMSE**	**Fixed**	**PPD**	**RMSE**
1	Burr	0.1	-2	0.0832	0.0677	0.0260	0.0660	0.0521	0.0228
				(0.0027)	(0.0023)		(0.0030)	(0.0024)	
2	Burr	0.1	-1	0.0524	0.0411	0.0203	0.0372	0.0274	0.0166
				(0.0020)	(0.0016)		(0.0022)	(0.0017)	
3	Burr	0.1	-0.5	0.1446	0.1812	0.0426	0.0916	0.1177	0.0343
				(0.0064)	(0.0064)		(0.0061)	(0.0063)	
4	Fréchet	0.1	-1	0.0659	0.0491	0.0222	0.0516	0.0350	0.0187
				(0.0022)	(0.0018)		(0.0026)	(0.0019)	
5	logΓ	0.1	0	0.0553	0.0406	0.0202	0.0435	0.0325	0.0180
				(0.0026)	(0.0019)		(0.0032)	(0.0024)	
γ = 0.1				**0.0803**	**0.0759**	**0.0276**	**0.0580**	**0.0529**	**0.0230**
				(0.0016)	(0.0015)		(0.0016)	(0.0015)	
6	Burr	0.5	-2	2.2085	1.8210	0.1349	1.6855	1.4283	0.1195
				(0.0641)	(0.0547)		(0.0749)	(0.0681)	
7	Burr	0.5	-1	1.2569	0.9261	0.0962	0.9130	0.7448	0.0863
				(0.0495)	(0.0390)		(0.0469)	(0.0411)	
8	Burr	0.5	-0.5	3.0273	3.8526	0.1963	2.2361	2.3898	0.1546
				(0.1409)	(0.1432)		(0.1416)	(0.1262)	
9	Fréchet	0.5	-1	1.8278	1.3757	0.1173	1.2656	0.9574	0.0978
				(0.0575)	(0.0469)		(0.0588)	(0.0492)	
10	*t*	0.5	-1	1.0614	0.8748	0.0935	0.6788	0.5890	0.0767
				(0.0479)	(0.0428)		(0.0471)	(0.0391)	
11	logΓ	0.5	0	1.2272	0.9493	0.0974	0.8112	0.6741	0.0821
				(0.0500)	(0.0437)		(0.0511)	(0.0458)	
γ = 0.5				**1.7682**	**1.6332**	**0.1278**	**1.2650**	**1.1306**	**0.1063**
				(0.0310)	(0.0293)		(0.0317)	(0.0280)	
12	Burr	1	-2	6.8492	8.1259	0.2851	5.5800	5.9039	0.2430
				(0.2321)	(0.2331)		(0.2617)	(0.2499)	
13	Burr	1	-1	4.4277	3.8879	0.1972	3.4861	2.9743	0.1725
				(0.1938)	(0.1571)		(0.1970)	(0.1608)	
14	Burr	1	-0.5	12.5338	12.3702	0.3517	6.8986	8.1449	0.2854
				(0.4969)	(0.5180)		(0.4291)	(0.4228)	
15	Fréchet	1	-1	5.8579	5.9964	0.2449	4.5796	4.3039	0.2075
				(0.2212)	(0.2113)		(0.2281)	(0.2047)	
16	*t*	1	-2	6.4216	7.2427	0.2691	5.5431	6.0318	0.2456
				(0.2170)	(0.2185)		(0.2540)	(0.2482)	
17	logΓ	1	0	4.5733	3.4014	0.1844	3.7183	2.4120	0.1553
				(0.2116)	(0.1525)		(0.2501)	(0.1645)	
γ = 1				**6.7773**	**6.8374**	**0.2615**	**4.9676**	**4.9618**	**0.2228**
				(0.1154)	(0.1134)		(0.1143)	(0.1051)	
Mean				**3.1024**	**3.0120**	**0.1736**	**2.2254**	**2.1658**	**0.1472**
				(0.0428)	(0.0413)		(0.0421)	(0.0384)	

Table 3.4.4.2 Fixed threshold and PPD estimator results for $n = 50$ and $n = 100$.

				n = 200			n = 500		
Distr.		γ	ρ	**Fixed**	**PPD**	**RMSE**	**Fixed**	**PPD**	**RMSE**
1	Burr	0.1	-2	0.0514	0.0405	0.0201	0.0329	0.0247	0.0157
				(0.0029)	(0.0022)		(0.0027)	(0.0019)	
2	Burr	0.1	-1	0.0311	0.0212	0.0146	0.0256	0.0130	0.0114
				(0.0017)	(0.0011)		(0.0021)	(0.0011)	
3	Burr	0.1	-0.5	0.0724	0.0833	0.0289	0.0307	0.0517	0.0227
				(0.0043)	(0.0041)		(0.0041)	(0.0038)	
4	Fréchet	0.1	-1	0.0388	0.0259	0.0161	0.0302	0.0168	0.0130
				(0.0020)	(0.0014)		(0.0029)	(0.0013)	
5	logΓ	0.1	0	0.0317	0.0286	0.0169	0.0285	0.0194	0.0139
				(0.0022)	(0.0022)		(0.0050)	(0.0023)	
γ **= 0.1**				**0.0451**	**0.0399**	**0.0200**	**0.0296**	**0.0251**	**0.0159**
				(0.0012)	(0.0011)		(0.0016)	(0.0010)	
6	Burr	0.5	-2	1.1216	0.9399	0.0970	0.7674	0.6359	0.0797
				(0.0514)	(0.0437)		(0.0656)	(0.0476)	
7	Burr	0.5	-1	0.7122	0.5805	0.0762	0.5768	0.3950	0.0628
				(0.0399)	(0.0329)		(0.0477)	(0.0392)	
8	Burr	0.5	-0.5	1.7391	1.6694	0.1292	0.9604	1.1387	0.1067
				(0.0943)	(0.0852)		(0.0715)	(0.0834)	
9	Fréchet	0.5	-1	0.9289	0.7253	0.0852	0.6675	0.4377	0.0662
				(0.0460)	(0.0380)		(0.0553)	(0.0333)	
10	*t*	0.5	-1	0.4378	0.4038	0.0635	0.3211	0.2722	0.0522
				(0.0320)	(0.0267)		(0.0262)	(0.0252)	
11	logΓ	0.5	0	0.7171	0.5613	0.0749	0.5285	0.3691	0.0608
				(0.0559)	(0.0482)		(0.0633)	(0.0364)	
γ **= 0.5**				**0.9428**	**0.8134**	**0.0902**	**0.6370**	**0.5414**	**0.0736**
				(0.0232)	(0.0202)		(0.0232)	(0.0196)	
12	Burr	1	-2	3.8053	4.3406	0.2083	2.2137	2.5581	0.1599
				(0.1846)	(0.1824)		(0.1761)	(0.1631)	
13	Burr	1	-1	2.5121	2.3573	0.1535	1.4392	1.5064	0.1227
				(0.1652)	(0.1260)		(0.1339)	(0.1408)	
14	Burr	1	-0.5	6.8924	5.9909	0.2448	4.7591	4.0734	0.2018
				(0.3808)	(0.3539)		(0.2936)	(0.2897)	
15	Fréchet	1	-1	3.1983	2.9867	0.1728	1.7037	1.6857	0.1298
				(0.2307)	(0.1754)		(0.1383)	(0.1251)	
16	*t*	1	-2	3.4984	4.1087	0.2027	2.1983	2.6006	0.1613
				(0.1761)	(0.1772)		(0.1807)	(0.1746)	
17	logΓ	1	0	3.1509	1.8962	0.1377	2.8202	1.5700	0.1253
				(0.2367)	(0.1386)		(0.3536)	(0.1965)	
γ **= 1**				**3.8429**	**3.6134**	**0.1901**	**2.5224**	**2.3324**	**0.1527**
				(0.0981)	(0.0843)		(0.0931)	(0.0773)	
Mean				**1.7270**	**1.5741**	**0.1255**	**1.1349**	**1.0217**	**0.1011**
				(0.0361)	(0.0306)		(0.0341)	(0.0281)	

Table 3.4.4.3 Fixed threshold and PPD estimator results for $n = 200$ and $n = 500$.

				***n* = 1000**			***n* = 2000**		
Distr.		γ	ρ	**Fixed**	**PPD**	**RMSE**	**Fixed**	**PPD**	**RMSE**
1	Burr	0.1	-2	0.0202	0.0167	0.0129	0.0136	0.0109	0.0105
				(0.0016)	(0.0011)		(0.0013)	(0.0008)	
2	Burr	0.1	-1	0.0158	0.0098	0.0099	0.0172	0.0085	0.0092
				(0.0013)	(0.0009)		(0.0013)	(0.0008)	
3	Burr	0.1	-0.5	0.0247	0.0372	0.0193	0.0146	0.0263	0.0162
				(0.0020)	(0.0035)		(0.0016)	(0.0025)	
4	Fréchet	0.1	-1	0.0189	0.0133	0.0115	0.0146	0.0078	0.0088
				(0.0018)	(0.0010)		(0.0012)	(0.0006)	
5	logΓ	0.1	0	0.0261	0.0167	0.0129	0.0234	0.0173	0.0132
				(0.0031)	(0.0017)		(0.0021)	(0.0014)	
γ = 0.1				**0.0212**	**0.0187**	**0.0137**	**0.0166**	**0.0141**	**0.0119**
				(0.0009)	(0.0009)		(0.0007)	(0.0006)	
6	Burr	0.5	-2	0.4716	0.3401	0.0583	0.2881	0.2247	0.0474
				(0.0422)	(0.0242)		(0.0249)	(0.0162)	
7	Burr	0.5	-1	0.2875	0.1955	0.0442	0.2773	0.1601	0.0400
				(0.0265)	(0.0194)		(0.0209)	(0.0152)	
8	Burr	0.5	-0.5	0.7987	0.9762	0.0988	0.5040	0.7235	0.0851
				(0.0592)	(0.0782)		(0.0429)	(0.0647)	
9	Fréchet	0.5	-1	0.3934	0.2352	0.0485	0.3098	0.1467	0.0383
				(0.0317)	(0.0200)		(0.0317)	(0.0116)	
10	*t*	0.5	-1	0.2043	0.2468	0.0497	0.2255	0.1578	0.0397
				(0.0184)	(0.0295)		(0.0191)	(0.0163)	
11	logΓ	0.5	0	0.6697	0.5327	0.0730	0.4538	0.3982	0.0631
				(0.0608)	(0.0401)		(0.0356)	(0.0234)	
γ = 0.5				**0.4709**	**0.4211**	**0.0649**	**0.3431**	**0.3018**	**0.0549**
				(0.0175)	(0.0166)		(0.0124)	(0.0125)	
12	Burr	1	-2	2.1398	1.5654	0.1251	1.7062	1.1343	0.1065
				(0.1828)	(0.1246)		(0.1827)	(0.0675)	
13	Burr	1	-1	1.2412	0.7953	0.0892	0.9888	0.5337	0.0731
				(0.0982)	(0.0747)		(0.0913)	(0.0511)	
14	Burr	1	-0.5	2.6523	3.7100	0.1926	1.6564	2.9970	0.1731
				(0.1837)	(0.2776)		(0.1628)	(0.2597)	
15	Fréchet	1	-1	1.4445	0.9030	0.0950	1.5941	0.7780	0.0882
				(0.1696)	(0.0820)		(0.1638)	(0.0624)	
16	*t*	1	-2	2.0298	1.5004	0.1225	1.4424	1.1532	0.1074
				(0.1872)	(0.1215)		(0.1219)	(0.0742)	
17	logΓ	1	0	2.5879	2.0135	0.1419	1.8078	1.5714	0.1254
				(0.2796)	(0.1671)		(0.1826)	(0.1208)	
γ = 1				**2.0159**	**1.7479**	**0.1322**	**1.5326**	**1.3613**	**0.1167**
				(0.0780)	(0.0640)		(0.0631)	(0.0523)	
Mean				**0.8847**	**0.7710**	**0.0878**	**0.6725**	**0.5911**	**0.0769**
				(0.0282)	(0.0234)		(0.0201)	(0.0190)	

Table 3.4.4.4 Fixed threshold and PPD estimator results for $n = 1000$ and $n = 2000$.

				***n* = 5000**			***n* = 10 000**		
Distr.		γ	ρ	**Fixed**	**PPD**	**RMSE**	**Fixed**	**PPD**	**RMSE**
1	Burr	0.1	-2	0.0115	0.0065	0.0081	0.0071	0.0042	0.0064
				(0.0017)	(0.0008)		(0.0010)	(0.0005)	
2	Burr	0.1	-1	0.0120	0.0059	0.0077	0.0076	0.0053	0.0073
				(0.0015)	(0.0007)		(0.0007)	(0.0006)	
3	Burr	0.1	-0.5	0.0078	0.0183	0.0135	0.0076	0.0111	0.0105
				(0.0015)	(0.0022)		(0.0010)	(0.0011)	
4	Fréchet	0.1	-1	0.0103	0.0056	0.0075	0.0073	0.0040	0.0063
				(0.0014)	(0.0006)		(0.0010)	(0.0005)	
5	logΓ	0.1	0	0.0175	0.0159	0.0126	0.0139	0.0129	0.0114
				(0.0025)	(0.0015)		(0.0018)	(0.0010)	
γ **= 0.1**				**0.0118**	**0.0104**	**0.0102**	**0.0087**	**0.0075**	**0.0087**
				(0.0008)	(0.0006)		(0.0005)	(0.0003)	
6	Burr	0.5	-2	0.1214	0.1095	0.0331	0.1415	0.0914	0.0302
				(0.0133)	(0.0114)		(0.0205)	(0.0098)	
7	Burr	0.5	-1	0.1750	0.1008	0.0317	0.1548	0.0793	0.0282
				(0.0185)	(0.0116)		(0.0184)	(0.0098)	
8	Burr	0.5	-0.5	0.3253	0.3770	0.0614	0.1924	0.2843	0.0533
				(0.0369)	(0.0433)		(0.0247)	(0.0343)	
9	Fréchet	0.5	-1	0.2022	0.1172	0.0342	0.1349	0.0843	0.0290
				(0.0228)	(0.0120)		(0.0160)	(0.0089)	
10	*t*	0.5	-1	0.1544	0.1356	0.0368	0.1635	0.1036	0.0322
				(0.0178)	(0.0197)		(0.0189)	(0.0128)	
11	logΓ	0.5	0	0.4174	0.4582	0.0677	0.3711	0.3863	0.0622
				(0.0428)	(0.0336)		(0.0518)	(0.0269)	
γ **= 0.5**				**0.2326**	**0.2164**	**0.0465**	**0.1930**	**0.1715**	**0.0414**
				(0.0112)	(0.0103)		(0.0114)	(0.0080)	
12	Burr	1	-2	1.1417	0.7888	0.0888	0.5650	0.3659	0.0605
				(0.1309)	(0.0761)		(0.0688)	(0.0376)	
13	Burr	1	-1	0.7985	0.4069	0.0638	0.5207	0.3506	0.0592
				(0.0854)	(0.0541)		(0.0654)	(0.0386)	
14	Burr	1	-0.5	1.0093	1.7775	0.1333	1.2134	1.1300	0.1063
				(0.1288)	(0.2028)		(0.1112)	(0.1286)	
15	Fréchet	1	-1	0.9403	0.5022	0.0709	0.5843	0.2354	0.0485
				(0.1029)	(0.0508)		(0.0719)	(0.0257)	
16	*t*	1	-2	0.9407	0.7132	0.0844	0.5196	0.4183	0.0647
				(0.1165)	(0.0724)		(0.0590)	(0.0435)	
17	logΓ	1	0	1.4298	1.3852	0.1177	1.0981	1.5004	0.1225
				(0.1868)	(0.1085)		(0.1220)	(0.1069)	
γ **= 1**				**1.0434**	**0.9289**	**0.0964**	**0.7502**	**0.6668**	**0.0817**
				(0.0527)	(0.0439)		(0.0353)	(0.0305)	
Mean				**0.4654**	**0.4073**	**0.0638**	**0.3423**	**0.2981**	**0.0546**
				(0.0192)	(0.0159)		(0.0134)	(0.0111)	

Table 3.4.4.5 Fixed threshold and PPD estimator results for $n = 5000$ and $n = 10\,000$.

				n = 20 000		
Distr.		γ	ρ	**Fixed**	**PPD**	**RMSE**
1	Burr	0.1	-2	0.0045	0.0027	0.0052
				(0.0006)	(0.0003)	
2	Burr	0.1	-1	0.0041	0.0030	0.0055
				(0.0005)	(0.0004)	
3	Burr	0.1	-0.5	0.0130	0.0098	0.0099
				(0.0009)	(0.0012)	
4	Fréchet	0.1	-1	0.0043	0.0024	0.0049
				(0.0006)	(0.0002)	
5	logΓ	0.1	0	0.0111	0.0132	0.0115
				(0.0011)	(0.0011)	
γ **= 0.1**				**0.0074**	**0.0062**	**0.0079**
				(0.0003)	(0.0004)	
6	Burr	0.5	-2	0.1137	0.1063	0.0326
				(0.0140)	(0.0157)	
7	Burr	0.5	-1	0.1455	0.1383	0.0372
				(0.0207)	(0.0311)	
8	Burr	0.5	-0.5	0.1268	0.1101	0.0332
				(0.0165)	(0.0129)	
9	Fréchet	0.5	-1	0.1472	0.1142	0.0338
				(0.0233)	(0.0128)	
10	*t*	0.5	-1	0.1203	0.0970	0.0311
				(0.0170)	(0.0119)	
11	logΓ	0.5	0	0.2618	0.2668	0.0516
				(0.0325)	(0.0301)	
γ **= 0.5**				**0.1526**	**0.1388**	**0.0373**
				(0.0088)	(0.0085)	
12	Burr	1	-2	0.5009	0.3776	0.0615
				(0.0777)	(0.0503)	
13	Burr	1	-1	0.4375	0.3830	0.0619
				(0.0533)	(0.0473)	
14	Burr	1	-0.5	0.4014	0.4330	0.0658
				(0.0533)	(0.0562)	
15	Fréchet	1	-1	0.5859	0.5100	0.0714
				(0.0779)	(0.0556)	
16	*t*	1	-2	0.4153	0.3610	0.0601
				(0.0550)	(0.0450)	
17	logΓ	1	0	1.1148	1.0225	0.1011
				(0.1210)	(0.0883)	
γ **= 1**				**0.5759**	**0.5145**	**0.0717**
				(0.0314)	(0.0241)	
Mean				**0.2670**	**0.2324**	**0.0482**
				(0.0123)	(0.0090)	

Table 3.4.4.6 Fixed threshold and PPD estimator results for $n = 20\,000$.

We conclude this section by considering some extra simulation results for very large samples.

For samples of size $n = 20\,000$, we apply Method 3 to the set of median estimates, with a fixed upper limit of 4. As pointed out, this suggests that the sample size becomes so large that any estimate with a threshold beyond $k = 20\%$ of n, need not be considered at all.

This idea was taken further by considering 30 values of k, namely $k = 1\%, 2\%, \dots, 30\%$ of n. In other words, we considered only large thresholds, with smaller increments between consecutive ones. We did this for samples of sizes $n = 20\,000$, $n = 50\,000$ and $n = 100\,000$.

For each sample size, a simulation study was done with 100 samples of that size from each of the 17 distributions used in the previous simulation studies. For each sample, the median and mode estimates were obtained at $k = 1\%, 2\%, \dots, 30\%$ of n. All three threshold selection methods (described in Section 3.4.2), were investigated.

The technique which consistently yielded good results over all sample sizes, distributions and EVI groups, was Method 1, with a window length of 29, applied to the median estimates.

This implies that for samples of sizes $n = 20\,000$, $n = 50\,000$ and $n = 100\,000$, the following method of estimating the EVI is recommended:

1. Given a sample of size n, calculate k multiplicative excesses at 30 values of k, namely $k = 1\%, 2\%, \dots, 30\%$ of n.
2. At each of these values, fit the PPD to the multiplicative excesses. Do this by determining the Bayesian estimates of γ, ρ and c, using Gibbs sampling.
3. This yields a set of 30 estimates of the EVI γ, by calculating the median of the Gibbs draws.
4. Apply Method 1 (described in Section 3.4.2) with a window length of 29. This entails calculating the instability measure of the median estimates at $k = 1\%, 2\%, \dots, 29\%$ of n, and at $k = 2\%, 3\%, \dots, 30\%$ of n, and choosing the set with the minimum instability as the stable region. Note that we still use $x = 1, 2, \dots, 29$, corresponding to $k = 1\%, 2\%, \dots, 29\%$ of n, when calculating the instability measure. That is, the increments in x stay 1, even though the increments in k were decreased from 5% to 1%.
5. The mean of the (median) estimates in the stable region is then the proposed benchmark estimate of the EVI.

Table 3.4.4.7 below contains the results obtained with the abovementioned method, as well as the information in Table 3.4.4.6 (for the sake of comparison). Table 3.4.4.8 below contains the results obtained for samples of sizes $n = 50\,000$ and $n = 100\,000$.

Only the final results of the simulation study are shown here. Also, we will not show results for samples of sizes $n = 50\,000$ and $n = 100\,000$ for any of the other estimation methods considered, including those in Chapter 4. Moreover, we will use the results of Table 3.4.4.6 for samples of size $n = 20\,000$ when comparing results with other methods of estimating the EVI.

				n = 20 000 (Old)			n = 20 000 (New)		
Distr.		γ	ρ	**Fixed**	**PPD**	**RMSE**	**Fixed**	**PPD**	**RMSE**
1	Burr	0.1	-2	0.0045	0.0027	0.0052	0.0045	0.0038	0.0062
				(0.0006)	(0.0003)		(0.0006)	(0.0005)	
2	Burr	0.1	-1	0.0041	0.0030	0.0055	0.0041	0.0062	0.0079
				(0.0005)	(0.0004)		(0.0005)	(0.0007)	
3	Burr	0.1	-0.5	0.0130	0.0098	0.0099	0.0130	0.0013	0.0036
				(0.0009)	(0.0012)		(0.0009)	(0.0002)	
4	Fréchet	0.1	-1	0.0043	0.0024	0.0049	0.0043	0.0049	0.0070
				(0.0006)	(0.0002)		(0.0006)	(0.0005)	
5	logΓ	0.1	0	0.0111	0.0132	0.0115	0.0111	0.0084	0.0091
				(0.0011)	(0.0011)		(0.0011)	(0.0011)	
γ = 0.1				**0.0074**	**0.0062**	**0.0079**	**0.0074**	**0.0049**	**0.0070**
				(0.0003)	(0.0004)		(0.0003)	(0.0003)	
6	Burr	0.5	-2	0.1137	0.1063	0.0326	0.1137	0.0974	0.0312
				(0.0140)	(0.0157)		(0.0140)	(0.0108)	
7	Burr	0.5	-1	0.1455	0.1383	0.0372	0.1455	0.1570	0.0396
				(0.0207)	(0.0311)		(0.0207)	(0.0128)	
8	Burr	0.5	-0.5	0.1268	0.1101	0.0332	0.1268	0.0352	0.0188
				(0.0165)	(0.0129)		(0.0165)	(0.0046)	
9	Fréchet	0.5	-1	0.1472	0.1142	0.0338	0.1472	0.1350	0.0367
				(0.0233)	(0.0128)		(0.0233)	(0.0139)	
10	*t*	0.5	-1	0.1203	0.0970	0.0311	0.1203	0.1713	0.0414
				(0.0170)	(0.0119)		(0.0170)	(0.0166)	
11	logΓ	0.5	0	0.2618	0.2668	0.0516	0.2618	0.2243	0.0474
				(0.0325)	(0.0301)		(0.0325)	(0.0306)	
γ = 0.5				**0.1526**	**0.1388**	**0.0373**	**0.1526**	**0.1367**	**0.0370**
				(0.0088)	(0.0085)		(0.0088)	(0.0069)	
12	Burr	1	-2	0.5009	0.3776	0.0615	0.5009	0.3972	0.0630
				(0.0777)	(0.0503)		(0.0777)	(0.0495)	
13	Burr	1	-1	0.4375	0.3830	0.0619	0.4375	0.5411	0.0736
				(0.0533)	(0.0473)		(0.0533)	(0.0527)	
14	Burr	1	-0.5	0.4014	0.4330	0.0658	0.4014	0.1282	0.0358
				(0.0533)	(0.0562)		(0.0533)	(0.0242)	
15	Fréchet	1	-1	0.5859	0.5100	0.0714	0.5859	0.6313	0.0795
				(0.0779)	(0.0556)		(0.0779)	(0.0679)	
16	*t*	1	-2	0.4153	0.3610	0.0601	0.4153	0.3681	0.0607
				(0.0550)	(0.0450)		(0.0550)	(0.0451)	
17	logΓ	1	0	1.1148	1.0225	0.1011	1.1148	0.7551	0.0869
				(0.1210)	(0.0883)		(0.1210)	(0.0775)	
γ = 1				**0.5759**	**0.5145**	**0.0717**	**0.5759**	**0.4702**	**0.0686**
				(0.0314)	(0.0241)		(0.0314)	(0.0227)	
Mean				**0.2670**	**0.2324**	**0.0482**	**0.2670**	**0.2156**	**0.0464**
				(0.0123)	(0.0090)		(0.0123)	(0.0084)	

Table 3.4.4.7 Fixed threshold and PPD estimator results for $n = 20\,000$.

				n = 50 000		n = 100 000	
Distr.		γ	ρ	**PPD**	**RMSE**	**PPD**	**RMSE**
1	Burr	0.1	-2	0.0023	0.0048	0.0024	0.0049
				(0.0003)		(0.0006)	
2	Burr	0.1	-1	0.0046	0.0068	0.0040	0.0063
				(0.0005)		(0.0003)	
3	Burr	0.1	-0.5	0.0008	0.0028	0.0004	0.0019
				(0.0001)		(0.0001)	
4	Fréchet	0.1	-1	0.0036	0.0060	0.0034	0.0059
				(0.0004)		(0.0003)	
5	logΓ	0.1	0	0.0042	0.0064	0.0047	0.0068
				(0.0004)		(0.0005)	
γ = 0.1				**0.0031**	**0.0056**	**0.0030**	**0.0055**
				(0.0002)		(0.0002)	
6	Burr	0.5	-2	0.0826	0.0287	0.0578	0.0240
				(0.0102)		(0.0072)	
7	Burr	0.5	-1	0.1189	0.0345	0.1329	0.0364
				(0.0112)		(0.0109)	
8	Burr	0.5	-0.5	0.0114	0.0107	0.0049	0.0070
				(0.0023)		(0.0007)	
9	Fréchet	0.5	-1	0.0984	0.0314	0.1099	0.0331
				(0.0109)		(0.0094)	
10	*t*	0.5	-1	0.1446	0.0380	0.1721	0.0415
				(0.0119)		(0.0130)	
11	logΓ	0.5	0	0.1115	0.0334	0.0945	0.0307
				(0.0113)		(0.0114)	
γ = 0.5				**0.0946**	**0.0308**	**0.0953**	**0.0309**
				(0.0042)		(0.0039)	
12	Burr	1	-2	0.2061	0.0454	0.2233	0.0473
				(0.0283)		(0.0293)	
13	Burr	1	-1	0.5560	0.0746	0.4480	0.0669
				(0.0576)		(0.0331)	
14	Burr	1	-0.5	0.0497	0.0223	0.0350	0.0187
				(0.0075)		(0.0052)	
15	Fréchet	1	-1	0.3474	0.0589	0.3512	0.0593
				(0.0381)		(0.0403)	
16	*t*	1	-2	0.2963	0.0544	0.2063	0.0454
				(0.0355)		(0.0327)	
17	logΓ	1	0	0.5329	0.0730	0.3164	0.0563
				(0.0556)		(0.0355)	
γ = 1				**0.3314**	**0.0576**	**0.2634**	**0.0513**
				(0.0166)		(0.0128)	
Mean				**0.1513**	**0.0389**	**0.1275**	**0.0357**
				(0.0061)		(0.0047)	

Table 3.4.4.8 PPD estimator results for $n = 50\,000$ and $n = 100\,000$.

3.4.5 Implied threshold and quantile estimation

The estimator of Section 3.4.4 is calculated as the average of EVI estimates over a stable region. While this is sufficient for the purpose of obtaining a point estimate of the EVI, we need a specific set of multiplicative excesses on which to base the calculation of extreme quantiles and hpd regions.

Figure 3.4.2.1 showed the results we obtained when applying Method 1 to mode estimates with a fixed region length of 12 for the Burr sample. Consider a graph similar to Figure 3.4.2.1 to illustrate the idea of an *implied threshold*.

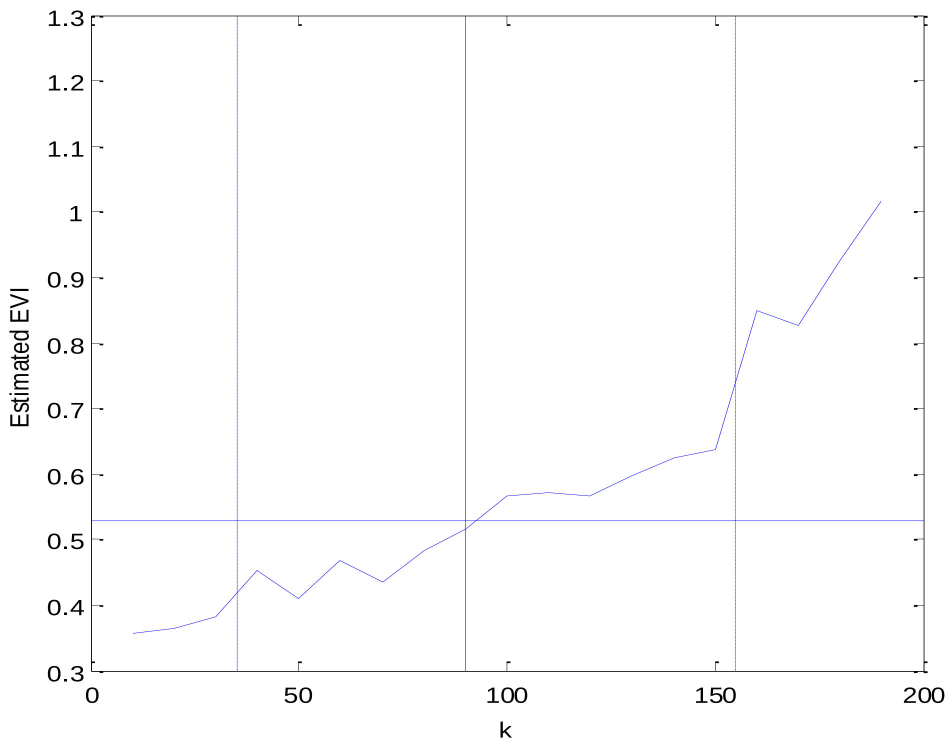

Figure 3.4.5.1 Implied threshold for the Burr sample.

The solid horizontal line represents our estimate of the EVI for this sample (0.5274). The two dashed vertical lines on the left and right indicate our stable region.

The solid vertical line indicates our implied threshold, which corresponds to $k = 90$. This value of $k = 90$ was obtained by finding the value of k in the stable region (from $k = 40$ to $k = 150$) for which the EVI estimate was closest to 0.5274. (The EVI estimate at $k = 90$ was 0.5153.) We therefore regard $k = 90$ as our threshold for the purpose of doing further calculations.

If there are two or more values of k which correspond to EVI estimates equally close to our final EVI estimate, the maximum of these values of k is chosen as our threshold. This happens for instance when the stable region consists of only two values of k. The reason we use the maximum k value is that the more observations (excesses) we include, the smaller the variance of the estimator, and the narrower the resulting hpd regions.

We now calculate 90%, 95% and 99% hpd regions for the EVI for the Burr sample, the estimates of the quantiles with upper probabilities $1/2n$ = $1/400$ and $1/10n$ = $1/2000$ and their respective 90%, 95% and 99% hpd regions. For these calculations we use the techniques described in Section 3.2. Note that we use the mode of the Gibbs draws of the quantiles, since we use the mode for the estimation of the EVI. The results are summarized in the following table:

			90% hpd		**95% hpd**		**99% hpd**	
	True value	**Estimate**	**Lower**	**Upper**	**Lower**	**Upper**	**Lower**	**Upper**
EVI	0.5000	0.5274	0.4105	0.6734	0.3969	0.7163	0.3693	0.8243
$\boldsymbol{x_{1/2n}}$	19.784	19.222	11.730	39.775	10.595	47.106	9.170	70.617
$\boldsymbol{x_{1/10n}}$	44.565	44.101	23.328	117.581	20.552	148.707	17.126	267.289

Table 3.4.5.1 Further calculations done at $k = 90$ for the Burr sample.

Here x_p denotes the quantile of the true underlying distribution, which has upper probability p. In the case of Table 3.4.5.1, the underlying distribution is the *Burr*$(\beta = 1, \gamma = 0.5, \rho = -0.7)$ distribution and the sample size is $n = 200$.

The next table shows the results of the simulation study for $n = 50$, followed by a description of the contents.

		Coverage probability			**Median interval length**		
Hpd region		**90%**	**95%**	**99%**	**90%**	**95%**	**99%**
$\hat{\gamma}$	γ **= 0.1**	97.6%	99.5%	100.0%	0.1137	0.1535	0.2618
	γ **= 0.5**	96.0%	98.8%	100.0%	0.5032	0.6537	1.0451
	γ **= 1**	90.5%	96.1%	99.7%	0.8220	1.0409	1.5950
$\boldsymbol{x_{1/2n}}$	γ **= 0.1**	94.0%	97.0%	99.7%	0.5426	0.6835	1.0351
	γ **= 0.5**	93.2%	96.9%	99.5%	20.9222	29.9892	62.8772
	γ **= 1**	87.6%	94.2%	99.0%	323.8504	530.0165	1576.8630
$\boldsymbol{x_{1/10n}}$	γ **= 0.1**	94.3%	98.0%	99.7%	0.9788	1.2730	2.0630
	γ **= 0.5**	95.0%	97.5%	99.6%	87.8707	142.1258	429.0643
	γ **= 1**	89.3%	95.4%	99.2%	3188.7079	6544.3668	35177.5012

Table 3.4.5.2 Coverage probability and length of hpd regions for samples of size $n = 50$.

Table 3.4.5.2 shows the results for the estimated hpd regions of three parameters, namely the EVI, $x_{1/2n}$ and $x_{1/10n}$. The results are subdivided into EVI groups.

For coverage probability the percentages given are the medians of the coverage probabilies obtained for the distributions in question. For instance, the first value of 97.6% was obtained in the following way:

1. For each of the first five distributions (for which $\gamma = 0.1$), we had 1000 samples (number of repetitions for $n = 50$). For each of these samples we calculated a 90% hpd region by means of Gibbs sampling.
2. For each of the first five distributions we calculated the proportion of times the value 0.1 was in the interval, yielding five coverage probabilities.
3. The median of these five probabilities was 97.6%.

The reason we calculated the median of the coverage probabilities, is to avoid the results being too adversely affected by the poor performance of a single family of distributions, in this case the Burr distributions with $\rho = -0.5$.

For median interval lengths the values given are the median of the median interval lengths for the distributions in question. For instance, the first value of 0.1137 was obtained in the following way:

1. For each of the 17 distributions, we had 1000 samples. For each of these samples we calculated a 90% hpd region by means of Gibbs sampling.
2. We calculated the median length for each of the 17 sets of 1000 hpd regions.
3. The median of the first five medians (distibutions for which $\gamma = 0.1$) was 0.1137.

The results pertaining to other sample sizes were obtained in a similar fashion. Since these results will not be used in subsequent chapters, they will not be discussed, but merely presented for the sake of completeness. Also, we will not cover quantile estimation or interval estimation for any of the estimators presented in Chapter 4.

Note that the results we give in this section were obtained by using 1000 Gibbs repetitions. Increasing the number of repetitions to a more significant number, say $100\,000$, improves the results quite substantially. (The results presented in Table 3.4.5.1 were obtained by using $100\,000$ repetitions.)

For samples of size $n = 100$ we obtained:

		Coverage probability			Median interval length		
Hpd region		**90%**	**95%**	**99%**	**90%**	**95%**	**99%**
$\hat{\gamma}$	γ = **0.1**	96.4%	97.6%	99.6%	0.0845	0.1126	0.2065
	γ = **0.5**	94.1%	97.8%	99.3%	0.3838	0.4850	0.7922
	γ = **1**	88.2%	95.1%	99.3%	0.6774	0.8511	1.3231
$x_{1/2n}$	γ = **0.1**	92.0%	96.2%	99.0%	0.4979	0.6163	0.9048
	γ = **0.5**	94.3%	96.8%	99.2%	23.5852	32.4656	60.0716
	γ = **1**	89.8%	95.7%	99.1%	571.8821	896.9684	2292.5194
$x_{1/10n}$	γ = **0.1**	93.6%	96.6%	99.4%	0.8514	1.0755	1.6724
	γ = **0.5**	94.9%	97.3%	99.5%	86.7417	130.0791	310.4036
	γ = **1**	91.3%	96.6%	99.3%	5007.2185	9544.1204	38482.7989

Table 3.4.5.3 Coverage probability and length of hpd regions for samples of size $n = 100$.

For samples of size $n = 200$ we obtained:

		Coverage probability			Median interval length		
Hpd region		**90%**	**95%**	**99%**	**90%**	**95%**	**99%**
$\hat{\gamma}$	γ = **0.1**	93.4%	96.2%	98.6%	0.0646	0.0817	0.1355
	γ = **0.5**	93.2%	96.5%	99.0%	0.3067	0.3836	0.6207
	γ = **1**	88.1%	94.0%	98.9%	0.5485	0.6940	1.0495
$x_{1/2n}$	γ = **0.1**	89.8%	94.4%	98.8%	0.4540	0.5583	0.7826
	γ = **0.5**	92.0%	95.8%	98.9%	28.2820	37.9132	65.2613
	γ = **1**	90.4%	95.9%	98.6%	1009.3777	1482.0383	3076.4167
$x_{1/10n}$	γ = **0.1**	90.6%	94.4%	98.6%	0.7482	0.9420	1.3598
	γ = **0.5**	92.2%	96.1%	99.0%	98.8983	140.8901	286.4945
	γ = **1**	91.2%	94.6%	98.8%	7871.0899	12997.3341	39619.3158

Table 3.4.5.4 Coverage probability and length of hpd regions for samples of size $n = 200$.

For samples of size $n = 500$ we obtained:

		Coverage probability			Median interval length		
Hpd region		**90%**	**95%**	**99%**	**90%**	**95%**	**99%**
$\hat{\gamma}$	γ = **0.1**	90.5%	92.0%	96.5%	0.0460	0.0559	0.0781
	γ = **0.5**	89.8%	94.3%	98.5%	0.2155	0.2615	0.3669
	γ = **1**	87.5%	93.0%	98.5%	0.4375	0.5381	0.7563
$x_{1/2n}$	γ = **0.1**	88.5%	92.5%	97.5%	0.3977	0.4731	0.6515
	γ = **0.5**	92.3%	95.0%	98.5%	33.6336	42.3798	63.7435
	γ = **1**	88.8%	96.3%	99.3%	2077.0734	2846.7786	5241.5063
$x_{1/10n}$	γ = **0.1**	88.5%	92.5%	97.5%	0.6310	0.7461	1.0517
	γ = **0.5**	91.5%	95.0%	99.0%	104.7711	136.5060	217.4860
	γ = **1**	89.3%	95.0%	99.0%	14814.3619	22139.6735	49540.9392

Table 3.4.5.5 Coverage probability and length of hpd regions for samples of size $n = 500$.

For samples of size $n = 1000$ we obtained:

		Coverage probability			Median interval length		
Hpd region		**90%**	**95%**	**99%**	**90%**	**95%**	**99%**
$\hat{\gamma}$	γ = **0.1**	84.5%	93.0%	97.5%	0.0356	0.0425	0.0568
	γ = **0.5**	89.0%	93.8%	98.5%	0.1834	0.2140	0.2857
	γ = **1**	84.5%	91.8%	97.5%	0.3381	0.3996	0.5152
$x_{1/2n}$	γ = **0.1**	88.5%	94.0%	97.0%	0.3449	0.4154	0.5619
	γ = **0.5**	88.5%	93.0%	96.8%	40.8569	50.7406	75.1961
	γ = **1**	87.3%	93.3%	97.5%	3254.3702	4243.5241	7366.8107
$x_{1/10n}$	γ = **0.1**	87.5%	94.5%	97.0%	0.5288	0.6383	0.8680
	γ = **0.5**	88.5%	94.3%	96.8%	123.5726	158.0292	245.1463
	γ = **1**	86.5%	93.3%	98.3%	21691.3309	30292.6145	60011.0165

Table 3.4.5.6 Coverage probability and length of hpd regions for samples of size $n = 1000$.

For samples of size $n = 2000$ we obtained:

		Coverage probability			Median interval length		
Hpd region		**90%**	**95%**	**99%**	**90%**	**95%**	**99%**
$\hat{\gamma}$	γ = **0.1**	86.0%	91.0%	97.5%	0.0292	0.0346	0.0444
	γ = **0.5**	86.8%	94.8%	98.5%	0.1529	0.1772	0.2197
	γ = **1**	79.0%	89.0%	97.0%	0.2771	0.3198	0.4031
$x_{1/2n}$	γ = **0.1**	87.5%	93.5%	98.0%	0.2991	0.3564	0.4683
	γ = **0.5**	89.0%	95.0%	97.5%	46.0590	56.9255	80.7567
	γ = **1**	86.0%	92.5%	98.0%	5446.2898	7027.9708	11069.9355
$x_{1/10n}$	γ = **0.1**	88.5%	93.0%	97.5%	0.4514	0.5352	0.7111
	γ = **0.5**	89.3%	94.8%	97.5%	134.9629	168.5044	251.5569
	γ = **1**	84.5%	92.5%	97.8%	35188.5014	46838.0423	81776.3331

Table 3.4.5.7 Coverage probability and length of hpd regions for samples of size $n = 2000$.

For samples of size $n = 5000$ we obtained:

		Coverage probability			Median interval length		
Hpd region		**90%**	**95%**	**99%**	**90%**	**95%**	**99%**
$\hat{\gamma}$	γ = **0.1**	77.0%	86.0%	94.0%	0.0213	0.0250	0.0310
	γ = **0.5**	84.5%	94.0%	98.5%	0.1171	0.1366	0.1658
	γ = **1**	75.0%	84.0%	95.5%	0.2317	0.2625	0.3221
$x_{1/2n}$	γ = **0.1**	85.0%	90.0%	96.0%	0.2680	0.3192	0.4251
	γ = **0.5**	86.5%	92.0%	95.0%	59.0858	72.4237	98.1175
	γ = **1**	83.0%	92.0%	96.5%	10857.2277	13829.2951	20503.7649
$x_{1/10n}$	γ = **0.1**	85.0%	91.0%	95.0%	0.4004	0.4772	0.6334
	γ = **0.5**	85.5%	92.5%	95.5%	170.1191	208.3107	291.8951
	γ = **1**	82.0%	92.0%	97.5%	67770.1084	89683.8169	140371.4286

Table 3.4.5.8 Coverage probability and length of hpd regions for samples of size $n = 5000$.

For samples of size $n = 10\,000$ we obtained:

Hpd region		Coverage probability 90%	95%	99%	Median interval length 90%	95%	99%
$\hat{\gamma}$	γ = **0.1**	78.0%	85.0%	96.0%	0.0180	0.0214	0.0265
	γ = **0.5**	85.5%	89.0%	98.5%	0.0926	0.1083	0.1351
	γ = **1**	80.0%	88.5%	97.0%	0.1833	0.2077	0.2574
$\boldsymbol{x_{1/2n}}$	γ = **0.1**	81.0%	87.0%	95.0%	0.2543	0.3109	0.4118
	γ = **0.5**	84.5%	89.5%	96.0%	67.7321	83.2114	114.0583
	γ = **1**	84.5%	90.5%	96.0%	18800.4229	23475.1619	34167.7872
$\boldsymbol{x_{1/10n}}$	γ = **0.1**	82.0%	87.0%	94.0%	0.3763	0.4568	0.6063
	γ = **0.5**	84.5%	90.0%	96.0%	187.1997	233.0747	328.3992
	γ = **1**	84.5%	91.0%	96.5%	115520.6841	148216.8292	224676.1505

Table 3.4.5.9 Coverage probability and length of hpd regions for samples of size $n = 10\,000$.

For samples of size $n = 20\,000$ we obtained:

Hpd region		Coverage probability 90%	95%	99%	Median interval length 90%	95%	99%
$\hat{\gamma}$	γ = **0.1**	66.0%	77.0%	95.0%	0.0135	0.0164	0.0222
	γ = **0.5**	75.5%	86.5%	94.5%	0.0848	0.1009	0.1284
	γ = **1**	74.0%	85.0%	92.5%	0.1618	0.1938	0.2460
$\boldsymbol{x_{1/2n}}$	γ = **0.1**	72.0%	81.0%	91.0%	0.2202	0.2729	0.3864
	γ = **0.5**	79.5%	87.5%	92.5%	83.8413	103.6563	142.1876
	γ = **1**	80.5%	86.5%	93.0%	30804.8561	39628.4977	56067.1341
$\boldsymbol{x_{1/10n}}$	γ = **0.1**	73.0%	79.0%	93.0%	0.3232	0.4027	0.5627
	γ = **0.5**	80.5%	88.0%	92.5%	232.3722	286.9152	399.1022
	γ = **1**	78.5%	85.5%	92.5%	185876.0885	240353.6663	351559.8011

Table 3.4.5.10 Coverage probability and length of hpd regions for samples of size $n = 20\,000$.

This concludes the section of this chapter devoted to the methodology and design of the simulation study, as well as the results pertaining to the PPD estimator.

The final section of this chapter will be devoted to procedures and approaches followed to ensure that such an extensive simulation study produces reliable results. The same principles were applied to the simulation studies conducted in Chapter 4.

3.5 Programming method

The extent of the simulation studies performed necessitated a disciplined and sometimes innovative approach to programming. In this section we will make some comments on the programming techniques used to ensure sound simulation results.

We will use the term *program* to refer to any coded algorithm. Sometimes we will make a distinction between functions and procedures. A procedure is simply a list of instructions which are performed. A function requires input parameters, and returns output parameters to the program which uses the function.

Programs were written in MATLAB.

After coding, each program is subjected to the following checklist, the details of which will be discussed below:

3.5.1 General principles

3.5.2 Commenting

3.5.3 Initializing the pseudo random number generator

3.5.4 Clearing variables

3.5.5 Checking input parameters

3.5.6 Checking output parameters

3.5.7 Log of negative numbers and division by zero

3.5.8 Checking functioning of code

3.5.9 Testing the general functioning of the program

3.5.10 Testing every line of code

3.5.11 Testing output data

3.5.12 Directory structure

In the final section (Section 3.5.13) we will deal with techniques applied to reduce computation time in large simulation studies.

3.5.1 General principles

After writing the programming code, it is always checked whether the program is specific to solve the problem at hand. In other words, we always avoided overcomplicating the code by trying to make it unnecessarily generalized. This approach makes programs simpler to write, check and test, and hence more reliable.

The prime example is the Gibbs sampler. A separate Gibbs sampler was written for the PPD and the GPD, even though the ideas are the same. Also, for each sampler, each conditional posterior log likelihood is calculated by a different program.

Only if an obvious generalization is implied by the individual programs, is it made. This was done, for example, in the function *Quantile.m*. We put all quantile functions in one program, which has as one of its parameters the name of the family of distributions under consideration.

The second general principle is to check whether any assumptions were made. If this is the case, the assumptions are stated in the comments of the program itself (see Section 3.5.2 below for more on commenting) and it is also explicitly stated in the thesis. We classified all assumptions as either a computational assumption or a theoretical/statistical assumption. Doing calculations accurately to only two decimal places to speed up computation time, is an example of a computational assumption. Restricting a parameter to a certain range, is an example of a theoretical/statistical assumption.

3.5.2 Commenting

Each function is supplied with a comment in the heading, which gives the name of the function, a brief description of what it does, syntax, description of input and output parameters, an example (if necessary) and assumptions made (if any).

The format is more or less similar to that of the built-in functions of MATLAB, and is accessible by using the help command. For instance, typing “help PPDBayes”:

```
%PPDBAYES   Calculates draws of marginal posteriors of Perturbed Pareto
%   Distribution parameters, using the Gibbs sampler.
%   DRAWS = PPDBAYES(DATA,K,[],REPS,BURNIN,P) returns in DRAWS the marginal
%      posterior draws of GAMMA, RHO and C, and draws of quantiles of the
%      original distribution.
%   DRAWS = PPDBAYES(DATA,K,RHO,REPS,BURNIN,P) returns the marginal
%      posterior draws of GAMMA and C, and draws of quantiles of the
%      original distribution.
%      DATA: Vector of positive observations, from the original
%         distribution G.
%      K: Number of excesses.
%      REPS: Number of MCMC draws (all which are retained).
%      BURNIN: Number of burn-in draws (not retained in output).
%      P: A vector of exceedance probabilities of original distribution.
%         Quantiles of G associated with these probabilities are drawn.
%         Empty if no quantiles is needed.
```

```
%   The PPD survival function is given by
%      1 - F(x) = (1 - C)x^(-1/GAMMA) + Cx^(-(1 - RHO)/GAMMA), where
%      x >= 1, GAMMA > 0, 1/RHO <= C <= 1 and RHO < 0.
%   The first order MDI prior (prior of the Pareto distribution) is used:
%      EXP(-GAMMA)/GAMMA
%   The following is returned, if RHO was specified empty:
%      DRAWS = [g1 r1 c1 x11 x21 ...
%               g2 r2 c2 x12 x22 ...
%               :  :  :   :   :   :], where g1, g2, ... are draws from
%         the marginal posterior of GAMMA. Similarly for RHO and C.
%         If P = [p1 p2], then x11, x12, ... are draws of quantiles,
%         such that G(X > x1i) = p1 for the original distribution G.
%   GAMMA initiated at Hill estimate and RHO at -0.5.
%   The value of RHO is restricted to -2 <= RHO <= -0.01.
%   If RHO is specified (not empty), the column [r1 r2 ...]' is not given
%      as output.
%   if P is empty, the columns of xij's are not given as output.
%   In order to calculate Bayesian estimates, issue the command
%      mean(DRAWS) assuming squared error loss (standard),
%      median(DRAWS) assuming absolute loss, and
%      mode(DRAWS) assuming 0-1 loss.
```

Also, commenting is done inside each program if any step is not straightforward.

3.5.3 Initializing the pseudo random number generator

The MATLAB random number generator always uses the same initial seed each time MATLAB is opened. The command *Randomize* is put as first command in any program which uses the random number generator, and executes the following program:

```
%RANDOMIZE   Randomizes pseudo random number generator.

global DIDRANDOMIZE

if isempty(DIDRANDOMIZE)
   rand('state',sum(100*clock));
   rand(50,1);
   DIDRANDOMIZE = 1;
end
```

This program specifies that the MATLAB 5 generator is used. It also ensures that initialization is done only once, by making use of a global variable.

3.5.4 Clearing variables

The command *clear* is used at the beginning of procedures. This makes procedures more reliable, in the sense that it ensures that values are assigned to variables used later on, and it avoids erroneously using previously assigned values of a variable.

3.5.5 Checking input parameters

Each input variable is checked with respect to the following:

- Integer: Some variables, for example sample size, should be an integer.
- Range: Some variables, for example the EVI should be positive. Also, relationships between variables are tested here. For instance, k (number of excesses) and n (sample size) should both be postive integers, but we also require $k < n$.
- Dimensions: Some variables should be scalars (variables of dimension 1×1), some vectors (dimension $1 \times m$ or $m \times 1$) and some matrices (dimension $m \times n$).

The function *Numerical.m* has also been written to check any parameter of any dimension, whether it is real, finite and not NaN (Not a Number). This function is mainly used to check output data (see Section 3.5.11), to pick up situations where the calculated values might have been the square root of a negative number (not real), too large (for example an infinite likelihood function value) or Not a Number (for example when trying to calculate the ratio of two infinite numbers).

3.5.6 Checking output parameters

This is done only to an extent similarly to the way input parameters are checked, but not to the same degree, since it will take up too much computation time, and also because the output data are tested thoroughly in any case (see Section 3.5.11).

Checks were always performed, which returned an error message when a likelihood function yielded an infinite value, or when an optimization algorithm (finding a root, maximum or minimum) did in fact not converge.

3.5.7 Log of negative numbers and division by zero

Every instance of log in the code is checked whether the argument of the log function can be negative, and every instance of division in the code is checked whether the denominator might be zero.

3.5.8 Checking functioning of code

Every line of code is then checked in detail by hand, writing down the names of variables that have been initialized, checking calculations, and checking the scope and updating of variables. We used a set of conventions for variable names, to ensure that the names are meaningful and unambiguous.

3.5.9 Testing the general functioning of the program

The program is then tested. First, all syntax errors are corrected. A check is then performed on the general functionality of the program. This is done by using one set of sample input and checking the output for correctness.

3.5.10 Testing every line of code

The program is then tested in detail, ensuring every line of code is executed at least once, and all exceptions and special cases are accounted for.

3.5.11 Testing output data

This step is quite extensive and time consuming, but essential. Many disasters have been avoided, and many insights gained, by performing these checks in a disciplined fashion. Each step in checking the data will now be described. We will use as an example the output obtained from the Gibbs sampler, specifically for the case where we estimated all three parameters of the PPD, namely γ, ρ and c.

The first major step is to check the integrity of the output. This involves running the same checks on the output data as the checks on the input and output parameters of functions, as mentioned in Sections 3.5.5 and 3.5.6. Here, however, the function *Numerical.m* is always used.

For the Gibbs sampler, the following checks were performed on the output:

- Dimensions of the output matrix.
- Running *Numerical.m*: All values should be finite, real and not NaN (Not a Number).
- Range of the estimates in isolation: $\hat{\gamma} > 0$, $-2 < \hat{\rho} < -0.01$, $\hat{c} \leq 1$ and quantile estimates positive.
- Range relationship between estimates: quantiles with smaller tail probabilities should be larger than quantiles with larger upper tail probabilities; the lower limits of the hpd regions should be below the upper limits; point estimates should fall inside the hpd regions; and hpd regions should get wider when moving from 90% to 99%.

We have here only stated the checks performed on the output. In Section 3.2.4 we discussed some cases of violations, by describing what steps were taken to remedy them. It was also shown that, by investigating these violations and their causes, one may obtain valuable insights.

The second major step is to check how meaningful the output is and to get an overall idea of how well the technique, for instance the estimation technique, performs.

The most important method here is to draw a histogram of output, for instance a histogram of the set of EVI estimates obtained from all the repetitions.

For the Gibbs sampler, we had to go even further. One needs to check individual repetitions, to see whether the relationship between estimates and hpd regions makes sense, and how these estimates compare with the true values of the parameters.

The vast amount of output necessitated a randomization approach to inspecting the output. This was done for the estimates of each of the three parameters of the PPD, as well as the two estimated quantiles. As an example, we consider what was done for the estimates of the EVI. The following procedure was executed, first a number of times for samples of size $n = 50$, then a number of times for samples of size $n = 100$, up to samples of size n = 20 000:

1. Select a distribution number from 1 to 17, randomly.
2. Select a repetition number from 1 to N, randomly, where N is based on Table 3.3.1.2. For example, $N = 500$ for samples of size $n = 100$.
3. Select a threshold number from 1 to 19, randomly. This corresponds, respectively, to $k = 5\%, 10\%, \ldots,\ 95\%$ of n.
4. Check whether the calculated values make sense by constructing the following graph:

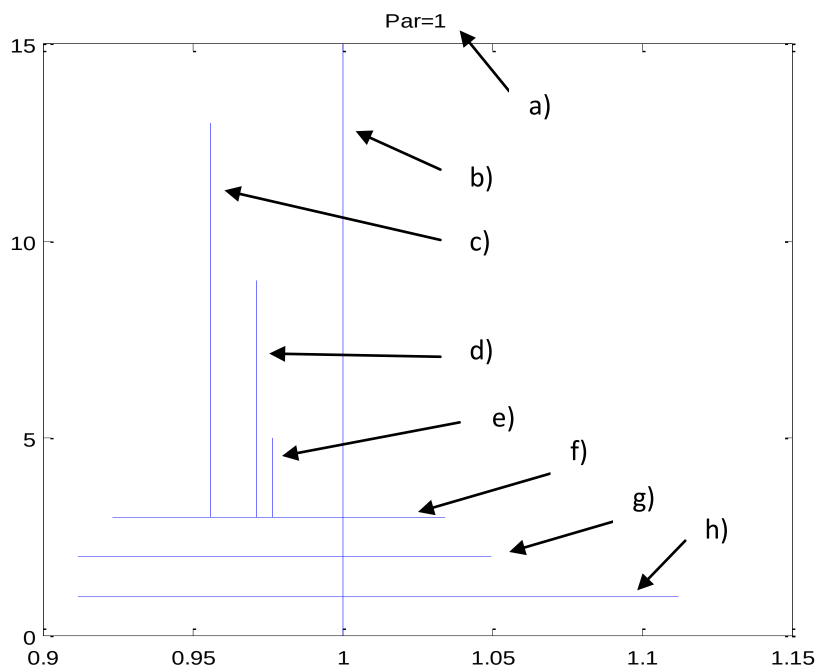

Figure 3.5.11 Checking the output of Gibbs sampler.

The following are shown in Figure 3.5.11:

a) Parameter 1 is the EVI.
b) True value of the EVI for this randomly selected distribution. In this case $\gamma = 1$.
c) Mode estimate of the EVI. (Mode of the Gibbs draws of the EVI.)

d) Median estimate of the EVI.
e) Mean estimate of the EVI.
f) 90% hpd region.
g) 95% hpd region.
h) 99% hpd region.

The third major step is to check the results overall. This entails the following:

1. Compare the simulation results to the results obtained in the literature, if available.
2. Checking whether the number of repetitions is sufficient to get a clear pattern across thresholds. This means that the average MSE should follow a smooth pattern, if plotted against the threshold fraction $k = 5\%, 10\%, \dots, 95\%$ of n, with a clear minimum value at some point between the two extreme thresholds $k = 5\%$ of n and $k = 95\%$ of n.
3. Checking whether the average MSE indeed becomes smaller as the sample size n gets larger.

3.5.12 Directory structure

Keeping track of all programs and data files becomes an important aspect. Different directories were created which contain only files with specific properties:

- A directory containing only standalone MATLAB programs, which have been fully checked and tested, using the methods described above.
- A directory containing MATLAB programs, used only once in the course of research, fully checked and tested, but not commented in as much detail.
- A directory containing data files used frequently throughout. The most important of these are the simulated samples from the 17 distributions on which all estimation techniques are tested, and the portfolio data, as described in Chapter 5.
- A directory containing data (output) files used only for a single simulation result. These files are logged on a sheet of paper, describing the contents and which testing procedures have already been done, if the checking has not been completed.
- A working directory containing all programs currently under construction or currently being used, and all data output files not yet processed into research results.

3.5.13 Reducing simulation time by using a network of computers

The Gibbs sampler is computationally intensive, and since we needed to calculate hundreds of thousands of sets of parameter estimates, a technique had to be developed for making use of a network of computers. In fact, at one stage 90 computers were run simultaneously.

The technique which we employed will now be described in detail.

One computer, which we will call the *hub*, sends commands, and receives and consolidates results from the other computers which perform the actual Gibbs sampling.

Each computer is assigned a number, ranging from 10 to 99. A file is created containing details of all the simulations which need to be performed. This file, called the *batch*, contains a matrix, which looks something like this:

1	1	1
1	1	2
:	:	:
1	1	100
1	2	1
1	2	2
:	:	:
1	17	100
2	1	1
2	1	2
:	:	:
9	17	100

Table 3.5.13 Contents of the batch file.

The first column gives the sample size code (ssc), the second the distribution number (d) and the third the sample/repetition number (rep). Each row is an instruction. The final row contains $ssc = 9$, $d = 17$ and $rep = 100$, which gives the following instruction: Determine the Gibbs estimates of the PPD parameters (at all 19 thresholds) for samples of size $n = 20\,000$ ($ssc = 1$ corresponding to $n = 50$, up to $ssc = 9$ corresponding to $n = 20\ 000$), using as data sample number 100 (the last sample in the case of samples of size $n = 20\ 000$) simulated from Distribution 17.

A network directory is used to connect the hub to the computers.

Computer 25, for example, will do the following until the entire simulation study is completed:

1. Write a request to obtain an instruction. This is done by writing an empty file with name r25 (“r” for “request”, followed by the computer number) to the network directory.
2. Read every 5 seconds the network directory for a file called g25 (“g” for “given instruction”, followed by the computer number).
3. Once the file g25 is available, Computer 25 reads the contents, which is one row of the batch, and deletes g25 from the network directory.
4. Suppose g25 contained $ssc = 3$, $d = 4$ and $rep = 7$. Computer 25 then determines estimates of the parameters of the PPD at 19 thresholds, using as data the seventh sample of size $n = 200$ ($ssc = 3$) simulated from Distribution 4.
5. The results are saved to the network directory in a file called 3047. (The distribution number is always two digits to avoid ambiguity.)
6. The process is repeated from Step 1.

The hub alternates between two tasks, namely giving instructions (Part 1), and loading and consolidating results (Part 2). Since Part 2 takes quite long, Part 1 is repeated for 10 minutes, after which the hub executes Part 2, and then returning to Part 1 for 10 minutes, and so on, until the entire simulation is done.

In Part 1 the following steps are performed:

1. The hub scans the network directory for any request file (a filename starting with "r").
2. Suppose it finds r25. It then loads the batch.
3. The hub then reads the first line of the batch, and saves that line in a file called g25 on the network directory.
4. It then removes the first line of the batch and saves the batch again.
5. It then deletes r25 from the network directory.

In Part 2 the following steps are performed:

1. The hub scans the network directory for any of the possible output files.
2. Suppose it finds 3047. It then opens the appropriate output data file where the simulation results are stored.
3. The hub consolidates the results of 3047 into the format of the output data file and enters them into the matrix containing the results.
4. It then saves and closes the output data file.
5. It then deletes file 3047 from the network directory.

Note that working directly from the batch (that is, each computer opening the batch, reading and deleting the first line), does not work, since violations occur when two or more computers attempt to open the batch file simultaneously. This is the reason the request file method had to be developed.

Conclusion

In this chapter we developed an estimator, which can be viewed as a benchmark estimator of EVI, to test the performance of other estimators against.

We defined the PPD and saw that the MLE of the parameter γ can be interpreted as a generalization of the Hill estimator. The methodology behind the Bayesian estimation of the PPD parameters were discussed in detail in Section 3.2, and the performance of the estimator at fixed thresholds was assessed by means of a simulation study, as discussed in Section 3.3.

In Section 3.4 we developed an adaptive threshold selection technique based on the stability of the estimates over a range of thresholds. We also showed through simulation that this technique improves on the performance of estimation at a fixed threshold.

Section 3.4.4 summarized what is meant by the *PPD estimator*, as we will refer to it in the rest of the thesis. We showed the construction of the estimator and presented tables which can be used to compare the performance of this estimator to that of other estimators.

In Section 3.5 we discussed some details of the programming methodology we applied.

This chapter has justified and established the design of a simulation study which was used to assess the performance of the newly constructed PPD estimator. In the next chapter we apply the same design to simulation studies performed on numerous other estimators in the literature, to assess the performance of the PPD estimator relative to them.

Chapter 4

Other estimators of the EVI

In this chapter we compare the performance of the PPD estimator to that of various other estimators in the literature. We also suggest a slight modification of the PPD estimator for samples of sizes $n = 2000$ and $n = 5000$, which leads to an improved PPD estimator.

In Section 4.1 we describe a method of estimating the second order parameter ρ, using the method by Gomes and Martins (2001), which we use as an external estimator of ρ in Section 4.2.

The other estimators we consider in this chapter are the following:

- In Section 4.2 we consider other estimators based on the PPD. The PPD is fitted to the multiplicative excesses, but using different methods of estimating the parameters.
- In Section 4.3 the performance of the Hill estimator is studied, where three methods of optimal threshold selection (and combinations of those) are considered.
- In Section 4.4 we consider estimators which can be seen as generalizations of the Hill estimator, in the sense that they are also based on the log-spacings of order statistics. This includes the moment estimator by Dekkers, Einmahl and De Haan (1989) and twelve estimators considered in the paper by Gomes and Martins (2002).
- Section 4.5 covers estimators involving the generalized Pareto distribution (GPD), specifically the MLE and Bayesian estimation of the GPD parameters.
- The Zipf estimator, the Pickands estimator and an estimator based on the exponential regression model are described in Section 4.6.

Section 4.7 shows the results of the simulation study of the estimators of Sections 4.4, 4.5 and 4.6, and these results are discussed in Section 4.8.

In Section 4.8 we also propose a modification of the PPD estimator for samples of sizes $n = 2000$ and $n = 5000$, which leads to an improved PPD estimator. Since the PPD estimator is computationally quite involved, we propose estimators which can be used as alternatives to the PPD estimator. Included are simulation study results, showing the performance of the estimator involving the PPD, and the alternative method of estimation.

In Section 4.9 we show some results concerning the bias of some of the estimators considered in Section 4.8, and we also give the justification for choosing $-2 \leq \rho \leq 0$ as the range of the second order parameter for the distributions from which we simulate data.

4.1 Estimation of the second order parameter

In Section 3.1.5 we mentioned that the method of estimation of the second order parameter ρ is one of the choices which has to be made when fitting the PPD to a set of mutliplicative excesses.

Particularly, we have to choose between three possible approaches:

- Restricting the range of ρ. We have followed this approach in Chapter 3 by applying the restriction $-2 \leq \rho < 0$.
- Not estimating ρ at all, but fixing it at some value, usually -1.
- Estimating ρ separately by a method suggested in the literature, and then estimating γ and c, given this value of ρ.

In this section we will describe a method for estimating ρ, which we will use when investigating the third approach mentioned above.

Several estimators exist in the literature. The method of estimation of ρ which we will describe here, and use in subsequent sections, is the one by Gomes and Martins (2001). We will only state the procedure here. For a full discussion of the estimator, see Gomes and Martins (2001).

Given a set of n positive, ordered observations $x_{1,n} \leq x_{2,n} \leq \ldots \leq x_{n,n}$ assumed to be from a second order regularly varying distribution, we define the following functions:

- For a given value of α and k, let $m_n^{(\alpha)}(k) = \frac{1}{k}\sum_{i=1}^{k} \left(\log x_{n-i+1,n} - \log x_{n-k,n}\right)^{\alpha}$.
- For a given value of α and k, let $\hat{\gamma}_n^{(\alpha)}(k) = \frac{m_n^{(\alpha)}(k)}{\Gamma(\alpha+1)(H_{k,n})^{\alpha-1}}$. Recall that $H_{k,n}$ is the Hill estimator, based on k excesses (refer to Section 3.1.3).
- For a given value of α, i_n and j_n, let $\chi(\alpha, i_n, j_n)$ be the sample median of $\hat{\gamma}_n^{(\alpha)}(k)$ for $i_n \leq k \leq j_n$.
- For a given value of α, i_n and j_n, let $\Sigma(\alpha, i_n, j_n) = \sum_{k=i_n}^{j_n} \left(\hat{\gamma}_n^{(\alpha)}(k) - \chi(\alpha, i_n, j_n)\right)^2$.

The procedure of estimating ρ is as follows:

1. Choose i_n and j_n which specify a range of k such that the estimates $\hat{\gamma}_n^{(\alpha)}(k)$ are "stable" for $i_n \leq k \leq j_n$. According to Gomes and Martins (2001) the technique is robust with respect to the choice of i_n and j_n. We will choose $i_n = \frac{n}{4}$ and $j_n = \frac{n}{2}$, rounded to the nearest integer.
2. Obtain an estimate of α as $\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \Sigma(\alpha, i_n, j_n)$.
3. Estimate ρ by $\hat{\rho}$ as the solution of $(1-\hat{\rho})^{\hat{\alpha}-1}\left(1+\hat{\rho}(\hat{\alpha}-2)\right) = 1$.

According to Gomes and Martins (2001), there is a one-to-one correspondence between the parameters α and ρ, given by the equation

$$(1-\rho)^{\alpha-1}\big(1+\rho(\alpha-2)\big)=1,$$

which is used in Step 3 to solve for $\hat{\rho}$ for a given value of $\hat{\alpha}$.

The restriction $\alpha \geq 2$ applies (Gomes and Martins, 2001). In particular, the value of $\alpha = 2.09272$ corresponds to $\rho = -10$ and $\alpha = 15.02746$ to $\rho = -0.01$.

The following graph shows the relationship between ρ and α. The graph of ρ as a function of α is not very informative. We show $\log(-\rho)$ as a function of α, for $2.1 \leq \alpha \leq 15.03$:

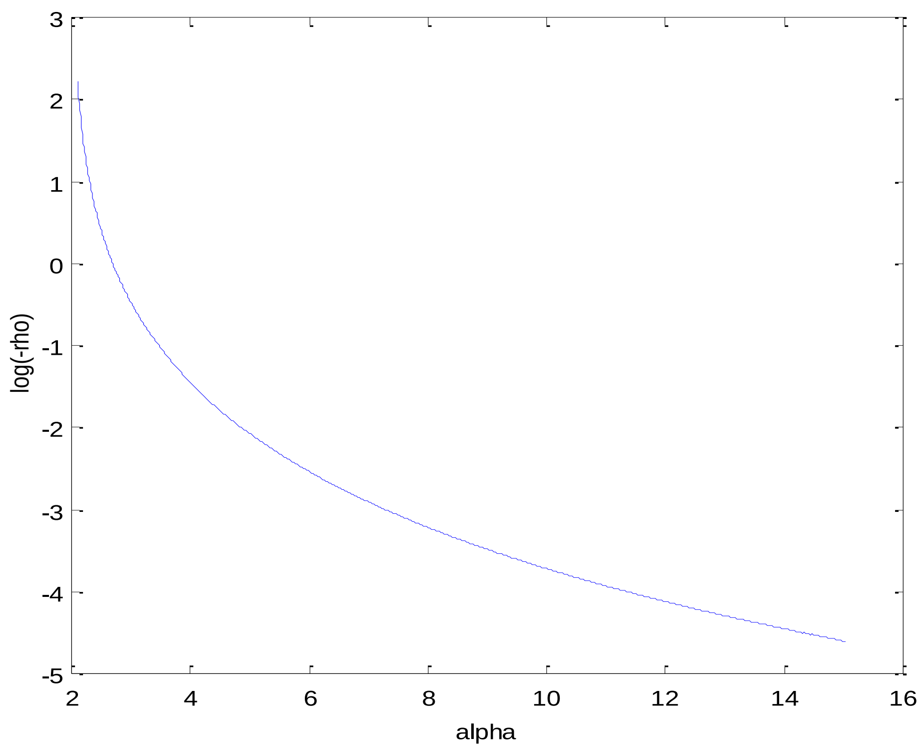

Figure 4.1.1 $\log(-\rho)$ as a function of α.

From the graph we can see that $\rho \to -\infty$ as $\alpha \to 2$, and $\rho \to 0$ as $\alpha \to \infty$. These limits can be obtained mathematically, even though it is not straightforward.

Computationally, we let $\hat{\rho} = -10$, if $2 \leq \hat{\alpha} < 2.1$, and $\hat{\rho} = -0.01$, if $\hat{\alpha} > 15$. For $2.1 \leq \hat{\alpha} \leq 15$ we solve for $\hat{\rho}$ numerically from the equation in Step 3. We therefore place the restriction $-10 \leq \hat{\rho} \leq -0.01$ on $\hat{\rho}$.

On another computational aspect, note that $\Sigma(\alpha, i_n, j_n)$ as a function of α will typically have one of three shapes. The shape of $\Sigma(\alpha, i_n, j_n)$ as a function of α should be taken into account, since it affects the way in which the minimization should be done.

The first shape sees $\Sigma(\alpha, i_n, j_n)$ decreasing as α increases (for all $2 < \alpha < 15$). In such a case we take $\hat{\alpha} = 15$, yielding $\hat{\rho} = -0.01$.

The second typical shape is shown in the following graph:

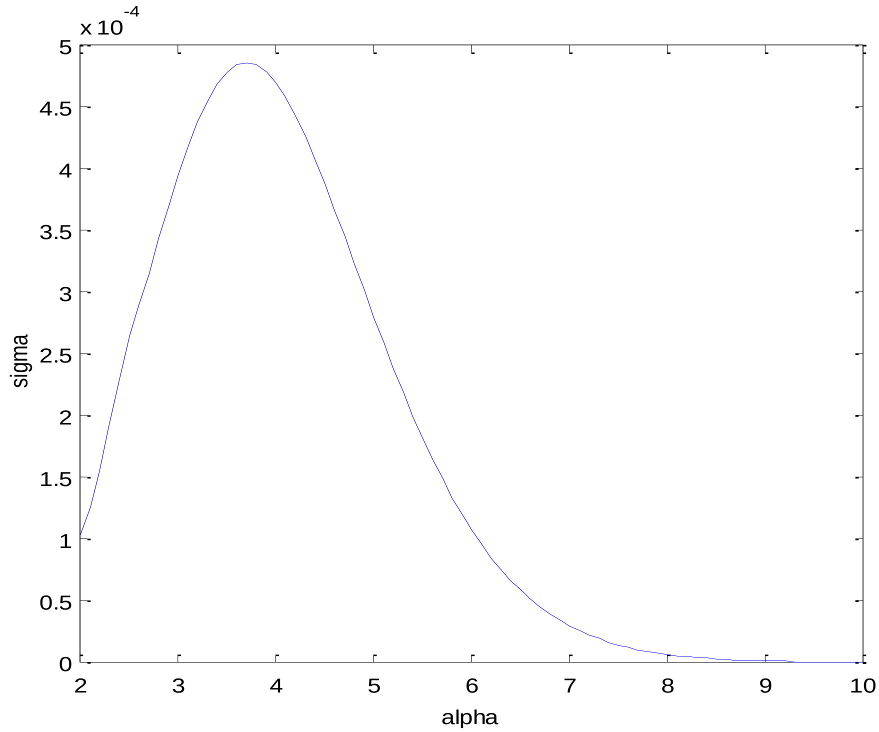

Figure 4.1.2 Typical behaviour of $\Sigma(\alpha, i_n, j_n)$ as a function of α.

In this case we do not regard the value of $\Sigma(\alpha, i_n, j_n)$ at $\alpha = 2$ as a local minimum. In this second case, we also take $\hat{\alpha} = 15$, yielding $\hat{\rho} = -0.01$.

The third typical shape is the following:

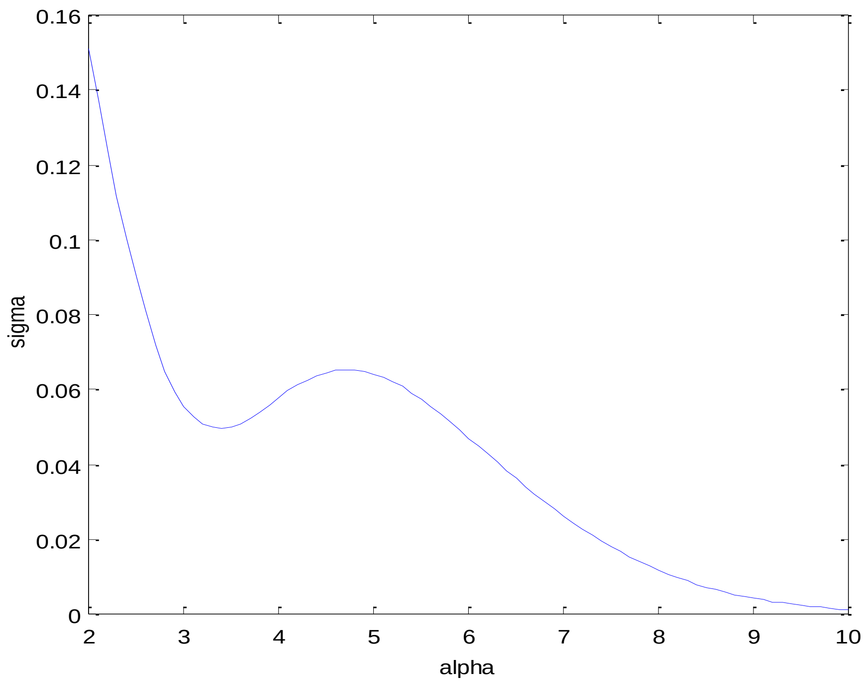

Figure 4.1.3 Typical behaviour of $\Sigma(\alpha, i_n, j_n)$ as a function of α.

Note that $\hat{\gamma}_n^{(\alpha)}(k) \to 0$ (and consequently $\Sigma(\alpha, i_n, j_n) \to 0$) as $\alpha \to \infty$, and that $\hat{\rho} = 0$ is always a (trivial) solution of $(1-\hat{\rho})^{\hat{\alpha}-1}\big(1+\hat{\rho}(\hat{\alpha}-2)\big) = 1$. In order to determine the value $\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \Sigma(\alpha, i_n, j_n)$, we have to take care to determine the local minimum of $\Sigma(\alpha, i_n, j_n)$ before the global minimum of 0 is reached when $\alpha \to \infty$.

In the case shown in Figure 4.1.3, the estimate of α is $\hat{\alpha} = 3.4$, which corresponds to $\hat{\rho} = -0.39$.

4.2 Other estimators based on the PPD

In Chapter 3 we constructed the benchmark estimator by determining the Bayesian estimates of all three PPD parameters, and by using an adaptive threshold selection technique based on the stability of the estimates of the EVI at fixed thresholds.

In this section we consider other options of obtaining an estimate of the PPD parameter γ. In Section 3.1.5 we mentioned some of the different options which can be considered when estimating the PPD parameter γ.

The first choice one has to make, involves the method of estimation of the second order parameter ρ. In this thesis, we consider three possible approaches:

1. Restricting the range of ρ. We have followed this approach in Chapter 3 by applying the restriction $-2 \leq \rho < 0$.
2. Not estimating ρ at all, but fixing it at some value, usually -1.
3. Estimating ρ separately by a method suggested in the literature, and then estimating γ and c, given this value of ρ. In Section 4.1 we discussed the method of estimating ρ as proposed by Gomes and Martins (2001).

The second choice, is the choice of threshold. In this section we do not derive a threshold selection technique, but determine which of the fixed thresholds ($k = 5\%, 10\%, \dots, 95\%$ of n) yields the lowest MSE.

The third choice which has to be made is between Bayesian estimation and MLE. We consider both these approaches in this section.

The computation of the MLEs of the parameters of the PPD is straightforward. The negative of the log likelihood is minimized across all parameters. Recall that the density function of the PPD is

$$f(z|\gamma,\rho,c) = \gamma^{-1} z^{-1/\gamma-1}(1 - c + c(1-\rho)z^{\rho/\gamma}).$$

For observed multiplicative excesses $z_1, z_2, \dots, z_k$, the log likelihood is given by

$$-k\log\gamma - (1 + \tfrac{1}{\gamma})\textstyle\sum_{i=1}^{k} \log z_i + \sum_{i=1}^{k} \log(1 - c + c(1-\rho)z_i^{\rho/\gamma}),$$

where $\gamma > 0$, $\rho < 0$, $\rho^{-1} \leq c \leq 1$ and $z_i \geq 1$ for all $i = 1,2,\dots,k$.

If ρ is estimated, we apply the same restriction to ρ as we did in the case of Bayesian estimation, namely $-2 \leq \rho < 0$. As mentioned in Section 3.1.5, it is difficult to estimate the second order parameter ρ accurately. One can see this when considering the likelihood as a function of ρ for given values of γ and c. The result is usually a very flat likelihood function, resulting in an unreliable MLE of ρ, since small changes in the parameters γ and c cause the value of ρ at which a maximum is reached, to change dramatically. As a consequence MLEs of ρ are close to -2 or 0 most of the time.

The results of the simulation study are shown in the tables which follow. Table 4.2.1 indicates which thresholds yielded the lowest MSE. Tables 4.2.2 to 4.2.10 show the actual lowest 100×MSE values which were obtained (and their corresponding standard errors), for sizes $n = 50$ to $n = 20\,000$, respectively.

The columns of Table 4.2.1 are divided into two parts, namely results obtained from Bayesian estimation, and results obtained from MLE, as indicated in the first row. The second row indicates the choice made with respect to ρ. "Est" indicates that ρ was estimated simultaneously with γ and c, and applying the restriction $-2 \leq \rho < 0$ (Option 1 above). "-1"indicates that ρ was not estimated, but fixed at -1 (Option 2 above). "Gomes" indicates that we estimated ρ using the technique by Gomes and Martins (2001), after which we estimate γ and c (Option 3 above).

In the case of Bayesian estimation, Table 4.2.1 also shows which method of estimation (mode, median or mean of Gibbs draws) yielded the lowest MSE. For instance, consider "Bayes Est" represented in the third and fourth columns. The first entry indicates that, when using the "Bayes Est" method of estimation, for samples of size $n = 50$ and distributions for which $\gamma = 0.1$, the average MSE was lowest when using a fixed threshold of 40% and calculating the mode of the Gibbs draws. Note that the information in the "Bayes Est" columns can be found in Tables 3.4.3.1 to 3.4.3.9.

For more detail (in the case of $n = 50$), we refer to Table 4.2.2. There we see that the average 100×MSE, for distributions of the EVI group $\gamma = 0.1$, was 0.0803 (in bold), with a standard error of 0.0016. Also, Distribution 2, for instance, had an average 100×MSE of 0.0524. Both values 0.0803 and 0.0016 (and all values for distributions of the EVI group $\gamma = 0.1$) were obtained at $k = 40\%$ of n, as indicated in Table 4.2.1.

Note that the first five columns of Table 4.2.2 are the same as that of Table 3.4.4.2. Similarly for the other sample sizes. This illustrates the way we use Tables 3.4.4.2 to 3.4.4.6 to compare the results of estimators. Also note that we use the results in the fifth column of Tables 3.4.4.2 to 3.4.4.6 (indicated "Fixed") to compare to the results of the other estimators based on the PPD. We do not use the benchmark PPD estimator (indicated "PPD"), since we wish to assess the effect of the estimation of ρ on the accuracy of estimation of γ, and therefore we do not want to bring the effect of threshold selection into play.

		Bayes						MLE		
Sample size		**Est**		**-1**		**Gomes**		**Est**	**-1**	**Gomes**
50	γ **= 0.1**	40%	Mode	30%	Mode	35%	Mode	65%	15%	40%
	γ **= 0.5**	40%	Mode	30%	Mode	40%	Mode	55%	30%	40%
	γ **= 1**	35%	Median	30%	Mode	40%	Median	65%	25%	45%
	Mean	35%	Median	30%	Mode	35%	Median	70%	25%	40%
100	γ **= 0.1**	35%	Mode	25%	Mode	35%	Mode	80%	10%	70%
	γ **= 0.5**	40%	Mode	25%	Mode	40%	Mode	70%	35%	40%
	γ **= 1**	30%	Median	25%	Mode	35%	Median	70%	30%	40%
	Mean	30%	Median	25%	Mode	35%	Median	70%	30%	40%
200	γ **= 0.1**	35%	Mode	25%	Mode	35%	Mode	75%	5%	5%
	γ **= 0.5**	40%	Mode	25%	Mode	35%	Mode	60%	40%	40%
	γ **= 1**	35%	Median	25%	Mode	35%	Median	65%	40%	35%
	Mean	35%	Median	25%	Mode	35%	Median	65%	40%	35%
500	γ **= 0.1**	25%	Mode	15%	Mode	25%	Mode	75%	60%	80%
	γ **= 0.5**	35%	Mode	25%	Mode	30%	Mode	65%	30%	30%
	γ **= 1**	35%	Median	15%	Mode	35%	Median	65%	40%	35%
	Mean	35%	Median	15%	Mode	35%	Median	65%	40%	35%
1000	γ **= 0.1**	30%	Mode	20%	Mode	35%	Mode	50%	50%	50%
	γ **= 0.5**	30%	Median	20%	Mode	35%	Mode	50%	35%	30%
	γ **= 1**	30%	Median	15%	Mode	30%	Median	50%	40%	30%
	Mean	30%	Median	15%	Mode	30%	Median	50%	40%	30%
2000	γ **= 0.1**	25%	Mode	15%	Mode	30%	Mode	50%	50%	5%
	γ **= 0.5**	25%	Median	15%	Mode	25%	Mode	50%	40%	30%
	γ **= 1**	20%	Median	10%	Mode	25%	Mode	45%	35%	25%
	Mean	20%	Median	10%	Mode	25%	Mode	45%	35%	25%
5000	γ **= 0.1**	15%	Mode	30%	Mode	15%	Mode	45%	50%	5%
	γ **= 0.5**	25%	Median	5%	Mode	25%	Mode	50%	35%	20%
	γ **= 1**	20%	Median	5%	Mode	20%	Mode	45%	35%	25%
	Mean	20%	Median	5%	Mode	20%	Mode	45%	35%	25%
10 000	γ **= 0.1**	15%	Median	15%	Median	15%	Mode	40%	50%	30%
	γ **= 0.5**	20%	Median	20%	Mode	15%	Mode	30%	30%	20%
	γ **= 1**	25%	Median	15%	Mode	15%	Mode	35%	30%	10%
	Mean	25%	Median	15%	Mode	15%	Mode	35%	30%	20%
20 000	γ **= 0.1**	25%	Median	15%	Mode	10%	Mode	35%	50%	25%
	γ **= 0.5**	15%	Median	15%	Mode	10%	Mode	35%	25%	15%
	γ **= 1**	10%	Median	10%	Mode	15%	Mode	35%	30%	15%
	Mean	15%	Median	15%	Mode	15%	Mode	35%	30%	15%

Table 4.2.1 Fixed thresholds for other estimators based on the PPD.

				Bayes			MLE		
Distr.		γ	ρ	**Est**	**-1**	**Gomes**	**Est**	**-1**	**Gomes**
1	Burr	0.1	-2	0.0832	0.0829	0.1144	0.2446	0.2961	0.2604
				(0.0027)	(0.0040)	(0.0037)	(0.0249)	(0.0236)	(0.0345)
2	Burr	0.1	-1	0.0524	0.0731	0.0700	0.1362	0.2962	0.3152
				(0.0020)	(0.0036)	(0.0027)	(0.0283)	(0.0240)	(0.0692)
3	Burr	0.1	-0.5	0.1446	0.2043	0.1163	0.3405	0.3988	0.5129
				(0.0064)	(0.0093)	(0.0062)	(0.0132)	(0.0301)	(0.1154)
4	Fréchet	0.1	-1	0.0659	0.0691	0.0930	0.2488	0.2627	0.2527
				(0.0022)	(0.0029)	(0.0030)	(0.0287)	(0.0163)	(0.0553)
5	logΓ	0.1	0	0.0553	0.0972	0.0634	0.3705	0.3260	0.3805
				(0.0026)	(0.0047)	(0.0029)	(0.0406)	(0.0221)	(0.0875)
γ **= 0.1**				**0.0803**	**0.1053**	**0.0914**	**0.2681**	**0.3160**	**0.3443**
				(0.0016)	(0.0024)	(0.0017)	(0.0127)	(0.0106)	(0.0346)
6	Burr	0.5	-2	2.2085	1.9619	2.7475	4.3180	3.0512	2.7242
				(0.0641)	(0.0679)	(0.0727)	(0.3880)	(0.1548)	(0.1403)
7	Burr	0.5	-1	1.2569	1.5785	1.5958	2.9903	3.2839	2.3295
				(0.0495)	(0.0697)	(0.0654)	(0.4277)	(0.2140)	(0.2904)
8	Burr	0.5	-0.5	3.0273	3.9853	2.9164	5.9794	7.0532	4.3073
				(0.1409)	(0.1797)	(0.1735)	(0.2791)	(0.4022)	(0.2227)
9	Fréchet	0.5	-1	1.8278	1.7321	2.3009	3.5867	3.0221	2.4366
				(0.0575)	(0.0659)	(0.0712)	(0.3201)	(0.1803)	(0.1577)
10	*t*	0.5	-1	1.0614	1.5456	1.2649	1.8171	3.2721	1.8495
				(0.0479)	(0.0797)	(0.0583)	(0.1934)	(0.2111)	(0.0929)
11	logΓ	0.5	0	1.2272	1.9442	1.4445	5.9238	4.8256	2.6296
				(0.0500)	(0.0988)	(0.0693)	(0.5403)	(0.2883)	(0.2135)
γ **= 0.5**				**1.7682**	**2.1246**	**2.0450**	**4.1025**	**4.0847**	**2.7128**
				(0.0310)	(0.0416)	(0.0383)	(0.1530)	(0.1043)	(0.0804)
12	Burr	1	-2	6.8492	8.1036	8.3039	11.7459	12.2713	9.1825
				(0.2321)	(0.2750)	(0.2501)	(1.1371)	(0.5845)	(0.3621)
13	Burr	1	-1	4.4277	5.8132	5.1039	7.0578	12.5479	7.2621
				(0.1938)	(0.2690)	(0.2211)	(1.1036)	(0.7427)	(0.3825)
14	Burr	1	-0.5	12.5338	11.7886	12.3072	30.6245	24.1105	18.4599
				(0.4969)	(0.5053)	(0.5589)	(0.8934)	(1.3639)	(0.8158)
15	Fréchet	1	-1	5.8579	7.2204	7.2687	13.7923	13.2101	9.6318
				(0.2212)	(0.2703)	(0.2680)	(1.6908)	(0.6922)	(0.5808)
16	*t*	1	-2	6.4216	7.8194	7.8315	10.2659	12.7102	9.3037
				(0.2170)	(0.2607)	(0.2411)	(1.5322)	(0.7070)	(0.3834)
17	logΓ	1	0	4.5733	6.0768	5.0719	22.3057	18.3367	9.7539
				(0.2116)	(0.3001)	(0.2697)	(2.6896)	(1.1991)	(0.6010)
γ **= 1**				**6.7773**	**7.8037**	**7.6479**	**15.9653**	**15.5311**	**10.5990**
				(0.1154)	(0.1327)	(0.1319)	(0.6614)	(0.3790)	(0.2229)
Mean				**3.1024**	**3.5351**	**3.4780**	**7.1837**	**7.0646**	**4.8406**
				(0.0428)	(0.0491)	(0.0466)	(0.2184)	(0.1392)	(0.0908)

Table 4.2.2 Results for $n = 50$ for other estimators based on the PPD.

				Bayes			MLE		
Distr.		γ	ρ	**Est**	**-1**	**Gomes**	**Est**	**-1**	**Gomes**
1	Burr	0.1	-2	0.0660	0.0580	0.0779	0.0483	0.2645	0.2634
				(0.0030)	(0.0032)	(0.0035)	(0.0062)	(0.0263)	(0.0685)
2	Burr	0.1	-1	0.0372	0.0501	0.0460	0.0305	0.3001	0.0571
				(0.0022)	(0.0038)	(0.0027)	(0.0023)	(0.0373)	(0.0053)
3	Burr	0.1	-0.5	0.0916	0.1366	0.0982	0.6677	0.3904	0.6268
				(0.0061)	(0.0095)	(0.0070)	(0.0139)	(0.0498)	(0.0253)
4	Fréchet	0.1	-1	0.0516	0.0513	0.0593	0.0660	0.2665	0.3085
				(0.0026)	(0.0033)	(0.0027)	(0.0238)	(0.0283)	(0.0747)
5	logΓ	0.1	0	0.0435	0.0781	0.0481	0.1848	0.2997	0.5017
				(0.0032)	(0.0063)	(0.0050)	(0.0399)	(0.0282)	(0.1432)
γ **= 0.1**				**0.0580**	**0.0748**	**0.0659**	**0.1994**	**0.3043**	**0.3515**
				(0.0016)	(0.0026)	(0.0020)	(0.0098)	(0.0157)	(0.0355)
6	Burr	0.5	-2	1.6855	1.5822	1.9005	2.7351	2.7116	1.7690
				(0.0749)	(0.0785)	(0.0786)	(0.5946)	(0.2088)	(0.1072)
7	Burr	0.5	-1	0.9130	1.1921	1.0812	0.6644	2.1456	1.2553
				(0.0469)	(0.0723)	(0.0572)	(0.0608)	(0.2464)	(0.0716)
8	Burr	0.5	-0.5	2.2361	2.5467	2.4417	7.6548	5.1114	2.9053
				(0.1416)	(0.1685)	(0.1645)	(0.2089)	(0.2620)	(0.1950)
9	Fréchet	0.5	-1	1.2656	1.3483	1.4771	1.5378	2.7898	1.6523
				(0.0588)	(0.0770)	(0.0733)	(0.2545)	(0.2364)	(0.1610)
10	*t*	0.5	-1	0.6788	1.2455	0.9103	1.7650	2.0575	1.3193
				(0.0471)	(0.1069)	(0.0605)	(0.0998)	(0.2218)	(0.1036)
11	logΓ	0.5	0	0.8112	1.6394	1.0450	3.4532	4.2184	1.8452
				(0.0511)	(0.1305)	(0.1032)	(0.6323)	(0.3722)	(0.2003)
γ **= 0.5**				**1.2650**	**1.5924**	**1.4760**	**2.9684**	**3.1724**	**1.7911**
				(0.0317)	(0.0454)	(0.0395)	(0.1559)	(0.1076)	(0.0604)
12	Burr	1	-2	5.5800	6.0703	6.5162	7.9588	9.3989	6.5686
				(0.2617)	(0.3007)	(0.2946)	(1.4370)	(0.7633)	(0.3341)
13	Burr	1	-1	3.4861	4.1184	3.6519	2.6298	8.0286	4.9980
				(0.1970)	(0.2319)	(0.1954)	(0.2569)	(0.7680)	(0.2998)
14	Burr	1	-0.5	6.8986	8.4193	9.9358	31.1774	18.8755	12.5259
				(0.4291)	(0.4971)	(0.6208)	(0.7999)	(1.1434)	(0.8333)
15	Fréchet	1	-1	4.5796	4.8936	5.0977	7.6203	8.2811	5.7158
				(0.2281)	(0.2434)	(0.2755)	(1.6890)	(0.7198)	(0.5298)
16	*t*	1	-2	5.5431	5.9004	6.0910	3.8303	9.4605	6.4950
				(0.2540)	(0.2812)	(0.2799)	(0.1645)	(0.9460)	(0.4130)
17	logΓ	1	0	3.7183	5.0517	3.6358	12.3425	16.9010	7.5199
				(0.2501)	(0.3310)	(0.3257)	(1.8665)	(1.4478)	(0.8553)
γ **= 1**				**4.9676**	**5.7423**	**5.8214**	**10.9265**	**11.8243**	**7.3039**
				(0.1143)	(0.1332)	(0.1463)	(0.5037)	(0.4079)	(0.2403)
Mean				**2.2254**	**2.6107**	**2.6383**	**4.9651**	**5.4741**	**3.3211**
				(0.0421)	(0.0497)	(0.0547)	(0.1862)	(0.1508)	(0.0887)

Table 4.2.3 Results for $n = 100$ for other estimators based on the PPD.

				Bayes			MLE		
Distr.		γ	ρ	**Est**	**-1**	**Gomes**	**Est**	**-1**	**Gomes**
1	Burr	0.1	-2	0.0514	0.0522	0.0610	0.0491	0.2356	0.3100
				(0.0029)	(0.0036)	(0.0034)	(0.0119)	(0.0178)	(0.0457)
2	Burr	0.1	-1	0.0311	0.0398	0.0310	0.0133	0.2741	0.4144
				(0.0017)	(0.0032)	(0.0017)	(0.0012)	(0.0259)	(0.0741)
3	Burr	0.1	-0.5	0.0724	0.1202	0.1002	0.4345	0.3012	0.3700
				(0.0043)	(0.0062)	(0.0070)	(0.0070)	(0.0355)	(0.0861)
4	Fréchet	0.1	-1	0.0388	0.0385	0.0381	0.0848	0.2437	0.3152
				(0.0020)	(0.0026)	(0.0019)	(0.0315)	(0.0263)	(0.0541)
5	logΓ	0.1	0	0.0317	0.0834	0.0274	0.1919	0.3189	0.3528
				(0.0022)	(0.0074)	(0.0022)	(0.0415)	(0.0322)	(0.0636)
γ **= 0.1**				**0.0451**	**0.0668**	**0.0515**	**0.1547**	**0.2747**	**0.3525**
				(0.0012)	(0.0022)	(0.0017)	(0.0108)	(0.0126)	(0.0296)
6	Burr	0.5	-2	1.1216	0.9241	1.3391	2.5955	2.3986	1.0242
				(0.0514)	(0.0536)	(0.0589)	(0.5341)	(0.2467)	(0.0548)
7	Burr	0.5	-1	0.7122	0.8427	0.8842	0.5947	1.2209	0.8685
				(0.0399)	(0.0564)	(0.0437)	(0.0428)	(0.1848)	(0.0466)
8	Burr	0.5	-0.5	1.7391	2.5665	2.0975	3.5388	5.5817	2.8242
				(0.0943)	(0.1381)	(0.1451)	(0.1058)	(0.1772)	(0.1638)
9	Fréchet	0.5	-1	0.9289	0.9376	1.0768	1.6637	2.2434	0.9792
				(0.0460)	(0.0676)	(0.0526)	(0.4040)	(0.2498)	(0.0814)
10	*t*	0.5	-1	0.4378	0.7792	0.7135	0.4562	1.3734	0.8554
				(0.0320)	(0.0637)	(0.0416)	(0.0427)	(0.2021)	(0.0590)
11	logΓ	0.5	0	0.7171	1.4881	0.6902	3.1599	3.2887	1.2817
				(0.0559)	(0.1256)	(0.0610)	(0.7376)	(0.3207)	(0.1530)
γ **= 0.5**				**0.9428**	**1.2564**	**1.1335**	**8.7342**	**2.6844**	**1.3055**
				(0.0232)	(0.0371)	(0.0310)	(2.1446)	(0.0961)	(0.0427)
12	Burr	1	-2	3.8053	3.7914	4.1197	8.7342	8.1560	4.4115
				(0.1846)	(0.2155)	(0.1976)	(2.1446)	(0.8823)	(0.3037)
13	Burr	1	-1	2.5121	3.0417	2.7449	1.8973	5.6397	3.7209
				(0.1652)	(0.2335)	(0.1643)	(0.1523)	(0.8793)	(0.2139)
14	Burr	1	-0.5	6.8924	9.0264	9.9437	20.1647	22.4159	9.4752
				(0.3808)	(0.5625)	(0.6032)	(0.4663)	(0.8054)	(0.6385)
15	Fréchet	1	-1	3.1983	3.4607	3.4458	7.5069	7.0009	4.1845
				(0.2307)	(0.2451)	(0.1981)	(1.9641)	(0.8144)	(0.2347)
16	*t*	1	-2	3.4984	3.9012	4.0026	3.5469	7.5696	4.4946
				(0.1761)	(0.2128)	(0.1891)	(0.4730)	(0.9534)	(0.2235)
17	logΓ	1	0	3.1509	4.8199	2.5553	15.3216	10.4374	5.1437
				(0.2367)	(0.3717)	(0.2557)	(3.6785)	(0.8520)	(0.4048)
γ **= 1**				**3.8429**	**4.6735**	**4.4687**	**9.5286**	**10.2033**	**5.2384**
				(0.0981)	(0.1355)	(0.1259)	(0.7897)	(0.3535)	(0.1504)
Mean				**1.7270**	**2.1126**	**2.0235**	**4.1891**	**4.6712**	**2.4691**
				(0.0361)	(0.0496)	(0.0463)	(0.2857)	(0.1294)	(0.0578)

Table 4.2.4 Results for $n = 200$ for other estimators based on the PPD.

				Bayes			MLE		
Distr.		γ	ρ	**Est**	**-1**	**Gomes**	**Est**	**-1**	**Gomes**
1	Burr	0.1	-2	0.0329	0.0345	0.0805	0.0249	0.0920	0.0366
				(0.0027)	(0.0035)	(0.0295)	(0.0017)	(0.0305)	(0.0016)
2	Burr	0.1	-1	0.0256	0.0267	0.0238	0.0046	0.0370	0.1591
				(0.0021)	(0.0035)	(0.0021)	(0.0008)	(0.0019)	(0.0114)
3	Burr	0.1	-0.5	0.0307	0.0660	0.0514	0.4057	0.5820	1.2041
				(0.0041)	(0.0070)	(0.0061)	(0.0069)	(0.0095)	(0.0433)
4	Fréchet	0.1	-1	0.0302	0.0388	0.0312	0.0202	0.2340	0.0569
				(0.0029)	(0.0060)	(0.0025)	(0.0011)	(0.0507)	(0.0313)
5	logΓ	0.1	0	0.0285	0.0573	0.0255	0.0377	0.4620	0.0361
				(0.0050)	(0.0068)	(0.0038)	(0.0037)	(0.0730)	(0.0040)
γ **= 0.1**				**0.0296**	**0.0447**	**0.0425**	**0.0986**	**0.2814**	**0.2986**
				(0.0016)	(0.0025)	(0.0061)	(0.0016)	(0.0189)	(0.0110)
6	Burr	0.5	-2	0.7674	0.6830	0.8632	1.3086	2.1766	0.7232
				(0.0656)	(0.0925)	(0.0722)	(0.7127)	(0.3701)	(0.0728)
7	Burr	0.5	-1	0.5768	0.4489	0.5710	0.2496	1.1156	0.6828
				(0.0477)	(0.0454)	(0.0459)	(0.0314)	(0.3148)	(0.0503)
8	Burr	0.5	-0.5	0.9604	2.1598	1.3059	4.8342	3.0282	1.4727
				(0.0715)	(0.1365)	(0.1119)	(0.1122)	(0.1432)	(0.1277)
9	Fréchet	0.5	-1	0.6675	0.5851	0.6873	0.5162	2.3966	0.6431
				(0.0553)	(0.0753)	(0.0585)	(0.0351)	(0.4454)	(0.0687)
10	*t*	0.5	-1	0.3211	0.3901	0.5226	0.4002	0.4596	0.5541
				(0.0262)	(0.0573)	(0.0467)	(0.0236)	(0.0863)	(0.0492)
11	logΓ	0.5	0	0.5285	1.3481	0.4763	1.0052	3.0421	0.7874
				(0.0633)	(0.1555)	(0.0453)	(0.3356)	(0.5161)	(0.0643)
γ **= 0.5**				**0.6370**	**0.9358**	**0.7377**	**1.3857**	**2.0364**	**0.8105**
				(0.0232)	(0.0416)	(0.0276)	(0.1329)	(0.1423)	(0.0314)
12	Burr	1	-2	2.2137	2.8880	2.0371	7.5851	8.6839	2.0399
				(0.1761)	(0.2888)	(0.1450)	(3.7930)	(1.7123)	(0.1705)
13	Burr	1	-1	1.4392	2.5463	1.6370	0.8142	3.0205	1.9151
				(0.1339)	(0.2529)	(0.1390)	(0.0897)	(1.2081)	(0.1778)
14	Burr	1	-0.5	4.7591	3.8679	7.3413	19.1007	21.1413	6.8039
				(0.2936)	(0.3285)	(0.5192)	(0.4214)	(0.6539)	(0.5000)
15	Fréchet	1	-1	1.7037	2.6406	1.6861	1.9868	3.7492	2.0934
				(0.1383)	(0.2406)	(0.1623)	(0.1274)	(1.1927)	(0.1840)
16	*t*	1	-2	2.1983	2.7300	1.7067	2.7478	4.1185	2.0164
				(0.1807)	(0.2823)	(0.1461)	(0.1658)	(1.2158)	(0.1712)
17	logΓ	1	0	2.8202	3.8364	2.5668	8.5097	8.5818	3.6317
				(0.3536)	(0.3983)	(0.3913)	(4.9351)	(1.4032)	(0.4485)
γ **= 1**				**2.5224**	**3.0849**	**2.8292**	**6.7907**	**8.2159**	**3.0834**
				(0.0931)	(0.1238)	(0.1191)	(1.0405)	(0.5188)	(0.1264)
Mean				**1.1349**	**1.4525**	**1.3335**	**2.9299**	**3.7508**	**1.5559**
				(0.0341)	(0.0464)	(0.0589)	(0.3703)	(0.1885)	(0.0520)

Table 4.2.5 Results for $n = 500$ for other estimators based on the PPD.

				Bayes			MLE		
Distr.		γ	ρ	**Est**	**-1**	**Gomes**	**Est**	**-1**	**Gomes**
1	Burr	0.1	-2	0.0202	0.0252	0.0223	0.1090	0.0675	0.0320
				(0.0016)	(0.0036)	(0.0039)	(0.0438)	(0.0210)	(0.0106)
2	Burr	0.1	-1	0.0158	0.0226	0.0112	0.0142	0.0280	0.2090
				(0.0013)	(0.0072)	(0.0011)	(0.0009)	(0.0137)	(0.1133)
3	Burr	0.1	-0.5	0.0247	0.0597	0.0656	0.0570	0.3366	0.1697
				(0.0020)	(0.0037)	(0.0040)	(0.0021)	(0.0059)	(0.0068)
4	Fréchet	0.1	-1	0.0189	0.0254	0.0136	0.0805	0.0909	0.3815
				(0.0018)	(0.0050)	(0.0014)	(0.0426)	(0.0286)	(0.1343)
5	logΓ	0.1	0	0.0261	0.0450	0.0222	0.4969	0.9480	1.5602
				(0.0031)	(0.0043)	(0.0018)	(0.1471)	(0.0945)	(0.3119)
γ = 0.1				**0.0212**	**0.0356**	**0.0270**	**0.1515**	**0.2942**	**0.4705**
				(0.0009)	(0.0022)	(0.0012)	(0.0319)	(0.0204)	(0.0716)
6	Burr	0.5	-2	0.4716	0.5576	0.4540	2.1528	3.0191	0.4487
				(0.0422)	(0.0717)	(0.0385)	(0.9392)	(0.5421)	(0.0436)
7	Burr	0.5	-1	0.2875	0.3802	0.2875	0.4145	0.2131	0.3571
				(0.0265)	(0.0839)	(0.0262)	(0.0224)	(0.0218)	(0.0307)
8	Burr	0.5	-0.5	0.7987	1.5964	1.7890	1.3901	3.7277	1.3206
				(0.0592)	(0.1947)	(0.1110)	(0.0618)	(0.1107)	(0.0964)
9	Fréchet	0.5	-1	0.3934	0.3809	0.4032	0.4902	0.5752	0.4326
				(0.0317)	(0.0667)	(0.0322)	(0.0345)	(0.2382)	(0.0378)
10	*t*	0.5	-1	0.2043	0.2738	0.3229	0.0970	0.4907	0.2909
				(0.0184)	(0.0299)	(0.0382)	(0.0142)	(0.0418)	(0.0293)
11	logΓ	0.5	0	0.6697	1.1902	0.5617	1.9076	1.2371	0.8125
				(0.0608)	(0.1117)	(0.0466)	(1.2317)	(0.0769)	(0.0632)
γ = 0.5				**0.4709**	**0.7298**	**0.6364**	**1.0753**	**1.5438**	**0.6104**
				(0.0175)	(0.0434)	(0.0231)	(0.2585)	(0.1015)	(0.0226)
12	Burr	1	-2	2.1398	1.8234	1.9163	18.0372	7.4978	2.3641
				(0.1828)	(0.1699)	(0.1945)	(6.2712)	(1.7893)	(0.4852)
13	Burr	1	-1	1.2412	1.5716	1.4773	1.6090	0.9002	1.6257
				(0.0982)	(0.1610)	(0.1410)	(0.0980)	(0.0939)	(0.1449)
14	Burr	1	-0.5	2.6523	3.7462	5.0169	5.3202	20.7014	4.7079
				(0.1837)	(0.3149)	(0.3710)	(0.1732)	(0.4548)	(0.3547)
15	Fréchet	1	-1	1.4445	1.6338	1.2680	6.0157	3.1868	1.4837
				(0.1696)	(0.1751)	(0.1181)	(3.3161)	(1.1504)	(0.1319)
16	*t*	1	-2	2.0298	1.9846	1.5474	2.1422	4.1143	1.6776
				(0.1872)	(0.1783)	(0.1695)	(0.1836)	(1.3131)	(0.1856)
17	logΓ	1	0	2.5879	4.3238	2.3543	25.0245	5.4636	3.3385
				(0.2796)	(0.4367)	(0.1695)	(9.9831)	(0.3687)	(0.2376)
γ = 1				**2.0159**	**2.5139**	**2.2634**	**9.6914**	**6.9774**	**2.5329**
				(0.0780)	(0.1063)	(0.0861)	(2.0417)	(0.4282)	(0.1167)
Mean				**0.8847**	**1.1694**	**1.0620**	**3.8446**	**3.2047**	**1.2845**
				(0.0282)	(0.0413)	(0.0460)	(0.7264)	(0.1554)	(0.0475)

Table 4.2.6 Results for $n = 1000$ for other estimators based on the PPD.

				Bayes			MLE		
Distr.		γ	ρ	**Est**	**-1**	**Gomes**	**Est**	**-1**	**Gomes**
1	Burr	0.1	-2	0.0136	0.0332	0.0104	0.0586	0.0026	0.2250
				(0.0013)	(0.0089)	(0.0011)	(0.0331)	(0.0003)	(0.0678)
2	Burr	0.1	-1	0.0172	0.0160	0.0073	0.0122	0.0467	0.6929
				(0.0013)	(0.0044)	(0.0007)	(0.0007)	(0.0230)	(0.1442)
3	Burr	0.1	-0.5	0.0146	0.0671	0.0472	0.0546	0.3354	0.2051
				(0.0016)	(0.0209)	(0.0028)	(0.0013)	(0.0039)	(0.0328)
4	Fréchet	0.1	-1	0.0146	0.0282	0.0101	0.0464	0.0750	0.4678
				(0.0012)	(0.0086)	(0.0010)	(0.0326)	(0.0273)	(0.1255)
5	logΓ	0.1	0	0.0234	0.0697	0.0292	0.2581	1.0815	1.1226
				(0.0021)	(0.0166)	(0.0021)	(0.1049)	(0.1003)	(0.1980)
γ **= 0.1**				**0.0166**	**0.0428**	**0.0209**	**0.0860**	**0.3083**	**0.5427**
				(0.0007)	(0.0060)	(0.0008)	(0.0229)	(0.0213)	(0.0571)
6	Burr	0.5	-2	0.2881	0.4007	0.3570	0.3502	0.9337	0.3595
				(0.0249)	(0.0754)	(0.0335)	(0.0332)	(0.3149)	(0.0852)
7	Burr	0.5	-1	0.2773	0.2763	0.2103	0.3729	0.1505	0.1849
				(0.0209)	(0.0419)	(0.0206)	(0.0179)	(0.0119)	(0.0179)
8	Burr	0.5	-0.5	0.5040	1.0420	0.8208	1.2705	4.9956	1.1762
				(0.0429)	(0.0738)	(0.0614)	(0.0310)	(0.0893)	(0.0632)
9	Fréchet	0.5	-1	0.3098	0.5305	0.2893	0.3883	0.2310	0.2747
				(0.0317)	(0.1178)	(0.0318)	(0.0269)	(0.1340)	(0.0270)
10	*t*	0.5	-1	0.2255	0.4267	0.3853	0.0755	0.5189	0.1565
				(0.0191)	(0.1539)	(0.2025)	(0.0079)	(0.0201)	(0.0185)
11	logΓ	0.5	0	0.4538	1.0025	0.5662	0.4950	1.0669	0.6646
				(0.0356)	(0.1436)	(0.0410)	(0.0474)	(0.0411)	(0.0479)
γ **= 0.5**				**0.3431**	**0.6131**	**0.4382**	**0.4921**	**1.3161**	**0.4694**
				(0.0124)	(0.0444)	(0.0369)	(0.0123)	(0.0595)	(0.0204)
12	Burr	1	-2	1.7062	1.8992	1.4221	1.8856	5.4914	1.1557
				(0.1827)	(0.2056)	(0.1432)	(0.3119)	(1.5980)	(0.1372)
13	Burr	1	-1	0.9888	1.2871	0.8216	1.4903	0.5857	0.7897
				(0.0913)	(0.1833)	(0.0806)	(0.0775)	(0.0598)	(0.0766)
14	Burr	1	-0.5	1.6564	3.1051	3.7781	3.1835	15.2803	3.9907
				(0.1628)	(0.2798)	(0.2492)	(0.0960)	(0.2904)	(0.2541)
15	Fréchet	1	-1	1.5941	1.9400	1.2113	1.7604	2.1150	1.3764
				(0.1638)	(0.2144)	(0.1381)	(0.1244)	(1.0066)	(0.2131)
16	*t*	1	-2	1.4424	1.5703	1.0490	1.4955	2.1405	0.8680
				(0.1219)	(0.2061)	(0.1092)	(0.1374)	(0.9704)	(0.1023)
17	logΓ	1	0	1.8078	2.9484	2.3688	2.1551	4.4463	2.9092
				(0.1826)	(0.2966)	(0.1850)	(0.2272)	(0.2329)	(0.2175)
γ **= 1**				**1.5326**	**2.1250**	**1.7751**	**1.9951**	**5.0099**	**1.8483**
				(0.0631)	(0.0958)	(0.0655)	(0.0742)	(0.3594)	(0.0731)
Mean				**0.6725**	**0.9836**	**0.7900**	**1.0485**	**2.3592**	**1.0265**
				(0.0201)	(0.0361)	(0.0266)	(0.0751)	(0.1309)	(0.0361)

Table 4.2.7 Results for $n = 2000$ for other estimators based on the PPD.

				Bayes			MLE		
Distr.		γ	ρ	**Est**	**-1**	**Gomes**	**Est**	**-1**	**Gomes**
1	Burr	0.1	-2	0.0115	0.0022	0.0086	0.0090	0.0017	0.3591
				(0.0017)	(0.0003)	(0.0016)	(0.0014)	(0.0002)	(0.1509)
2	Burr	0.1	-1	0.0120	0.0025	0.0074	0.0126	0.0129	0.5273
				(0.0015)	(0.0003)	(0.0010)	(0.0008)	(0.0007)	(0.1409)
3	Burr	0.1	-0.5	0.0078	0.1093	0.0169	0.0331	0.3332	0.1641
				(0.0015)	(0.0025)	(0.0022)	(0.0010)	(0.0036)	(0.0398)
4	Fréchet	0.1	-1	0.0103	0.0024	0.0095	0.0116	0.0012	0.3060
				(0.0014)	(0.0004)	(0.0034)	(0.0011)	(0.0002)	(0.1004)
5	logΓ	0.1	0	0.0175	0.0389	0.0273	0.0170	1.2323	1.7837
				(0.0025)	(0.0020)	(0.0030)	(0.0023)	(0.1430)	(0.4784)
γ **= 0.1**				**0.0118**	**0.0310**	**0.0139**	**0.0167**	**0.3163**	**0.6281**
				(0.0008)	(0.0007)	(0.0011)	(0.0006)	(0.0286)	(0.1064)
6	Burr	0.5	-2	0.1214	0.6262	0.0946	0.1189	1.2864	0.1147
				(0.0133)	(0.1853)	(0.0160)	(0.0260)	(0.5499)	(0.0241)
7	Burr	0.5	-1	0.1750	0.3377	0.1087	0.3364	0.0838	0.1376
				(0.0185)	(0.0515)	(0.0157)	(0.0179)	(0.0095)	(0.0187)
8	Burr	0.5	-0.5	0.3253	0.4990	0.7706	1.2085	3.5821	0.5385
				(0.0369)	(0.0791)	(0.0561)	(0.0216)	(0.0691)	(0.0447)
9	Fréchet	0.5	-1	0.2022	0.4278	0.1212	0.3677	0.0333	0.3581
				(0.0228)	(0.1216)	(0.0230)	(0.0374)	(0.0054)	(0.2208)
10	*t*	0.5	-1	0.1544	0.4698	0.1161	0.0497	0.3285	0.1388
				(0.0178)	(0.0854)	(0.0133)	(0.0043)	(0.0192)	(0.0172)
11	logΓ	0.5	0	0.4174	0.9659	0.7150	0.4391	1.0935	0.7463
				(0.0428)	(0.1330)	(0.0621)	(0.0609)	(0.0469)	(0.0638)
γ **= 0.5**				**0.2326**	**0.5544**	**0.3211**	**0.4200**	**1.0680**	**0.3390**
				(0.0112)	(0.0480)	(0.0151)	(0.0135)	(0.0928)	(0.0395)
12	Burr	1	-2	1.1417	1.6517	0.7450	1.1510	2.3404	1.1582
				(0.1309)	(0.1908)	(0.1534)	(0.1828)	(1.4622)	(0.5701)
13	Burr	1	-1	0.7985	0.9219	0.4289	1.4920	0.3622	0.3730
				(0.0854)	(0.1234)	(0.0589)	(0.0931)	(0.0364)	(0.0497)
14	Burr	1	-0.5	1.0093	1.3410	2.1486	3.0402	14.6282	3.0442
				(0.1288)	(0.1629)	(0.2232)	(0.0828)	(0.2618)	(0.2016)
15	Fréchet	1	-1	0.9403	1.3442	0.5606	1.3987	0.1170	0.4617
				(0.1029)	(0.2400)	(0.0980)	(0.1436)	(0.0171)	(0.0749)
16	*t*	1	-2	0.9407	1.8300	0.3854	0.9108	1.3748	0.3369
				(0.1165)	(0.4342)	(0.0644)	(0.1724)	(1.1646)	(0.0528)
17	logΓ	1	0	1.4298	2.9033	2.4687	1.7103	3.8572	2.6190
				(0.1868)	(0.4156)	(0.2307)	(0.2452)	(0.1564)	(0.2348)
γ **= 1**				**1.0434**	**1.6653**	**1.1229**	**1.6172**	**3.7800**	**1.3322**
				(0.0527)	(0.1175)	(0.0632)	(0.0665)	(0.3157)	(0.1095)
Mean				**0.4654**	**0.7932**	**0.5207**	**0.7761**	**1.8472**	**0.7905**
				(0.0192)	(0.0448)	(0.0258)	(0.0769)	(0.1163)	(0.0548)

Table 4.2.8 Results for $n = 5000$ for other estimators based on the PPD.

				Bayes			MLE		
Distr.		γ	ρ	**Est**	**-1**	**Gomes**	**Est**	**-1**	**Gomes**
1	Burr	0.1	-2	0.0071	0.0161	0.0027	0.0053	0.0013	0.0405
				(0.0010)	(0.0088)	(0.0004)	(0.0010)	(0.0001)	(0.0383)
2	Burr	0.1	-1	0.0076	0.0026	0.0029	0.0139	0.0117	0.0021
				(0.0007)	(0.0008)	(0.0004)	(0.0009)	(0.0005)	(0.0003)
3	Burr	0.1	-0.5	0.0076	0.0324	0.0106	0.0198	0.3388	0.0444
				(0.0010)	(0.0015)	(0.0011)	(0.0007)	(0.0032)	(0.0015)
4	Fréchet	0.1	-1	0.0073	0.0061	0.0123	0.0095	0.0009	0.0463
				(0.0010)	(0.0036)	(0.0065)	(0.0012)	(0.0001)	(0.0309)
5	logΓ	0.1	0	0.0139	0.0298	0.0220	0.0116	0.9612	1.4122
				(0.0018)	(0.0019)	(0.0020)	(0.0018)	(0.1376)	(0.2172)
γ = 0.1				**0.0087**	**0.0174**	**0.0101**	**0.0120**	**0.2628**	**0.3091**
				(0.0005)	(0.0020)	(0.0014)	(0.0005)	(0.0275)	(0.0445)
6	Burr	0.5	-2	0.1415	0.0440	0.0767	0.1984	0.2858	0.0508
				(0.0205)	(0.0050)	(0.0122)	(0.0306)	(0.2599)	(0.0092)
7	Burr	0.5	-1	0.1548	0.0373	0.0626	0.3055	0.0424	0.0497
				(0.0184)	(0.0050)	(0.0095)	(0.0312)	(0.0045)	(0.0072)
8	Burr	0.5	-0.5	0.1924	1.2175	0.2451	0.0714	2.6465	0.3965
				(0.0247)	(0.0349)	(0.0249)	(0.0061)	(0.0450)	(0.0272)
9	Fréchet	0.5	-1	0.1349	0.0405	0.0720	0.2359	0.0256	0.0570
				(0.0160)	(0.0056)	(0.0102)	(0.0278)	(0.0035)	(0.0066)
10	*t*	0.5	-1	0.1635	0.0509	0.0809	0.3877	0.1495	0.0574
				(0.0189)	(0.0070)	(0.0097)	(0.0212)	(0.0093)	(0.0081)
11	logΓ	0.5	0	0.3711	0.7685	0.5819	0.2999	0.9377	0.6038
				(0.0518)	(0.0445)	(0.0511)	(0.0456)	(0.0376)	(0.0525)
γ = 0.5				**0.1930**	**0.3598**	**0.1865**	**0.2498**	**0.6813**	**0.2025**
				(0.0114)	(0.0096)	(0.0101)	(0.0121)	(0.0444)	(0.0102)
12	Burr	1	-2	0.5650	0.3122	0.2725	0.7385	2.2157	0.2458
				(0.0688)	(0.0395)	(0.0368)	(0.1134)	(1.4750)	(0.0493)
13	Burr	1	-1	0.5207	0.1397	0.2354	1.4853	0.1664	1.4647
				(0.0654)	(0.0196)	(0.0352)	(0.1202)	(0.0192)	(1.0793)
14	Burr	1	-0.5	1.2134	3.2001	1.1292	0.6867	10.2394	0.8954
				(0.1112)	(0.1396)	(0.1045)	(0.0325)	(0.1605)	(0.0993)
15	Fréchet	1	-1	0.5843	1.7092	0.3352	0.9014	0.1057	0.4579
				(0.0719)	(1.0300)	(0.0638)	(0.1032)	(0.0164)	(0.0820)
16	*t*	1	-2	0.5196	0.3148	0.2300	0.6522	0.1098	0.2431
				(0.0590)	(0.0414)	(0.0395)	(0.1213)	(0.0163)	(0.0517)
17	logΓ	1	0	1.0981	2.8410	2.1824	1.4325	3.6755	2.2929
				(0.1220)	(0.1940)	(0.2171)	(0.2101)	(0.1288)	(0.2276)
γ = 1				**0.7502**	**1.4195**	**0.7308**	**0.9828**	**2.7521**	**0.9333**
				(0.0353)	(0.1765)	(0.0429)	(0.0521)	(0.2483)	(0.1855)
Mean				**0.3423**	**0.6477**	**0.3267**	**0.4496**	**1.3565**	**0.5108**
				(0.0134)	(0.0687)	(0.0156)	(0.0190)	(0.0891)	(0.0427)

Table 4.2.9 Results for $n = 10\,000$ for other estimators based on the PPD.

				Bayes			MLE		
Distr.		γ	ρ	**Est**	**-1**	**Gomes**	**Est**	**-1**	**Gomes**
1	Burr	0.1	-2	0.0045	0.0014	0.0016	0.0046	0.0014	0.0010
				(0.0006)	(0.0002)	(0.0003)	(0.0008)	(0.0001)	(0.0002)
2	Burr	0.1	-1	0.0041	0.0011	0.0023	0.0110	0.0122	0.0012
				(0.0005)	(0.0002)	(0.0003)	(0.0010)	(0.0002)	(0.0002)
3	Burr	0.1	-0.5	0.0130	0.0343	0.0111	0.0104	0.3317	0.0305
				(0.0009)	(0.0010)	(0.0054)	(0.0004)	(0.0028)	(0.0010)
4	Fréchet	0.1	-1	0.0043	0.0010	0.0019	0.0082	0.0008	0.0694
				(0.0006)	(0.0001)	(0.0003)	(0.0010)	(0.0001)	(0.0679)
5	logΓ	0.1	0	0.0111	0.0264	0.0227	0.0059	0.9482	0.9115
				(0.0011)	(0.0012)	(0.0047)	(0.0012)	(0.1349)	(0.1494)
γ = 0.1				**0.0074**	**0.0128**	**0.0079**	**0.0080**	**0.2588**	**0.2027**
				(0.0003)	(0.0003)	(0.0014)	(0.0004)	(0.0270)	(0.0328)
6	Burr	0.5	-2	0.1137	0.0433	0.0274	0.1367	1.0115	0.0205
				(0.0140)	(0.0069)	(0.0038)	(0.0221)	(0.4900)	(0.0033)
7	Burr	0.5	-1	0.1455	0.0226	0.0412	0.3656	0.0246	0.0303
				(0.0207)	(0.0033)	(0.0061)	(0.0262)	(0.0029)	(0.0044)
8	Burr	0.5	-0.5	0.1268	0.7999	0.1479	0.1964	1.8871	0.2444
				(0.0165)	(0.0252)	(0.0161)	(0.0066)	(0.0309)	(0.0172)
9	Fréchet	0.5	-1	0.1472	0.0248	0.0417	0.2349	0.0144	0.0340
				(0.0233)	(0.0035)	(0.0059)	(0.0285)	(0.0015)	(0.0051)
10	*t*	0.5	-1	0.1203	0.0314	0.0597	0.3345	0.0784	0.0339
				(0.0170)	(0.0038)	(0.0081)	(0.0108)	(0.0054)	(0.0045)
11	logΓ	0.5	0	0.2618	0.6780	0.5035	0.1986	0.8614	0.5000
				(0.0325)	(0.0339)	(0.0433)	(0.0406)	(0.0242)	(0.0386)
γ = 0.5				**0.1526**	**0.2667**	**0.1369**	**0.2444**	**0.6462**	**0.1438**
				(0.0088)	(0.0072)	(0.0080)	(0.0103)	(0.0819)	(0.0072)
12	Burr	1	-2	0.5009	0.9359	0.1280	0.4523	2.0677	0.0933
				(0.0777)	(0.7260)	(0.0315)	(0.0855)	(1.4170)	(0.0278)
13	Burr	1	-1	0.4375	0.1439	0.1246	1.2562	0.1479	0.1230
				(0.0533)	(0.0209)	(0.0189)	(0.1106)	(0.0125)	(0.0177)
14	Burr	1	-0.5	0.4014	1.7781	0.8860	0.7704	10.6803	0.8960
				(0.0533)	(0.0937)	(0.0665)	(0.0285)	(0.1179)	(0.0662)
15	Fréchet	1	-1	0.5859	0.2341	0.1927	1.0258	0.0457	1.1094
				(0.0779)	(0.0381)	(0.0317)	(0.1208)	(0.0065)	(0.9299)
16	*t*	1	-2	0.4153	0.1936	0.1004	0.3168	0.0642	0.0808
				(0.0550)	(0.0259)	(0.0139)	(0.0816)	(0.0099)	(0.0115)
17	logΓ	1	0	1.1148	2.5108	2.0497	1.0052	3.5489	2.1254
				(0.1210)	(0.1253)	(0.1338)	(0.1695)	(0.0823)	(0.1344)
γ = 1				**0.5759**	**0.9661**	**0.5802**	**0.8044**	**2.7591**	**0.7380**
				(0.0314)	(0.1241)	(0.0263)	(0.0442)	(0.2374)	(0.1571)
Mean				**0.2670**	**0.4721**	**0.2581**	**0.3726**	**1.3946**	**0.3853**
				(0.0123)	(0.0091)	(0.0096)	(0.0160)	(0.0881)	(0.0567)

Table 4.2.10 Results for $n = 20\,000$ for other estimators based on the PPD.

From Tables 4.2.2 to 4.2.10 we can see that the "Bayes Est" method (estimating all three parameters of the PPD, using Bayes) performs best, overall and for all EVI groups, except for samples of sizes $n = 10\ 000$ and $n = 20\ 000$, where it seems to be marginally better to first estimate ρ using Gomes and Martins (2001), and then determining the Bayesian estimates of the two remaining parameters. The reason for this might be that the limiting assumptions required for the estimator by Gomes and Martins (2001) are only satisfied for very large samples. We will, however, not explore this idea further.

We will now proceed to estimators which do not involve the estimation of the PPD parameters.

4.3 The Hill estimator

The most famous estimator of the EVI, the Hill estimator, was defined in Section 3.1.3 as follows:

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent, identically distributed random variables with common Pareto type distribution function F concentrated on $[0, \infty)$. Let k $(2 \leq k \leq n-1)$ denote the number of excesses, and let $Z_i = X_{n-i+1,n}/X_{n-k,n}$, $i = 1,2,\ldots,k$, be the multiplicative excesses.

The Hill estimator of the (positive) EVI, denoted by $H_{k,n}$, is defined as

$$H_{k,n} = \frac{1}{k}\sum_{i=1}^{k} \log Z_i.$$

Since the Hill estimator has been studied extensively, several methods of threshold selection have been proposed in the literature. We consider three of those methods in this thesis. For these methods we only state the procedure used to obtain a threshold. For detail on the derivation of these methods and explanations of how they work, see the references.

4.3.1 Guillou and Hall's method for selecting an optimal threshold

Guillou and Hall (2001) developed a method of selecting an optimal threshold for the Hill estimator. For an ordered set of observations $x_{1,n} \leq x_{2,n} \leq \ldots \leq x_{n,n}$ and a given value of k, calculate the following:

1. $H_{k,n}$ (the Hill estimate as defined above)
2. $y_i = i\log(x_{n-i+1,n}/x_{n-i,n})$ for $i = 1,2,..,k$
3. $w_i = k - 2i + 1$ for $i = 1,2,..,k$
4. $u_k = \sum_{i=1}^{k} w_i y_i$
5. $t_k = \sqrt{\frac{3}{k^3}}\frac{u_k}{H_{k,n}}$

6. $m = \lfloor k/2 \rfloor$ (the integer part of $k/2$)
7. $q_k = \sqrt{\frac{1}{2m+1}\sum_{j=-m}^{m} t_{k+j}^2}$

We define the optimal choice k_{opt} as the least integer k such that $q_r > c_{crit}$ for all $r > k$. The standard choices of the critical value c_{crit} include 1.25 and 1.5. In this thesis we will examine both these choices.

In some instances $q_k \leq c_{crit}$ for all $k = 2,3,\ldots,k_{\max}$, in which case we choose $k = k_{\max}$, where $k_{\max}$ is the maximum value of k for which $\sum_{j=-m}^{m} t_{k+j}^2$ can be calculated. (From $k + j < n$ we see that $k_{\max} < 2n/3$.) If $q_k > c_{crit}$ for all $k = 2,3,\ldots,k_{\max}$, the procedure fails.

As far as the simulation study was concerned, the procedure never failed for c_{crit} = 1.5. There were only two samples (of size $n = 50$) for which c_{crit} = 1.25 failed, out of all the thousands of samples. In these two cases we used c_{crit} = 1.5 instead of c_{crit} = 1.25.

The results of the simulation study are shown in Section 4.3.4.

4.3.2 Selecting an optimal threshold by minimizing the AMSE

Beirlant et al. (2004) discuss a technique to determine an optimal threshold by estimating the asymptotic mean square error (AMSE) of the Hill estimator at each value of k and choosing as an optimal value of k that value which minimizes this estimated AMSE.

For $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ and a given value of k, the AMSE of the Hill estimator $H_{k,n}$ is the sum of the asymptotic variance and the square of the asymptotic bias:

$$\mathrm{AMSE}(H_{k,n}) = \mathrm{AVar}(H_{k,n}) + \ \mathrm{ABias}^2(H_{k,n}) = \frac{\gamma^2}{k} + \left(\frac{b_{n,k}}{1-\rho}\right)^2.$$

Here $\gamma > 0$ is the EVI, $\rho < 0$ the second order parameter, and $-\infty < b_{n,k} < \infty$ (Beirlant et al., 2004).

In order to reduce computation time, ρ is chosen at a fixed value. The value recommended by Beirlant et al. (2004), is $\rho = -1$.

The two remaining parameters γ and $b_{n,k}$ are estimated by using either MLE or least squares estimation. According to Beirlant et al. (2004), the resulting estimates of these two estimation techniques are very similar. Least squares estimation will be applied here.

For a given value of k, we need to minimize $\sum_{i=1}^{k}\left(y_i-\gamma-b_{n,k}\left(\frac{i}{k+1}\right)\right)^2$ with respect to γ and $b_{n,k}$, where $y_i = i\log(x_{n-i+1,n}/x_{n-i,n})$ (Beirlant et al., 2004).

Taking partial derivatives with respect to γ and $b_{n,k}$, equating the resulting expressions to 0 and solving for $\hat{\gamma}$ and $\hat{b}_{n,k}$, we obtain estimates $\hat{b}_{n,k} = \frac{6\left(2\sum_{i=1}^{k} iy_i-k(k+1)\overline{y}\right)}{k(k-1)}$ and $\hat{\gamma} = \overline{y} - \frac{\hat{b}_{n,k}}{2}$, where $\overline{y} = \frac{1}{k}\sum_{i=1}^{k} y_i = H_{k,n}$.

Beirlant et al. (2004:115) also give advice on some computational aspects. To avoid instabilities arising from the optimization procedure, they apply a smoothness condition by linking estimates at subsequent values of k. The restriction $|\hat{b}_{n,k}| \le 1.1|\hat{b}_{n,k+1}|$ is applied.

In summary, the procedure therefore involves letting k assume all values from $n-1$, $n-2$ down to 2, and doing the following at each value of k:

1. Calculate $y_i = i\log(x_{n-i+1,n}/x_{n-i,n})$ for $i = 1,2,\ldots,k$.
2. Calculate $\hat{b}_{n,k}$ and $\hat{\gamma}$ as above. If $\hat{\gamma} < 0$, set $\hat{\gamma} = 0$ and $\hat{b}_{n,k} = 2\overline{y}$.
3. If $|\hat{b}_{n,k}| > 1.1|\hat{b}_{n,k+1}|$, set $\hat{b}_{n,k} = 1.1\text{sign}(\hat{b}_{n,k})|\hat{b}_{n,k+1}|$. (There is no restriction on the first value, namely $\hat{b}_{n,n-1}$.)
4. Calculate $\widehat{AMSE(H_{k,n})} = \frac{\hat{\gamma}^2}{k} + \left(\frac{\hat{b}_{n,k}}{2}\right)^2$.

The optimal threshold is the value of k for which $\widehat{AMSE(H_{k,n})}$ is a minimum.

Since the AMSE can be estimated at all values of k, there are no instances where this procedure does not provide an optimal threshold. The results of the simulation study are shown in Section 4.3.4.

4.3.3 Drees and Kaufmann's method for selecting an optimal threshold

Drees and Kaufmann (1998) developed a method of selecting an optimal threshold for the Hill estimator. For n observations, the procedure is the following:

1. Obtain an initial estimate of the EVI, by calculating $\hat{\gamma}_0 = H_{2\sqrt{n},n}$.
2. For $r_n = 2.5\hat{\gamma}_0 n^{0.25}$, compute $\hat{k}_n(r_n)$, which is the minimum value of k for which there is an i such that $|\sqrt{i}(H_{i,n} - H_{k,n})| > r_n$. Here $k = 1, 2, \ldots, (n-1)$ and $i = 1, 2, \ldots, k$.
3. Similarly, compute $\hat{k}_n(r_n^{0.7})$.
4. The optimal choice of k is given as $\hat{k}_{opt} = \frac{1}{3}\left[\frac{\hat{k}_n(r_n^{0.7})}{\left(\hat{k}_n(r_n)\right)^{0.7}}\right]^{\frac{10}{3}}(2\hat{\gamma}_0)^{\frac{1}{3}}$.

The procedure fails if $\hat{k}_n(r_n)$ or $\hat{k}_n(r_n^{0.7})$ does not exist, or if $\hat{k}_{opt} < 2$, or if $\hat{k}_{opt} > n-1$.

The number of times the method by Drees and Kaufmann fails, is much more substantial than that of the method by Guillou and Hall of Section 4.3.1. Table 4.3.3 below shows the percentage of samples for which the method by Drees and Kaufmann failed. For example, the method failed for 39.7% of the $17\ 000$ samples of size $n = 50$. (There were 1000 samples from each of 17 distributions.)

Sample size	50	100	200	500	1000	2000	>2000
Percentage failure	39.7%	28.4%	20.7%	17%	15.3%	8.6%	0%

Table 4.3.3 Percentage of samples for which method by Drees and Kaufmann failed.

To be able to compare the results of this technique of threshold selection to the previous two techniques, one has to be able to calculate a threshold for all samples. The way in which we manage this, is to use two approaches of optimal threshold calculation for the method by Drees and Kaufmann. The first is to calculate the optimal threshold by using Drees and Kaufmann, and to use the method by Guillou and Hall with c_{crit} = 1.25 to calculate the optimal threshold for the instances where the method by Drees and Kaufmann fails. (We use c_{crit} = 1.25, since this choice of c_{crit} yielded better results than c_{crit} = 1.5, as can be seen in Section 4.3.4.) The second is to calculate the optimal threshold by using Drees and Kaufmann, and to use the method of minimizing the AMSE (of Section 4.3.2) to calculate the optimal threshold for the instances where the method by Drees and Kaufmann fails.

The simulation study results of both these approaches are shown in Section 4.3.4.

4.3.4 Simulation study results

Tables 4.3.4.1 to 4.3.4.9 below show the simulation study results obtained for the Hill estimator, by applying the threshold selection techniques discussed in the previous three sections. (The results obtained for the Hill estimator without the application of a threshold selection procedure, are shown in Section 4.7.) We compare the performance of these techniques to that of the benchmark PPD estimator, as shown in Tables 3.4.4.2 to 3.4.4.6.

The tables below are in the same format as that of Tables 3.4.4.2 to 3.4.4.6, with the following last six columns:

PPD The benchmark PPD estimator.

G&H 1.25 Method by Guillou and Hall with c_{crit} = 1.25.

G&H 1.5 Method by Guillou and Hall with c_{crit} = 1.5.

AMSE Method of minimizing the AMSE.

D&K G&H Method by Drees and Kaufmann. If the method failed, the method by Guillou and Hall with $c_{crit} = 1.25$ was used.

D&K AMSE Method by Drees and Kaufmann. If the method failed, the method of minimizing the AMSE was used.

As pointed out in Section 4.3.3, the method by Drees and Kaufmann never failed for samples of sizes $n > 2000$. For this reason the final three tables (Tables 4.3.4.7 to 4.3.4.9) only have four columns (instead of five) pertaining to the Hill estimator. In those tables the last two columns (*D&K G&H* and *D&K AMSE*) are replaced by one column, namely:

D&K: Method by Drees and Kaufmann.

Only the results pertaining to these threshold selection techniques are presented below. Due to lack of space the results of the Hill estimator obtained without applying a threshold selection technique, will be presented in Section 4.7.

					G&H		AMSE	D&K	
Distr.		γ	ρ	**PPD**	**1.25**	**1.5**		**G&H**	**AMSE**
1	Burr	0.1	-2	0.0677	0.1001	0.0827	0.1440	0.1016	0.1140
				(0.0023)	(0.0070)	(0.0045)	(0.0068)	(0.0067)	(0.0052)
2	Burr	0.1	-1	0.0411	0.1974	0.2198	0.1986	0.1389	0.1347
				(0.0016)	(0.0119)	(0.0082)	(0.0074)	(0.0070)	(0.0067)
3	Burr	0.1	-0.5	0.1812	0.6702	0.8473	0.4249	0.4658	0.4581
				(0.0064)	(0.0229)	(0.0268)	(0.0152)	(0.0194)	(0.0189)
4	Fréchet	0.1	-1	0.0491	0.1034	0.1103	0.1591	0.1045	0.1490
				(0.0018)	(0.0047)	(0.0044)	(0.0069)	(0.0047)	(0.0065)
5	logΓ	0.1	0	0.0406	0.2418	0.2295	0.2214	0.2417	0.2213
				(0.0019)	(0.0208)	(0.0074)	(0.0070)	(0.0208)	(0.0070)
γ = 0.1				**0.0759**	**0.2626**	**0.2979**	**0.2296**	**0.2105**	**0.2154**
				(0.0015)	(0.0068)	(0.0059)	(0.0041)	(0.0061)	(0.0046)
6	Burr	0.5	-2	1.8210	2.3289	1.9375	3.4698	2.0470	2.1975
				(0.0547)	(0.1601)	(0.0900)	(0.1677)	(0.1460)	(0.1435)
7	Burr	0.5	-1	0.9261	4.6944	5.5002	5.0457	3.5288	3.4760
				(0.0390)	(0.1738)	(0.1960)	(0.1794)	(0.1469)	(0.1429)
8	Burr	0.5	-0.5	3.8526	16.8610	20.5762	11.4342	13.4500	13.2335
				(0.1432)	(0.5948)	(0.6434)	(0.4236)	(0.5082)	(0.4904)
9	Fréchet	0.5	-1	1.3757	2.8913	2.7421	3.6831	2.8052	3.0796
				(0.0469)	(0.1393)	(0.1126)	(0.1629)	(0.1345)	(0.1400)
10	*t*	0.5	-1	0.8748	6.8244	8.2975	5.9781	4.8233	4.8072
				(0.0428)	(0.2945)	(0.2969)	(0.2034)	(0.2059)	(0.2059)
11	logΓ	0.5	0	0.9493	6.1096	6.3251	5.7691	6.1189	5.7374
				(0.0437)	(0.2351)	(0.2083)	(0.1887)	(0.2351)	(0.1877)
γ = 0.5				**1.6332**	**6.6183**	**7.5631**	**5.8967**	**5.4622**	**5.4218**
				(0.0293)	(0.1259)	(0.1296)	(0.0976)	(0.1076)	(0.1026)
12	Burr	1	-2	8.1259	8.0540	7.4595	13.2209	7.6748	9.0173
				(0.2331)	(0.3891)	(0.3539)	(0.6457)	(0.3432)	(0.4650)
13	Burr	1	-1	3.8879	19.4536	21.5813	18.7033	14.7567	14.3458
				(0.1571)	(0.8505)	(0.8478)	(0.7363)	(0.6893)	(0.6512)
14	Burr	1	-0.5	12.3702	66.9450	87.1917	40.5876	57.2861	53.0072
				(0.5180)	(4.1188)	(3.2842)	(1.4455)	(4.0797)	(1.9902)
15	Fréchet	1	-1	5.9964	12.0935	10.7673	14.9394	12.2096	12.5887
				(0.2113)	(0.8119)	(0.4043)	(0.6233)	(0.8040)	(0.5274)
16	*t*	1	-2	7.2427	14.0227	11.1011	15.3821	10.4309	8.5804
				(0.2185)	(1.3754)	(0.5766)	(0.6676)	(0.9555)	(0.3787)
17	logΓ	1	0	3.4014	22.8151	22.9917	22.2403	22.8579	22.1045
				(0.1525)	(0.9330)	(0.7466)	(0.7532)	(0.9349)	(0.7516)
γ = 1				**6.8374**	**23.8973**	**26.8488**	**20.8456**	**20.8693**	**19.9407**
				(0.1134)	(0.7685)	(0.5936)	(0.3516)	(0.7392)	(0.3940)
Mean				**3.0120**	**10.8474**	**12.2330**	**9.5060**	**9.3554**	**9.0148**
				(0.0413)	(0.2749)	(0.2144)	(0.1288)	(0.2636)	(0.1437)

Table 4.3.4.1 Results for $n = 50$ for the Hill estimator.

					G&H		AMSE	D&K	
Distr.		γ	ρ	**PPD**	**1.25**	**1.5**		**G&H**	**AMSE**
1	Burr	0.1	-2	0.0521	0.0587	0.0483	0.0992	0.0524	0.0509
				(0.0024)	(0.0066)	(0.0040)	(0.0086)	(0.0042)	(0.0040)
2	Burr	0.1	-1	0.0274	0.1108	0.1231	0.1427	0.0745	0.0743
				(0.0017)	(0.0063)	(0.0063)	(0.0091)	(0.0047)	(0.0047)
3	Burr	0.1	-0.5	0.1177	0.4096	0.5014	0.3142	0.3068	0.3068
				(0.0063)	(0.0192)	(0.0217)	(0.0151)	(0.0163)	(0.0163)
4	Fréchet	0.1	-1	0.0350	0.0866	0.0799	0.1050	0.0691	0.0733
				(0.0019)	(0.0123)	(0.0102)	(0.0085)	(0.0074)	(0.0060)
5	logΓ	0.1	0	0.0325	0.1692	0.1866	0.1827	0.1692	0.1826
				(0.0024)	(0.0080)	(0.0077)	(0.0092)	(0.0080)	(0.0092)
γ **= 0.1**				**0.0529**	**0.1670**	**0.1879**	**0.1688**	**0.1344**	**0.1376**
				(0.0015)	(0.0052)	(0.0052)	(0.0047)	(0.0041)	(0.0041)
6	Burr	0.5	-2	1.4283	1.7070	1.4170	2.5154	1.2423	1.2487
				(0.0681)	(0.1546)	(0.1009)	(0.2188)	(0.0794)	(0.0877)
7	Burr	0.5	-1	0.7448	2.9235	3.1919	3.6513	1.9024	1.9024
				(0.0411)	(0.2220)	(0.1817)	(0.2375)	(0.1207)	(0.1207)
8	Burr	0.5	-0.5	2.3898	9.6180	12.1339	7.6414	7.3940	7.3940
				(0.1262)	(0.4533)	(0.5387)	(0.3357)	(0.3747)	(0.3747)
9	Fréchet	0.5	-1	0.9574	2.1252	1.9810	2.8455	1.6799	1.8770
				(0.0492)	(0.2143)	(0.1222)	(0.2127)	(0.1065)	(0.1464)
10	*t*	0.5	-1	0.5890	4.0332	4.7462	4.6105	2.8419	2.8419
				(0.0391)	(0.2285)	(0.2408)	(0.2496)	(0.1680)	(0.1680)
11	logΓ	0.5	0	0.6741	4.3481	4.6239	4.4629	4.3428	4.4422
				(0.0458)	(0.2135)	(0.1716)	(0.2179)	(0.2136)	(0.2179)
γ **= 0.5**				**1.1306**	**4.1258**	**4.6823**	**4.2878**	**3.2339**	**3.2844**
				(0.0280)	(0.1083)	(0.1100)	(0.1017)	(0.0827)	(0.0850)
12	Burr	1	-2	5.9039	7.2639	4.7759	9.8040	4.6396	4.5764
				(0.2499)	(1.2389)	(0.3360)	(0.8558)	(0.3378)	(0.3160)
13	Burr	1	-1	2.9743	11.9673	13.7786	16.3911	9.4207	9.4149
				(0.1608)	(0.6291)	(0.6903)	(0.9700)	(0.6157)	(0.6159)
14	Burr	1	-0.5	8.1449	38.9480	50.1057	31.5774	31.8209	31.8174
				(0.4228)	(1.8365)	(2.1709)	(1.3938)	(1.5986)	(1.5987)
15	Fréchet	1	-1	4.3039	6.9040	6.6428	12.2253	6.5115	7.8048
				(0.2047)	(0.6082)	(0.3464)	(0.9916)	(0.5786)	(0.6775)
16	*t*	1	-2	6.0318	6.7276	5.9297	11.6052	4.7522	4.7542
				(0.2482)	(0.5927)	(0.3960)	(0.9125)	(0.2910)	(0.2911)
17	logΓ	1	0	2.4120	18.0256	19.3161	18.4860	18.0748	18.1982
				(0.1645)	(0.7425)	(0.7244)	(0.9015)	(0.7416)	(0.8882)
γ **= 1**				**4.9618**	**14.9727**	**16.7581**	**16.6815**	**12.5366**	**12.7610**
				(0.1051)	(0.4274)	(0.4118)	(0.4165)	(0.3341)	(0.3483)
Mean				**2.1658**	**6.7898**	**7.6225**	**7.4506**	**5.6056**	**5.7035**
				(0.0384)	(0.1556)	(0.1504)	(0.1513)	(0.1215)	(0.1265)

Table 4.3.4.2 Results for $n = 100$ for the Hill estimator.

					G&H		AMSE	D&K	
Distr.		γ	ρ	**PPD**	**1.25**	**1.5**		**G&H**	**AMSE**
1	Burr	0.1	-2	0.0405	0.0366	0.0318	0.0606	0.0324	0.0325
				(0.0022)	(0.0037)	(0.0029)	(0.0068)	(0.0021)	(0.0021)
2	Burr	0.1	-1	0.0212	0.0963	0.0879	0.1047	0.0424	0.0424
				(0.0011)	(0.0145)	(0.0059)	(0.0082)	(0.0028)	(0.0028)
3	Burr	0.1	-0.5	0.0833	0.2764	0.3437	0.2677	0.2146	0.2146
				(0.0041)	(0.0132)	(0.0148)	(0.0123)	(0.0104)	(0.0104)
4	Fréchet	0.1	-1	0.0259	0.0467	0.0484	0.0743	0.0315	0.0306
				(0.0014)	(0.0032)	(0.0030)	(0.0078)	(0.0022)	(0.0029)
5	logΓ	0.1	0	0.0286	0.1486	0.1650	0.1555	0.1486	0.1538
				(0.0022)	(0.0055)	(0.0056)	(0.0078)	(0.0055)	(0.0077)
γ **= 0.1**				**0.0399**	**0.1209**	**0.1354**	**0.1326**	**0.0939**	**0.0948**
				(0.0011)	(0.0042)	(0.0035)	(0.0039)	(0.0025)	(0.0027)
6	Burr	0.5	-2	0.9399	0.7078	0.6143	1.2358	0.5199	0.5390
				(0.0437)	(0.0543)	(0.0365)	(0.1546)	(0.0319)	(0.0389)
7	Burr	0.5	-1	0.5805	2.0151	2.2048	2.6794	1.2812	1.2812
				(0.0329)	(0.1176)	(0.1119)	(0.2029)	(0.0809)	(0.0809)
8	Burr	0.5	-0.5	1.6694	6.7391	7.9853	5.9276	5.2570	5.2570
				(0.0852)	(0.3427)	(0.3322)	(0.2583)	(0.2465)	(0.2465)
9	Fréchet	0.5	-1	0.7253	1.1640	1.1107	1.8383	0.8728	0.8271
				(0.0380)	(0.0712)	(0.0561)	(0.1858)	(0.0519)	(0.0565)
10	*t*	0.5	-1	0.4038	2.5765	2.8295	3.4243	1.5129	1.5129
				(0.0267)	(0.1715)	(0.1445)	(0.2268)	(0.0855)	(0.0855)
11	logΓ	0.5	0	0.5613	3.4527	3.7934	3.4448	3.4532	3.3971
				(0.0482)	(0.1502)	(0.1400)	(0.1876)	(0.1502)	(0.1838)
γ **= 0.5**				**0.8134**	**2.7759**	**3.0897**	**3.0917**	**2.1495**	**2.1357**
				(0.0202)	(0.0729)	(0.0683)	(0.0838)	(0.0529)	(0.0561)
12	Burr	1	-2	4.3406	3.4572	3.0072	5.9196	2.2121	2.2201
				(0.1824)	(0.5000)	(0.3238)	(0.7187)	(0.1635)	(0.1628)
13	Burr	1	-1	2.3573	7.5437	7.4945	10.1546	5.4943	5.2706
				(0.1260)	(0.5947)	(0.3983)	(0.8081)	(0.3989)	(0.3244)
14	Burr	1	-0.5	5.9909	24.3611	30.4142	23.3378	20.3640	20.3640
				(0.3539)	(1.2375)	(1.3799)	(1.1232)	(1.0770)	(1.0770)
15	Fréchet	1	-1	2.9867	4.7245	4.7403	7.0122	3.8100	4.1583
				(0.1754)	(0.3543)	(0.3317)	(0.7247)	(0.3108)	(0.4084)
16	*t*	1	-2	4.1087	5.7724	3.5962	6.2146	3.7796	2.5170
				(0.1772)	(1.3511)	(0.2530)	(0.6961)	(1.2753)	(0.1786)
17	logΓ	1	0	1.8962	13.2999	14.3966	13.5175	13.2822	13.5005
				(0.1386)	(0.6598)	(0.4904)	(0.7809)	(0.6606)	(0.7817)
γ **= 1**				**3.6134**	**9.8598**	**10.6082**	**11.0261**	**8.1570**	**8.0051**
				(0.0843)	(0.3544)	(0.2678)	(0.3354)	(0.3120)	(0.2416)
Mean				**1.5741**	**4.4952**	**4.8743**	**5.0217**	**3.6652**	**3.6070**
				(0.0306)	(0.1277)	(0.0976)	(0.1220)	(0.1117)	(0.0875)

Table 4.3.4.3 Results for $n = 200$ for the Hill estimator.

					G&H		AMSE	D&K	
Distr.		γ	ρ	**PPD**	**1.25**	**1.5**		**G&H**	**AMSE**
1	Burr	0.1	-2	0.0247	0.0151	0.0126	0.0235	0.0115	0.0115
				(0.0019)	(0.0020)	(0.0014)	(0.0067)	(0.0013)	(0.0013)
2	Burr	0.1	-1	0.0130	0.0439	0.0436	0.0619	0.0239	0.0239
				(0.0011)	(0.0052)	(0.0035)	(0.0111)	(0.0023)	(0.0023)
3	Burr	0.1	-0.5	0.0517	0.1530	0.1760	0.1579	0.1344	0.1344
				(0.0038)	(0.0125)	(0.0113)	(0.0139)	(0.0085)	(0.0085)
4	Fréchet	0.1	-1	0.0168	0.0304	0.0251	0.0376	0.0144	0.0144
				(0.0013)	(0.0044)	(0.0027)	(0.0085)	(0.0013)	(0.0013)
5	logΓ	0.1	0	0.0194	0.0984	0.1144	0.0852	0.0970	0.0832
				(0.0023)	(0.0063)	(0.0055)	(0.0060)	(0.0063)	(0.0060)
γ **= 0.1**				**0.0251**	**0.0682**	**0.0744**	**0.0732**	**0.0562**	**0.0535**
				(0.0010)	(0.0031)	(0.0027)	(0.0043)	(0.0022)	(0.0022)
6	Burr	0.5	-2	0.6359	0.3577	0.2788	0.6194	0.2674	0.2674
				(0.0476)	(0.0678)	(0.0272)	(0.1929)	(0.0260)	(0.0260)
7	Burr	0.5	-1	0.3950	0.9771	1.0437	1.7552	0.7082	0.7082
				(0.0392)	(0.1088)	(0.1059)	(0.3027)	(0.0614)	(0.0614)
8	Burr	0.5	-0.5	1.1387	4.1123	4.5609	3.7651	3.0381	3.0381
				(0.0834)	(0.5103)	(0.2514)	(0.3119)	(0.1854)	(0.1854)
9	Fréchet	0.5	-1	0.4377	0.6255	0.5680	0.8695	0.3454	0.3454
				(0.0333)	(0.0937)	(0.0478)	(0.2065)	(0.0298)	(0.0298)
10	*t*	0.5	-1	0.2722	1.2219	1.1602	1.9364	0.6819	0.6819
				(0.0252)	(0.1312)	(0.0891)	(0.3082)	(0.0602)	(0.0602)
11	logΓ	0.5	0	0.3691	2.5170	2.7797	2.7185	2.5309	2.7333
				(0.0364)	(0.1618)	(0.1229)	(0.2821)	(0.1613)	(0.2818)
γ **= 0.5**				**0.5414**	**1.6352**	**1.7319**	**1.9440**	**1.2620**	**1.2957**
				(0.0196)	(0.0956)	(0.0528)	(0.1110)	(0.0439)	(0.0584)
12	Burr	1	-2	2.5581	1.5446	1.1758	2.3138	0.9063	0.9063
				(0.1631)	(0.3374)	(0.1246)	(0.7219)	(0.0860)	(0.0860)
13	Burr	1	-1	1.5064	3.7820	4.4651	5.8672	2.8600	2.8600
				(0.1408)	(0.2952)	(0.3334)	(1.0722)	(0.2435)	(0.2435)
14	Burr	1	-0.5	4.0734	15.6007	18.3697	16.8343	12.1155	12.1155
				(0.2897)	(1.0059)	(1.0649)	(1.3519)	(0.8304)	(0.8304)
15	Fréchet	1	-1	1.6857	2.7274	2.6428	3.8166	1.9276	1.9267
				(0.1251)	(0.2935)	(0.2566)	(0.8769)	(0.2010)	(0.2010)
16	*t*	1	-2	2.6006	2.5448	2.3597	1.9166	1.4088	1.4088
				(0.1746)	(0.8850)	(0.7149)	(0.4733)	(0.1424)	(0.1424)
17	logΓ	1	0	1.5700	9.8628	11.0765	10.0801	9.8601	9.4009
				(0.1965)	(0.5485)	(0.5258)	(0.9245)	(0.5484)	(0.8385)
γ **= 1**				**2.3324**	**6.0104**	**6.6816**	**6.8048**	**4.8464**	**4.7697**
				(0.0773)	(0.2573)	(0.2423)	(0.3854)	(0.1762)	(0.2055)
Mean				**1.0217**	**2.7185**	**2.9913**	**3.1094**	**2.1724**	**2.1565**
				(0.0281)	(0.0969)	(0.0875)	(0.1415)	(0.0641)	(0.0754)

Table 4.3.4.4 Results for $n = 500$ for the Hill estimator.

					G&H		AMSE	D&K	
Distr.		γ	ρ	**PPD**	**1.25**	**1.5**		**G&H**	**AMSE**
1	Burr	0.1	-2	0.0167	0.0088	0.0068	0.0200	0.0056	0.0056
				(0.0011)	(0.0012)	(0.0006)	(0.0074)	(0.0005)	(0.0005)
2	Burr	0.1	-1	0.0098	0.0233	0.0227	0.0462	0.0127	0.0127
				(0.0009)	(0.0027)	(0.0018)	(0.0106)	(0.0011)	(0.0011)
3	Burr	0.1	-0.5	0.0372	0.0841	0.1120	0.1066	0.1011	0.1011
				(0.0035)	(0.0061)	(0.0071)	(0.0113)	(0.0054)	(0.0054)
4	Fréchet	0.1	-1	0.0133	0.0126	0.0135	0.0199	0.0078	0.0078
				(0.0010)	(0.0012)	(0.0011)	(0.0065)	(0.0007)	(0.0007)
5	logΓ	0.1	0	0.0167	0.0756	0.0898	0.0785	0.0762	0.0757
				(0.0017)	(0.0039)	(0.0041)	(0.0078)	(0.0039)	(0.0070)
γ = 0.1				**0.0187**	**0.0409**	**0.0490**	**0.0543**	**0.0407**	**0.0406**
				(0.0009)	(0.0016)	(0.0017)	(0.0040)	(0.0014)	(0.0018)
6	Burr	0.5	-2	0.3401	0.2109	0.1689	0.4264	0.1310	0.1310
				(0.0242)	(0.0282)	(0.0168)	(0.1603)	(0.0137)	(0.0137)
7	Burr	0.5	-1	0.1955	0.6159	0.6289	0.7741	0.3987	0.3987
				(0.0194)	(0.0637)	(0.0551)	(0.1561)	(0.0372)	(0.0372)
8	Burr	0.5	-0.5	0.9762	2.7089	3.1923	2.1754	2.1012	2.1012
				(0.0782)	(0.3737)	(0.3256)	(0.2106)	(0.1468)	(0.1468)
9	Fréchet	0.5	-1	0.2352	0.3488	0.3652	0.6894	0.1912	0.1912
				(0.0200)	(0.0361)	(0.0324)	(0.2119)	(0.0177)	(0.0177)
10	*t*	0.5	-1	0.2468	0.7166	0.7981	1.4041	0.5008	0.5008
				(0.0295)	(0.0700)	(0.0687)	(0.2893)	(0.0488)	(0.0488)
11	logΓ	0.5	0	0.5327	2.0119	2.2735	2.1890	2.0177	2.0182
				(0.0401)	(0.1004)	(0.0992)	(0.2251)	(0.0911)	(0.1851)
γ = 0.5				**0.4211**	**1.1022**	**1.2378**	**1.2764**	**0.8901**	**0.8902**
				(0.0166)	(0.0668)	(0.0589)	(0.0872)	(0.0308)	(0.0409)
12	Burr	1	-2	1.5654	0.9544	0.8439	0.9014	0.6857	0.6857
				(0.1246)	(0.1130)	(0.0939)	(0.1134)	(0.0727)	(0.0727)
13	Burr	1	-1	0.7953	2.3075	2.6395	2.7230	1.8996	1.8996
				(0.0747)	(0.1888)	(0.2006)	(0.5732)	(0.1796)	(0.1796)
14	Burr	1	-0.5	3.7100	9.0334	10.4975	11.9787	7.2513	7.2513
				(0.2776)	(0.6424)	(0.6562)	(1.3457)	(0.5310)	(0.5310)
15	Fréchet	1	-1	0.9030	1.8034	1.6676	2.7304	1.1058	1.1058
				(0.0820)	(0.2074)	(0.1642)	(0.7609)	(0.1033)	(0.1033)
16	*t*	1	-2	1.5004	1.9352	1.2337	3.2143	0.8483	0.8483
				(0.1215)	(0.5307)	(0.2358)	(0.9481)	(0.1372)	(0.1372)
17	logΓ	1	0	2.0135	7.9558	9.1027	8.5292	8.2529	8.4614
				(0.1671)	(0.3912)	(0.3936)	(0.7942)	(0.3767)	(0.7311)
γ = 1				**1.7479**	**3.9983**	**4.3308**	**5.0128**	**3.3406**	**3.3753**
				(0.0640)	(0.1615)	(0.1411)	(0.3440)	(0.1168)	(0.1567)
Mean				**0.7710**	**1.8122**	**1.9798**	**2.2357**	**1.5052**	**1.5174**
				(0.0234)	(0.0617)	(0.0540)	(0.1253)	(0.0426)	(0.0571)

Table 4.3.4.5 Results for $n = 1000$ for the Hill estimator.

					G&H		AMSE	D&K	
Distr.		γ	ρ	**PPD**	**1.25**	**1.5**		**G&H**	**AMSE**
1	Burr	0.1	-2	0.0109	0.0043	0.0030	0.0075	0.0028	0.0028
				(0.0008)	(0.0010)	(0.0002)	(0.0046)	(0.0003)	(0.0003)
2	Burr	0.1	-1	0.0085	0.0124	0.0141	0.0153	0.0095	0.0095
				(0.0008)	(0.0009)	(0.0010)	(0.0045)	(0.0008)	(0.0008)
3	Burr	0.1	-0.5	0.0263	0.0646	0.0669	0.0708	0.0758	0.0758
				(0.0025)	(0.0071)	(0.0042)	(0.0100)	(0.0040)	(0.0040)
4	Fréchet	0.1	-1	0.0078	0.0154	0.0077	0.0061	0.0044	0.0044
				(0.0006)	(0.0067)	(0.0007)	(0.0007)	(0.0004)	(0.0004)
5	logΓ	0.1	0	0.0173	0.0657	0.0743	0.0635	0.0714	0.0707
				(0.0014)	(0.0032)	(0.0030)	(0.0052)	(0.0026)	(0.0048)
γ = 0.1				**0.0141**	**0.0325**	**0.0332**	**0.0326**	**0.0328**	**0.0327**
				(0.0006)	(0.0021)	(0.0011)	(0.0026)	(0.0010)	(0.0013)
6	Burr	0.5	-2	0.2247	0.1217	0.0828	0.1793	0.0641	0.0641
				(0.0162)	(0.0259)	(0.0081)	(0.0996)	(0.0058)	(0.0058)
7	Burr	0.5	-1	0.1601	0.6888	0.3669	0.7350	0.2283	0.2283
				(0.0152)	(0.2617)	(0.0289)	(0.2157)	(0.0200)	(0.0200)
8	Burr	0.5	-0.5	0.7235	1.5201	1.8518	2.3048	1.4107	1.4107
				(0.0647)	(0.1089)	(0.1111)	(0.3076)	(0.0920)	(0.0920)
9	Fréchet	0.5	-1	0.1467	0.2879	0.2698	0.8636	0.1715	0.1715
				(0.0116)	(0.0327)	(0.0187)	(0.2614)	(0.0139)	(0.0139)
10	*t*	0.5	-1	0.1578	0.4440	0.4335	0.6178	0.2880	0.2880
				(0.0163)	(0.0621)	(0.0336)	(0.1785)	(0.0300)	(0.0300)
11	logΓ	0.5	0	0.3982	1.4505	1.7122	1.6289	1.7320	1.6300
				(0.0234)	(0.0631)	(0.0654)	(0.1830)	(0.0563)	(0.0546)
γ = 0.5				**0.3018**	**0.7522**	**0.7862**	**1.0549**	**0.6491**	**0.6321**
				(0.0125)	(0.0500)	(0.0230)	(0.0889)	(0.0191)	(0.0190)
12	Burr	1	-2	1.1343	0.6294	0.3542	1.2789	0.2334	0.2334
				(0.0675)	(0.1365)	(0.0450)	(0.6098)	(0.0272)	(0.0272)
13	Burr	1	-1	0.5337	1.5024	1.6781	3.0290	0.9942	0.9942
				(0.0511)	(0.1603)	(0.1565)	(0.9073)	(0.0834)	(0.0834)
14	Burr	1	-0.5	2.9970	6.6233	8.3157	8.9053	5.9706	5.9706
				(0.2597)	(0.4296)	(0.4863)	(1.0922)	(0.4028)	(0.4028)
15	Fréchet	1	-1	0.7780	0.9653	0.9142	1.2689	0.6711	0.6711
				(0.0624)	(0.0984)	(0.0715)	(0.3986)	(0.0559)	(0.0559)
16	*t*	1	-2	1.1532	0.5702	0.4833	1.3770	0.3493	0.3493
				(0.0742)	(0.0667)	(0.0491)	(0.6204)	(0.0329)	(0.0329)
17	logΓ	1	0	1.5714	6.3303	7.1256	5.5599	7.0225	6.5330
				(0.1208)	(0.3930)	(0.3842)	(0.3832)	(0.2613)	(0.2355)
γ = 1				**1.3613**	**2.7702**	**3.1452**	**3.5698**	**2.5402**	**2.4586**
				(0.0523)	(0.1051)	(0.1078)	(0.2924)	(0.0821)	(0.0799)
Mean				**0.5911**	**1.2527**	**1.3973**	**1.6419**	**1.1353**	**1.1004**
				(0.0190)	(0.0411)	(0.0389)	(0.1079)	(0.0297)	(0.0290)

Table 4.3.4.6 Results for $n = 2000$ for the Hill estimator.

					G&H		AMSE	D&K
Distr.		γ	ρ	**PPD**	**1.25**	**1.5**		
1	Burr	0.1	-2	0.0065	0.0029	0.0022	0.0027	0.0013
				(0.0008)	(0.0008)	(0.0004)	(0.0008)	(0.0002)
2	Burr	0.1	-1	0.0059	0.0154	0.0100	0.0256	0.0057
				(0.0007)	(0.0058)	(0.0015)	(0.0133)	(0.0006)
3	Burr	0.1	-0.5	0.0183	0.0507	0.0504	0.0349	0.0570
				(0.0022)	(0.0055)	(0.0043)	(0.0039)	(0.0032)
4	Fréchet	0.1	-1	0.0056	0.0044	0.0045	0.0136	0.0033
				(0.0006)	(0.0005)	(0.0005)	(0.0095)	(0.0004)
5	logΓ	0.1	0	0.0159	0.0490	0.0579	0.0471	0.0601
				(0.0015)	(0.0032)	(0.0031)	(0.0028)	(0.0022)
γ **= 0.1**				**0.0104**	**0.0245**	**0.0250**	**0.0248**	**0.0255**
				(0.0006)	(0.0017)	(0.0011)	(0.0034)	(0.0008)
6	Burr	0.5	-2	0.1095	0.0444	0.0462	0.0325	0.0370
				(0.0114)	(0.0064)	(0.0065)	(0.0045)	(0.0052)
7	Burr	0.5	-1	0.1008	0.2167	0.2172	0.1831	0.1244
				(0.0116)	(0.0286)	(0.0250)	(0.0267)	(0.0148)
8	Burr	0.5	-0.5	0.3770	0.9400	1.1530	0.8943	1.0266
				(0.0433)	(0.0927)	(0.0960)	(0.0823)	(0.0771)
9	Fréchet	0.5	-1	0.1172	0.0819	0.0961	0.0755	0.0733
				(0.0120)	(0.0101)	(0.0112)	(0.0100)	(0.0102)
10	*t*	0.5	-1	0.1356	0.5067	0.5088	0.6931	0.2140
				(0.0197)	(0.1635)	(0.1689)	(0.2982)	(0.0255)
11	logΓ	0.5	0	0.4582	1.3045	1.4287	1.4231	1.6446
				(0.0336)	(0.0760)	(0.0777)	(0.1951)	(0.0561)
γ **= 0.5**				**0.2164**	**0.5157**	**0.5750**	**0.5502**	**0.5200**
				(0.0103)	(0.0342)	(0.0352)	(0.0611)	(0.0167)
12	Burr	1	-2	0.7888	0.3484	0.2507	0.2605	0.1536
				(0.0761)	(0.0749)	(0.0644)	(0.0686)	(0.0226)
13	Burr	1	-1	0.4069	0.6690	0.8410	1.4208	0.5700
				(0.0541)	(0.0717)	(0.0847)	(0.8421)	(0.0631)
14	Burr	1	-0.5	1.7775	3.9258	4.6158	5.1370	3.3122
				(0.2028)	(0.3958)	(0.4116)	(1.1942)	(0.3300)
15	Fréchet	1	-1	0.5022	0.5354	0.3800	0.2465	0.2916
				(0.0508)	(0.1860)	(0.0358)	(0.0302)	(0.0342)
16	*t*	1	-2	0.7132	0.2418	0.2522	0.2407	0.1932
				(0.0724)	(0.0397)	(0.0311)	(0.0446)	(0.0369)
17	logΓ	1	0	1.3852	4.9985	5.4382	4.5661	5.8141
				(0.1085)	(0.4633)	(0.2640)	(0.2495)	(0.2248)
γ **= 1**				**0.9289**	**1.7865**	**1.9630**	**1.9786**	**1.7225**
				(0.0439)	(0.1078)	(0.0838)	(0.2475)	(0.0680)
Mean				**0.4073**	**0.8197**	**0.9031**	**0.8998**	**0.7990**
				(0.0159)	(0.0399)	(0.0321)	(0.0900)	(0.0247)

Table 4.3.4.7 Results for $n = 5000$ for the Hill estimator.

					G&H		AMSE	D&K
Distr.		γ	ρ	**PPD**	**1.25**	**1.5**		
1	Burr	0.1	-2	0.0042	0.0012	0.0012	0.0010	0.0007
				(0.0005)	(0.0003)	(0.0002)	(0.0001)	(0.0001)
2	Burr	0.1	-1	0.0053	0.0044	0.0038	0.0031	0.0030
				(0.0006)	(0.0008)	(0.0005)	(0.0004)	(0.0003)
3	Burr	0.1	-0.5	0.0111	0.0450	0.0305	0.0327	0.0400
				(0.0011)	(0.0141)	(0.0023)	(0.0099)	(0.0022)
4	Fréchet	0.1	-1	0.0040	0.0026	0.0026	0.0020	0.0020
				(0.0005)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
5	logΓ	0.1	0	0.0129	0.0395	0.0448	0.0368	0.0532
				(0.0010)	(0.0026)	(0.0026)	(0.0020)	(0.0015)
γ **= 0.1**				**0.0075**	**0.0186**	**0.0166**	**0.0151**	**0.0198**
				(0.0003)	(0.0029)	(0.0007)	(0.0020)	(0.0005)
6	Burr	0.5	-2	0.0914	0.0334	0.0308	0.0322	0.0180
				(0.0098)	(0.0052)	(0.0035)	(0.0064)	(0.0026)
7	Burr	0.5	-1	0.0793	0.1317	0.1007	0.0705	0.0788
				(0.0098)	(0.0320)	(0.0098)	(0.0099)	(0.0085)
8	Burr	0.5	-0.5	0.2843	0.6887	0.7936	0.8542	0.6138
				(0.0343)	(0.0626)	(0.0614)	(0.2373)	(0.0492)
9	Fréchet	0.5	-1	0.0843	0.0702	0.0750	0.0564	0.0610
				(0.0089)	(0.0073)	(0.0074)	(0.0055)	(0.0066)
10	*t*	0.5	-1	0.1036	0.1430	0.1583	0.1108	0.1020
				(0.0128)	(0.0188)	(0.0195)	(0.0145)	(0.0125)
11	logΓ	0.5	0	0.3863	1.0131	1.1370	1.1601	1.3402
				(0.0269)	(0.0549)	(0.0571)	(0.2246)	(0.0498)
γ **= 0.5**				**0.1715**	**0.3467**	**0.3825**	**0.3807**	**0.3690**
				(0.0080)	(0.0153)	(0.0145)	(0.0546)	(0.0120)
12	Burr	1	-2	0.3659	0.1884	0.1145	0.1182	0.0885
				(0.0376)	(0.0514)	(0.0148)	(0.0205)	(0.0122)
13	Burr	1	-1	0.3506	0.3569	0.3975	0.2822	0.2547
				(0.0386)	(0.0506)	(0.0409)	(0.0368)	(0.0301)
14	Burr	1	-0.5	1.1300	2.7380	3.1145	3.8372	2.1348
				(0.1286)	(0.2898)	(0.2564)	(1.1892)	(0.1834)
15	Fréchet	1	-1	0.2354	0.9065	0.3487	1.2449	0.2779
				(0.0257)	(0.4067)	(0.0319)	(0.9604)	(0.0293)
16	*t*	1	-2	0.4183	0.3132	0.1829	0.1122	0.0988
				(0.0435)	(0.1675)	(0.0607)	(0.0168)	(0.0143)
17	logΓ	1	0	1.5004	3.9507	4.3612	3.8596	5.2412
				(0.1069)	(0.2293)	(0.2288)	(0.2086)	(0.1737)
γ **= 1**				**0.6668**	**1.4089**	**1.4199**	**1.5757**	**1.3493**
				(0.0305)	(0.0965)	(0.0589)	(0.2572)	(0.0428)
Mean				**0.2981**	**0.6251**	**0.6410**	**0.6949**	**0.6123**
				(0.0111)	(0.0345)	(0.0214)	(0.0928)	(0.0157)

Table 4.3.4.8 Results for $n = 10\,000$ for the Hill estimator.

					G&H		AMSE	D&K
Distr.		γ	ρ	**PPD**	**1.25**	**1.5**		
1	Burr	0.1	-2	0.0027	0.0009	0.0005	0.0005	0.0004
				(0.0003)	(0.0003)	(0.0001)	(0.0001)	(0.0001)
2	Burr	0.1	-1	0.0030	0.0033	0.0036	0.0023	0.0024
				(0.0004)	(0.0007)	(0.0007)	(0.0003)	(0.0002)
3	Burr	0.1	-0.5	0.0098	0.0213	0.0222	0.0150	0.0311
				(0.0012)	(0.0036)	(0.0021)	(0.0013)	(0.0013)
4	Fréchet	0.1	-1	0.0024	0.0020	0.0018	0.0015	0.0015
				(0.0002)	(0.0004)	(0.0002)	(0.0002)	(0.0001)
5	logΓ	0.1	0	0.0132	0.0315	0.0361	0.0416	0.0484
				(0.0011)	(0.0017)	(0.0016)	(0.0096)	(0.0011)
γ = 0.1				**0.0062**	**0.0118**	**0.0128**	**0.0122**	**0.0168**
				(0.0004)	(0.0008)	(0.0005)	(0.0019)	(0.0003)
6	Burr	0.5	-2	0.1063	0.0137	0.0117	0.0115	0.0095
				(0.0157)	(0.0020)	(0.0014)	(0.0015)	(0.0014)
7	Burr	0.5	-1	0.1383	0.0663	0.0658	0.0458	0.0454
				(0.0311)	(0.0078)	(0.0071)	(0.0056)	(0.0048)
8	Burr	0.5	-0.5	0.1101	0.4802	0.4935	0.3636	0.4594
				(0.0129)	(0.0560)	(0.0393)	(0.0356)	(0.0361)
9	Fréchet	0.5	-1	0.1142	0.0394	0.0424	0.0296	0.0297
				(0.0128)	(0.0043)	(0.0042)	(0.0033)	(0.0031)
10	*t*	0.5	-1	0.0970	0.0892	0.1027	0.0738	0.0705
				(0.0119)	(0.0092)	(0.0101)	(0.0082)	(0.0077)
11	logΓ	0.5	0	0.2668	0.8139	0.9114	0.8229	1.1546
				(0.0301)	(0.0460)	(0.0449)	(0.0420)	(0.0304)
γ = 0.5				**0.1388**	**0.2505**	**0.2713**	**0.2245**	**0.2948**
				(0.0085)	(0.0123)	(0.0102)	(0.0093)	(0.0080)
12	Burr	1	-2	0.3776	0.1039	0.1034	0.0593	0.0416
				(0.0503)	(0.0405)	(0.0405)	(0.0107)	(0.0062)
13	Burr	1	-1	0.3830	0.3231	0.3627	0.2237	0.1939
				(0.0473)	(0.0378)	(0.0480)	(0.0232)	(0.0217)
14	Burr	1	-0.5	0.4330	1.9494	1.9859	1.5192	1.5278
				(0.0562)	(0.2703)	(0.1677)	(0.1561)	(0.1505)
15	Fréchet	1	-1	0.5100	0.2581	0.2082	1.0981	0.1492
				(0.0556)	(0.0530)	(0.0188)	(0.9269)	(0.0165)
16	*t*	1	-2	0.3610	0.0694	0.0721	0.0622	0.0603
				(0.0450)	(0.0077)	(0.0077)	(0.0076)	(0.0082)
17	logΓ	1	0	1.0225	3.4926	3.9046	3.4737	4.3372
				(0.0883)	(0.1509)	(0.1468)	(0.1433)	(0.1207)
γ = 1				**0.5145**	**1.0328**	**1.1061**	**1.0727**	**1.0517**
				(0.0241)	(0.0532)	(0.0387)	(0.1585)	(0.0325)
Mean				**0.2324**	**0.4564**	**0.4899**	**0.4614**	**0.4802**
				(0.0090)	(0.0193)	(0.0141)	(0.0561)	(0.0118)

Table 4.3.4.9 Results for $n = 20\,000$ for the Hill estimator.

From these tables one can see that the PPD estimator significantly outperforms the Hill estimator.

Of the threshold selection techniques considered for the Hill estimator, the one which consistently yielded the lowest MSE, was the method by Drees and Kaufmann, except for samples of size $n = 20\,000$, where the method by Guillou and Hall performed the best overall. Note, however, that for samples of sizes $n = 5000$ and greater, the difference between the MSEs yielded by method by Drees and Kaufmann and the method by Guillou and Hall, became insignificant.

It is a known fact in EVT that the Hill estimator does not perform very well. In order to assess the true relative performance of the PPD, it has to be compared to much wider variety of estimators in the literature. Some of these estimators are considered in the next section.

4.4 Other estimators based on log-spacings of order statistics

In Section 4.3 we restated the definition of the Hill estimator. The Hill estimator is calculated as follows:

$$H_{k,n} = \frac{1}{k}\sum_{i=1}^{k} \log\left(\frac{X_{n-i+1,n}}{X_{n-k,n}}\right).$$

In Section 3.1.3 an equivalent definition of the Hill estimator was also given:

$$H_{k,n} = \frac{1}{k}\sum_{i=1}^{k} i\log\left(\frac{X_{n-i+1,n}}{X_{n-i,n}}\right).$$

These formulas are based on the $k+1$ largest of the n order statistics $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$. Specifically, the first is based on $\log\left(\frac{X_{n-i+1,n}}{X_{n-k,n}}\right) = \log(X_{n-i+1,n}) - \log(X_{n-k,n})$, which is the log-spacing (distance between the logs) of the i-th largest and the $(k+1)$-th largest order statistics. The second formula is based on $\log\left(\frac{X_{n-i+1,n}}{X_{n-i,n}}\right) = \log(X_{n-i+1,n}) - \log(X_{n-i,n})$, which is the log-spacing of the i-th largest and the $(i+1)$-th largest order statistics.

In this section we consider a number of estimators which are based on either of these log-spacings, or moments of them. The estimators we consider in this section are those considered by Gomes and Martins (2002). The only exception is the moment estimator, which was introduced by Dekkers, Einmahl and De Haan (1989).

All estimators considered in this section are subject to the same assumptions made on the Hill estimator:

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent, identically distributed random variables with common Pareto type distribution function F concentrated on $[0, \infty)$. Let k $(2 \leq k \leq n-1)$ denote the number of excesses.

We assume the distribution F to be of Pareto type, and therefore the EVI to be positive. If we obtain a negative estimate of the EVI during the course of the simulation study, the estimate will be set to zero for the sake of calculating the MSE. This is done to avoid unfairly penalising estimators applicable to all domains of attraction (where the EVI can also assume 0 or negative values), for instance the moment estimator.

As in previous sections of this chapter, we only give the method of calculating the estimators. For more detail and discussions on how these estimators were derived, the references as mentioned in the text, may be consulted.

Most of the estimators in this section are based on the moment statistics

$$M_n^{(j)}(k) \coloneqq \frac{1}{k}\sum_{i=1}^{k}\left(\log\frac{X_{n-i+1,n}}{X_{n-k,n}}\right)^j, j \geq 1.$$

Note that $M_n^{(1)}(k)$ is the Hill estimator. In this section we also denote the Hill estimator by $\gamma_n^{(1)}(k)$, to keep the notation consistent with that of Gomes and Martins (2002).

The moment estimator of Dekkers et al. (1989) is defined as

$$\gamma_n^{MOM}(k) \coloneqq M_n^{(1)}(k) + 1 - \frac{1}{2}\left(1 - \frac{\left(M_n^{(1)}(k)\right)^2}{M_n^{(2)}(k)}\right)^{-1}.$$

Gomes and Martins (2002) also considered the following estimators:

$$\gamma_n^{(2)}(k) \coloneqq \frac{M_n^{(2)}(k)}{2M_n^{(1)}(k)}$$

$$\gamma_n^{(3)}(k) \coloneqq \sqrt{\frac{M_n^{(2)}(k)}{2}}$$

$$\gamma_n^{GJ}(k) \coloneqq 3\gamma_n^{(2)}(k) - 2\gamma_n^{(3)}(k)$$

$$\gamma_n^{GJ(\hat{\rho})}(k) \coloneqq \frac{-(2-\hat{\rho})\gamma_n^{(2)}(k) + 2\gamma_n^{(3)}(k)}{\hat{\rho}}$$

$$\gamma_n^{NGJ}(k) \coloneqq 2\gamma_n^{(1)}\left(\frac{k}{2}\right) - \gamma_n^{(1)}(k)$$

$\gamma_n^{NGJ(\hat{\rho})}(k) := \frac{\gamma_n^{(1)}(k) - 2^{-\hat{\rho}}\gamma_n^{(1)}([k/2])}{1-2^{-\hat{\rho}}}$

$\gamma_n^{P}(k) := 2\gamma_n^{(2)}(k) - \gamma_n^{(1)}(k)$

In the tables in Section 4.7, these estimators will be denoted by γ_2, γ_3, GJ, GJ(ρ), NGJ, NGJ(ρ) and Peng, respectively.

Note that in Gomes and Martins (2002) the definition of $\gamma_n^{(3)}(k)$ is erroneously stated as $\frac{\sqrt{M_n^{(2)}(k)}}{2}$. The correct definition is found in Gomes, Martins and Neves (2000).

The abbreviations GJ and NGJ refer to *generalized jackknife* and *natural generalized jackknife*, respectively.

The estimators $\gamma_n^{GJ(\hat{\rho})}(k)$ and $\gamma_n^{NGJ(\hat{\rho})}(k)$ are functions of $\hat{\rho}$, an estimator of the second order parameter ρ. In Section 4.1 we considered one such estimator of ρ in detail. In Section 4.2 we considered some estimators which were constructed by first estimating ρ for a given data set, and then substituting this estimated value into the PPD density function as a given value of ρ. This *external estimation* of the the second order parameter is also applied here.

As is often the case, ρ is sometimes set to -1, as an alternative to estimating ρ. The estimators $\gamma_n^{GJ}(k)$ and $\gamma_n^{NGJ}(k)$ can be obtained by substituting $\hat{\rho} = -1$ into the formulas for $\gamma_n^{GJ(\hat{\rho})}(k)$ and $\gamma_n^{NGJ(\hat{\rho})}(k)$, respectively.

The estimator of ρ in Section 4.1 can also be used here, but we will limit ourselves to the estimators of ρ considered by Gomes and Martins (2002).

The following estimator is suggested by Peng (1998):

$$\hat{\rho}_2 := -\frac{1}{\log 2}\left|\log\left|\frac{M_n^{(2)}\left(\left[\frac{n}{2\log n}\right]\right) - 2\left(M_n^{(1)}\left(\left[\frac{n}{2\log n}\right]\right)\right)^2}{M_n^{(2)}\left(\left[\frac{n}{\log n}\right]\right) - 2\left(M_n^{(1)}\left(\left[\frac{n}{\log n}\right]\right)\right)^2}\right|\right|.$$

Peng (1998) also derived an “asymptotically unbiased” estimator,

$$\gamma_n^{P(\hat{\rho})}(k) := \frac{\gamma_n^{(1)}(k) - (1-\hat{\rho})\gamma_n^{(2)}(k)}{\hat{\rho}}.$$

If one substitutes $\hat{\rho} = -1$ into the formula for $\gamma_n^{P(\hat{\rho})}(k)$, one arrives at

$$\gamma_n^{P}(k) := 2\gamma_n^{(2)}(k) - \gamma_n^{(1)}(k),$$

which is the estimator we have in the list above.

We will only consider the estimator $\gamma_n^{P}(k)$, and not $\gamma_n^{P(\hat{\rho}_2)}(k)$. The latter yields very poor results from an MSE point of view. These (poor) results were also obtained by Gomes and Martins (2002), where they remarked that $\hat{\rho}_2$ induces high variances when incorporated into the estimators considered. Refer to Remark 3.3 on p. 18 in the article by Gomes and Martins (2002).

The following two estimators of ρ are particular members of estimators proposed by Fraga Alves, Gomes and De Haan (2003). They are

$$\hat{\rho}_3 := -\left|\frac{3\left(T_n^{(0)}(k_1)-1\right)}{T_n^{(0)}(k_1)-3}\right| \quad \text{and}$$

$$\hat{\rho}_4 := -\left|\frac{3\left(T_n^{(1)}(k_1)-1\right)}{T_n^{(1)}(k_1)-3}\right|,$$

where $k_1 = \min(n-1, [2n/\log\log n])$, and

$$T_n^{(\tau)}(k) := \begin{cases} \dfrac{\left(M_n^{(1)}(k)\right)^{\tau} - \left(M_n^{(1)}(k)/2\right)^{\tau/2}}{\left(M_n^{(2)}(k)/2\right)^{\tau/2} - \left(M_n^{(3)}(k)/6\right)^{\tau/3}} & \text{if } \tau > 0 \\ \dfrac{\log\left(M_n^{(1)}(k)\right) - \frac{1}{2}\log\left(M_n^{(1)}(k)/2\right)}{\frac{1}{2}\log\left(M_n^{(2)}(k)/2\right) - \frac{1}{3}\log\left(M_n^{(3)}(k)/6\right)} & \text{if } \tau = 0 \end{cases}.$$

We set the minimum value of the estimators $\hat{\rho}_3$ and $\hat{\rho}_4$ at -10, since this restriction was applied to the estimator of Section 4.1, and also to avoid numerical problems.

Finally, we will also consider a maximum likelihood estimator of the EVI, as it appears in Gomes and Martins (2002), namely

$$\gamma_n^{ML}(k) := \frac{1}{k}\sum_{i=1}^{k} U_i - \left(\frac{1}{k}\sum_{i=1}^{k} iU_i\right)\frac{\sum_{i=1}^{k}(2i-k-1)U_i}{\sum_{i=1}^{k} i(2i-k-1)U_i},$$

where $U_i = i\log\left(\frac{X_{n-i+1,n}}{X_{n-i,n}}\right)$, $1 \le i \le k$.

Note that the first part of the formula for $\gamma_n^{ML}(k)$ is $\frac{1}{k}\sum_{i=1}^{k} U_i$, which is the equivalent definition of the Hill estimator, as stated at the beginning of this section.

The MLE $\gamma_n^{ML}(k)$ is obtained by substituting $\hat{\rho} = -1$ into the more general expression of the MLE,

$$\gamma_n^{ML(\hat{\rho})}(k) := \frac{1}{k}\sum_{i=1}^{k} U_i - \left(\frac{1}{k}\sum_{i=1}^{k} i^{-\hat{\rho}}U_i\right)\frac{\left(\sum_{i=1}^{k} i^{-\hat{\rho}}\right)\left(\sum_{i=1}^{k} U_i\right) - k\left(\sum_{i=1}^{k} i^{-\hat{\rho}}U_i\right)}{\left(\sum_{i=1}^{k} i^{-\hat{\rho}}\right)\left(\sum_{i=1}^{k} i^{-\hat{\rho}}U_i\right) - k\left(\sum_{i=1}^{k} i^{-2\hat{\rho}}U_i\right)}.$$

As far as these MLEs are concerned, we will consider three estimators, namely $\gamma_n^{ML}(k)$, $\gamma_n^{ML(\hat{\rho}_3)}(k)$ and $\gamma_n^{ML(\hat{\rho}_4)}(k)$, which will be denoted by ML, ML(ρ_3) and ML(ρ_4), respectively, in the tables in Section 4.7.

In Section 4.2 estimators based on the PPD were considered. The PPD was fitted to multplicative excesses. In the next section the focus will be on estimators of the EVI obtained by fitting the GPD to additive excesses.

4.5 Estimators based on the GPD

The generalized Pareto distribution (GPD) was defined in Section 2.4.3. In this section we apply the restriction $\gamma > 0$.

The GPD has distribution function

$$F(z) = 1 - \left(1 + \frac{\gamma z}{\sigma}\right)^{-\frac{1}{\gamma}},$$

where $z \geq 0$, $\gamma > 0$ and $\sigma > 0$.

The density function of the GPD is

$$f(z) = \frac{1}{\sigma}\left(1 + \frac{\gamma z}{\sigma}\right)^{-\frac{1}{\gamma}-1}.$$

Given n observations from a heavy-tailed distribution, and number of excesses k (implying $x_{n-k,n}$ as the threshold), we calculate additive excesses $z_i = x_{n-i+1,n} - x_{n-k,n}$, $i = 1,2,\ldots,k$. We then fit the GPD to these excesses to obtain estimates of γ and σ, where γ is the EVI.

In this section we consider both Bayesian and maximum likelihood estimation of the GPD parameters.

4.5.1 Bayesian estimation of GPD parameters

Section 3.2 discussed the Bayesian estimation of PPD parameters in detail. We apply exactly the same concepts to the estimation of the GPD parameters. In this section we will only state the general ideas and how they correspond to or differ from those in Section 3.2. We will also discuss aspects of simulation in the same order as that of the PPD, in order to ease comparison.

In Section 3.2.2 we discussed the MDI prior. We also use this prior for the GPD. For the GPD the MDI prior is given by $\frac{1}{\sigma}e^{-\gamma}$ (Beirlant et al., 2004).

This results in a posterior distribution given by

$$\frac{1}{\sigma^{k+1}} e^{-\gamma} \prod_{i=1}^{k} \left(1 + \frac{\gamma z_i}{\sigma}\right)^{-\frac{1}{\gamma}-1}.$$

We also applied Gibbs sampling to determine the estimates. As an initial value for γ, we use the Hill estimate. We use 10 burn-in draws and 1000 draws which are retained. We also apply the rejection method to obtain individual draws of parameters.

In the detailed description of Step 2.2 in Section 3.2.3 we showed how a more efficient procedure was obtained by making use of two iterations, on each iteration dividing the range of values for γ into 20 intervals. Also, we based our initial range of a parameter on the range obtained by the previous Gibbs cycle. Both these methods are applied here as well.

The computationally intensive part of the posterior, namely

$$\prod_{i=1}^{k} \left(1 + \frac{\gamma z_i}{\sigma}\right)^{-\frac{1}{\gamma}-1},$$

is also approximated in a fashion similar to the approximation used for the PPD.

The Bayesian estimate can again be calculated as either the mean, median or mode of the Gibbs draws.

The Bayesian estimator of γ is denoted by GPD Bayes in the tables of Section 4.7.

4.5.2 Maximum likelihood estimation of GPD parameters

The likelihood function of the GPD is given by

$$\frac{1}{\sigma^{k}} \prod_{i=1}^{k} \left(1 + \frac{\gamma z_i}{\sigma}\right)^{-\frac{1}{\gamma}-1}.$$

The MLEs of γ and σ are obtained by numerically maximizing the log likelihood function

$$\ell(\gamma, \sigma) = -k \log \sigma - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{k} \log\left(1 + \frac{\gamma z_i}{\sigma}\right)$$

over γ and σ.

Since we apply the restriction $\gamma > 0$, we set $\hat{\gamma} = 0$ if we obtain $\hat{\gamma} \leq 0$.

We applied two methods of maximizing over γ and σ.

The first method is to determine

$$\hat{\sigma} = \underset{\sigma>0}{\operatorname{argmax}} \max_{\gamma>0} \ell(\gamma, \sigma),$$

and then

$$\hat{\gamma} = \underset{\gamma>0}{\operatorname{argmax}} \, \ell(\gamma, \hat{\sigma}).$$

The second method is to determine

$$\hat{\gamma} = \underset{\gamma>0}{\operatorname{argmax}} \max_{\sigma>0} \ell(\gamma, \sigma),$$

and then

$$\hat{\sigma} = \underset{\sigma>0}{\operatorname{argmax}} \, \ell(\hat{\gamma}, \sigma).$$

The first method computes faster for smaller sample sizes, and the second method faster for larger samples. We applied the first method if $k \leq 2000$, and the second method if $k > 2000$.

The GPD is notorious for causing numerical difficulties. Except for the ability to choose the most efficient method, another advantage of having two methods is that an alternative method is available if the original method runs into numerical problems.

The difference in speed of execution and numerical stability is not only due to the order of maximization of γ and σ as shown above. Different techniques of maximization were used.

The first method was used directly as stated above, namely by simply calculating the log likelihood at different values of γ and σ, and finding the values of γ and σ yielding the maximum.

The second method used a derivative calculation for the “inner” maximization, that is for determining

$$\max_{\sigma>0} \ell(\gamma, \sigma).$$

Here we regard γ as a constant. Taking the derivative of $\ell(\gamma, \sigma)$ with respect to σ, we have

$$\frac{\partial \ell(\gamma, \sigma)}{\partial \sigma} = -\frac{k}{\sigma} + \left(\frac{1}{\gamma} + 1\right) \frac{\gamma}{\sigma^2} \sum_{i=1}^{k} \frac{z_i}{\left(1 + \frac{\gamma z_i}{\sigma}\right)}.$$

Setting this expression equal to zero, yields

$$\sum_{i=1}^{k} \frac{z_i}{(\hat{\sigma} + \gamma z_i)} - \frac{k}{1+\gamma} = 0,$$

the solution of which is obtained numerically.

The "outer" maximization (relative to γ) in

$$\hat{\gamma} = \underset{\gamma>0}{\operatorname{argmax}} \max_{\sigma>0} \ell(\gamma, \sigma)$$

is then performed directly on the log likelihood, that is by calculating the log likelihood at different values of γ, and finding the value of γ yielding the maximum.

There are still many more estimators in the literature. In the following section, we consider some estimators which do not fall in the classes of estimators we have discussed in Sections 4.2 to 4.5.

4.6 Other well-known estimators

In this section we will consider the following estimators:

- The Zipf estimator.
- The Pickands estimator.
- An estimator based on the exponential regression model.

These estimators will be described in this section, and the simulation results pertaining to these estimators, will be presented in Section 4.7.

4.6.1 The Zipf estimator

For $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ and a given value of k, Zipf (1949) defined the following estimator of the EVI:

$$\hat{\gamma}_n^Z(k) := \frac{\frac{1}{k}\sum_{i=1}^{k}\left(\log \frac{k+1}{i+1} - c_1\right)\log UH_{i,n}}{c_2 - c_1^2},$$

where $c_1 := \frac{1}{k}\sum_{i=1}^{k} \log \frac{k+1}{i+1}$, $c_2 := \frac{1}{k}\sum_{i=1}^{k}\left(\log \frac{k+1}{i+1}\right)^2$ and $UH_{i,n} := X_{n-i,n} H_{i,n}$.

$H_{i,n}$ again denotes the Hill estimator based on the i largest order statistics from n observations, and k again denotes the number of largest order statistics used in the calculation of the estimate.

The Zipf estimator is applicable to all domains of attraction, but we apply the restriction $\gamma > 0$. Following the same reasoning as in Section 4.4, we set $\hat{\gamma}_n^Z(k) = 0$ if we obtain $\hat{\gamma}_n^Z(k) < 0$, for the purpose of calculating the MSE.

Simulation results will be given in Section 4.7.

4.6.2 The Pickands estimator

Pickands (1975) defined the following estimator of the EVI for $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ and a given value of k:

$$\hat{\gamma}_n^{Pi}(k) := \frac{1}{\log 2} \log \frac{X_{n-k+1,n} - X_{n-2k+1,n}}{X_{n-2k+1,n} - X_{n-4k+1,n}},$$

where $k \leq n/4$.

The Pickands estimator is applicable to all domains of attraction, but we apply the restriction $\gamma > 0$. Following the same reasoning as in Section 4.4, we set $\hat{\gamma}_n^{Pi}(k) = 0$ for the purpose of calculating the MSE, if we obtain $\hat{\gamma}_n^{Pi}(k) < 0$.

The Pickands estimator is known to be inferior and is generally not used in comparisons. We also found this to be the case, and therefore the simulation results pertaining to the Pickands estimator will be excluded from Section 4.7.

4.6.3 An estimator based on the exponential regression model

Matthys and Beirlant (2003) show that, for $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ and a given value of k,

$$Y_i := i \log \frac{X_{n-i+1,n} - X_{n-k,n}}{X_{n-i,n} - X_{n-k,n}} \quad \overset{\mathcal{D}}{\approx} \quad \frac{\gamma}{1 - \left(\frac{i}{k+1}\right)^{\gamma}} \; E_i,$$

for $i = 1,2, \ldots, k-1$, where γ is the EVI, and $E_1, E_2, \ldots, E_{k-1}$ are independent standard exponential random variables.

Let $\lambda(\gamma, i, k) = \frac{1-\left(\frac{i}{k+1}\right)^{\gamma}}{\gamma}$. We therefore have

$$Y_i \approx \frac{1}{\lambda(\gamma, i, k)} \; E_i \sim exp\big(\lambda(\gamma, i, k)\big),$$

for $i = 1,2, \ldots, k-1$.

For observations $y_1, y_2, \dots, y_{k-1}$, we have the log likelihood function

$$\ell(\gamma) = \log\left(\prod_{i=1}^{k-1} \lambda(\gamma, i, k) \exp(-\lambda(\gamma, i, k) y_i)\right) = \sum_{i=1}^{k-1} (\log \lambda(\gamma, i, k) - \lambda(\gamma, i, k) y_i),$$

which is numerically maximized with respect to γ, yielding an estimate of the EVI which we will denote by $\hat{\gamma}_n^{ERM}(k)$.

The exponential regression model (ERM) estimator is applicable to all domains of attraction, but we apply the restriction $\gamma > 0$. Following the same reasoning as in Section 4.4, we set $\hat{\gamma}_n^{ERM}(k) = 0$ for the purpose of calculating the MSE, if we obtain $\hat{\gamma}_n^{ERM}(k) < 0$.

In the tables in Section 4.7, we will denote this estimator by ERM.

The simulation results pertaining to the estimators mentioned in this and the previous two sections, will be presented in the section to follow.

4.7 Simulation results for Sections 4.4 to 4.6

In this section we present the simulation results pertaining to the estimators discussed in Sections 4.4 to 4.6.

The number of estimators considered in these sections is so large, that we will present it in a format which differs from the previous formats. Each combination of sample size and EVI group will now be presented in a separate table. There are nine sample sizes ($n = 50$ to $n = 20\,000$) and three EVI groups ($\gamma = 0.1$, $\gamma = 0.5$ and $\gamma = 1$), which result in 27 tables, namely Tables 4.7.1 to 4.7.27 below.

The estimators are now given by rows and the distributions by columns, whereas previously we had it the other way round. The first two rows show the distribution from which samples were generated. Specifically, the first row shows the name of the family of distributions, and the second row the value of the second order parameter ρ. (The value of the EVI is given in the title of the table.) For more details on the distributions and number of repetitions per sample size, see Tables 3.3.2.5 and 3.3.1.2, respectively.

The results contained in rows three and four (PPD) correspond to the PPD columns of Tables 3.4.4.1 to 3.4.4.6, which are the results of the benchmark estimator.

As previously, the entries in the table are 100×MSEs, and their corresponding standard errors given below them, in brackets. The second last column shows the mean (in bold) for that EVI group (the mean of the values in that row), together with the corresponding standard error below it, in brackets.

The final column indicates which fixed threshold yielded the lowest MSE for that sample size – EVI group – estimator combination, expressed as percentage of the sample size. As an example, consider the last two columns in the seventh row of Table 4.7.1. The value of 0.9086 was obtained as follows:

1. At $k = 5\%$ of n, obtain the Moment estimate of the EVI for each of 1000 samples of size $n = 50$ from Distribution 1 (Burr distribution with $\gamma = 0.1$ and $\rho = -2$), and calculate the MSE.
2. Repeat Step 1, but for Distributions 2 to 5.
3. Calculate the mean MSE for Distributions 1 to 5 (at $k = 5\%$ of n).
4. Repeat Steps 1 to 3 for $k = 10\%, 15\%, \dots, 95\%$ of n.
5. Find the lowest of the means over all values of k obtained in Steps 3 and 4. In this case the lowest value was 0.9086, and that was at $k = 60\%$ of n.

The point to note is that the values 0.9030, 0.9179, …, 0.9042 in the same row were all at at $k = 60\%$ of n.

Note that the PPD estimator does not have an entry in the last column, since it makes use of threshold selection. The results shown for the Hill estimator does not include the threshold selection procedures of Section 4.3.

For the Bayesian estimation of the GPD parameter γ (denoted GPD Bayes in the table), one can calculate the mean, median or mode of the Gibbs draws. We only show the best perfoming choice in the tables below. This choice is indicated in the tables. Specifically, for EVI groups $\gamma = 0.1$ and $\gamma = 0.5$, the lowest MSE for all sample sizes was obtained by calculating the mean. For the EVI group $\gamma = 1$, the lowest MSE was obtained by calculating the median for sample sizes of up to $n = 1000$, and calculating the mode for sample sizes $n = 2000$ and larger.

The disadvantage of presenting the tables in the format below, is that the mean performance across all EVI groups is not included. However, since the EVI group $\gamma = 1$ is by far the greatest contributor to the mean MSE over all EVI groups, one could use the results from the EVI group $\gamma = 1$ as an indication of the relative performance across all EVI groups.

When comparing the simulation results, it is important to keep in mind that several estimators are designed for the general case where the EVI can also be zero or negative, for example the Zipf estimator described in Section 4.6.1. These estimators have the disadvantage of not making the *a priori* assumption that the EVI is positive, and should be expected to perform somewhat inferior to those assuming data to be from a Pareto type distribution.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**0**		
PPD	0.0677	0.0411	0.1812	0.0491	0.0406	**0.0759**	
	(0.0023)	(0.0016)	(0.0064)	(0.0018)	(0.0019)	(0.0015)	
Hill	0.1047	0.1289	0.4151	0.1036	0.2169	**0.1938**	20%
	(0.0057)	(0.0071)	(0.0166)	(0.0057)	(0.0105)	(0.0045)	
Moment	0.9030	0.9179	0.9353	0.8828	0.9042	**0.9086**	60%
	(0.0236)	(0.0103)	(0.0072)	(0.0200)	(0.0261)	(0.0085)	
γ_2	0.1071	0.0971	0.2497	0.0879	0.1472	**0.1378**	30%
	(0.0075)	(0.0065)	(0.0115)	(0.0050)	(0.0096)	(0.0037)	
γ_3	0.0862	0.0940	0.3244	0.0783	0.1483	**0.1463**	25%
	(0.0053)	(0.0051)	(0.0129)	(0.0041)	(0.0073)	(0.0034)	
GJ	0.0826	0.0384	0.1826	0.0593	0.0913	**0.0908**	90%
	(0.0036)	(0.0015)	(0.0058)	(0.0027)	(0.0058)	(0.0019)	
GJ(ρ_3)	0.1208	0.1234	0.1232	0.1150	0.2496	**0.1464**	75%
	(0.0053)	(0.0048)	(0.0058)	(0.0050)	(0.0183)	(0.0042)	
GJ(ρ_4)	0.0950	0.0807	0.2539	0.0858	0.2052	**0.1441**	50%
	(0.0075)	(0.0049)	(0.0105)	(0.0046)	(0.0204)	(0.0050)	
NGJ	0.1850	0.2653	0.4361	0.1794	0.2461	**0.2624**	70%
	(0.0080)	(0.0099)	(0.0148)	(0.0075)	(0.0126)	(0.0049)	
NGJ(ρ_3)	0.2053	0.3294	0.5973	0.2491	0.7228	**0.4208**	50%
	(0.0092)	(0.0123)	(0.0240)	(0.0102)	(0.1182)	(0.0244)	
NGJ(ρ_4)	0.0940	0.1326	0.3543	0.1058	0.2865	**0.1946**	80%
	(0.0060)	(0.0063)	(0.0148)	(0.0049)	(0.0260)	(0.0063)	
Peng	0.1048	0.0693	0.0823	0.0712	0.0930	**0.0841**	90%
	(0.0040)	(0.0022)	(0.0042)	(0.0029)	(0.0057)	(0.0018)	
ML	0.0563	0.0315	0.2764	0.0407	0.0858	**0.0981**	95%
	(0.0019)	(0.0014)	(0.0074)	(0.0017)	(0.0050)	(0.0019)	
ML(ρ_3)	0.0769	0.0574	0.1254	0.0703	0.1554	**0.0971**	90%
	(0.0028)	(0.0024)	(0.0057)	(0.0028)	(0.0140)	(0.0032)	
ML(ρ_4)	0.0631	0.1075	0.5394	0.0741	0.2199	**0.2008**	65%
	(0.0032)	(0.0055)	(0.0155)	(0.0038)	(0.0183)	(0.0050)	
GPD Bayes	0.1140	0.1417	0.1571	0.1175	0.3486	**0.1758**	90%
Mean	(0.0072)	(0.0024)	(0.0021)	(0.0122)	(0.0289)	(0.0065)	
GPD MLE	0.9271	0.9720	0.9782	0.9246	0.8774	**0.9359**	85%
	(0.0069)	(0.0023)	(0.0012)	(0.0105)	(0.0167)	(0.0042)	
Zipf	0.9972	0.9232	0.9400	0.9823	1.0955	**0.9876**	95%
	(0.0359)	(0.0105)	(0.0083)	(0.0300)	(0.0439)	(0.0131)	
ERM	0.9450	0.9952	0.9986	0.9369	0.8980	**0.9547**	90%
	(0.0073)	(0.0017)	(0.0009)	(0.0122)	(0.0232)	(0.0054)	

Table 4.7.1 Simulation results for $n = 50$ and EVI group $\gamma = 0.1$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	***t***	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	1.8210	0.9261	3.8526	1.3757	0.8748	0.9493	**1.6332**	
	(0.0547)	(0.0390)	(0.1432)	(0.0469)	(0.0428)	(0.0437)	(0.0293)	
Hill	2.5106	3.4406	10.6748	2.5838	4.0936	5.2822	**4.7642**	20%
	(0.1208)	(0.1835)	(0.4667)	(0.1362)	(0.2124)	(0.2624)	(0.1052)	
Moment	6.5136	6.8734	9.6331	5.8528	7.4830	4.8302	**6.8644**	60%
	(0.2461)	(0.2565)	(0.3677)	(0.2355)	(0.2664)	(0.1978)	(0.1089)	
$\boldsymbol{\gamma_2}$	2.4937	2.5211	6.0893	2.2447	2.6025	3.6584	**3.2683**	30%
	(0.1357)	(0.1581)	(0.2738)	(0.1255)	(0.1758)	(0.2211)	(0.0771)	
$\boldsymbol{\gamma_3}$	2.0354	2.5403	8.0377	1.9726	2.8684	3.8106	**3.5441**	25%
	(0.0929)	(0.1347)	(0.3110)	(0.1033)	(0.1517)	(0.1921)	(0.0734)	
GJ	2.0572	0.8454	4.6647	1.5416	0.8768	2.4340	**2.0699**	90%
	(0.0594)	(0.0398)	(0.1476)	(0.0617)	(0.0443)	(0.1452)	(0.0386)	
GJ($\boldsymbol{\rho_3}$)	2.9346	3.0465	3.5501	3.0255	3.1532	6.2638	**3.6623**	70%
	(0.1084)	(0.1290)	(0.1782)	(0.1235)	(0.1229)	(0.3502)	(0.0769)	
GJ($\boldsymbol{\rho_4}$)	2.1642	2.1817	6.3607	2.2012	2.2057	5.2029	**3.3861**	50%
	(0.1012)	(0.1395)	(0.2585)	(0.1230)	(0.1463)	(0.4510)	(0.0967)	
NGJ	4.6352	6.2602	10.8012	4.7262	8.1458	6.9766	**6.9242**	60%
	(0.2001)	(0.2737)	(0.4621)	(0.2110)	(0.2898)	(0.3722)	(0.1286)	
NGJ($\boldsymbol{\rho_3}$)	5.3662	8.8204	14.8374	5.9308	11.4516	15.7978	**10.3674**	50%
	(0.2255)	(0.3336)	(0.5151)	(0.2375)	(0.3952)	(1.4865)	(0.2813)	
NGJ($\boldsymbol{\rho_4}$)	2.6888	3.9599	10.5215	3.1941	4.2222	10.0072	**5.7656**	50%
	(0.1271)	(0.1947)	(0.4352)	(0.1498)	(0.2167)	(0.6814)	(0.1469)	
Peng	2.6398	1.5400	2.1611	1.8527	1.2246	2.4857	**1.9840**	90%
	(0.0737)	(0.0567)	(0.1052)	(0.0677)	(0.0475)	(0.1439)	(0.0362)	
ML	1.4606	0.7004	6.9854	1.0552	1.4246	2.3463	**2.3287**	95%
	(0.0445)	(0.0335)	(0.1848)	(0.0403)	(0.0607)	(0.1374)	(0.0413)	
ML($\boldsymbol{\rho_3}$)	2.0107	1.1330	3.1101	1.7632	0.8424	3.1841	**2.0072**	95%
	(0.0635)	(0.0526)	(0.1323)	(0.0673)	(0.0402)	(0.1933)	(0.0434)	
ML($\boldsymbol{\rho_4}$)	1.5181	2.5915	13.9336	1.6682	3.9921	6.2056	**4.9848**	65%
	(0.0812)	(0.1253)	(0.4246)	(0.0871)	(0.1487)	(1.5006)	(0.2627)	
GPD Bayes	5.3062	5.2869	5.4388	4.5944	5.3851	5.5701	**5.2636**	75%
Mean	(0.1563)	(0.1581)	(0.2342)	(0.1460)	(0.1533)	(0.2568)	(0.0773)	
GPD MLE	9.5631	9.8955	7.0238	8.3569	10.2314	6.6059	**8.6128**	75%
	(0.2799)	(0.2923)	(0.2713)	(0.2668)	(0.2910)	(0.2790)	(0.1144)	
Zipf	7.1576	6.7649	14.6305	6.5155	6.9328	7.3160	**8.2195**	80%
	(0.2591)	(0.2546)	(0.5282)	(0.2446)	(0.2758)	(0.2716)	(0.1313)	
ERM	8.1895	8.2344	6.7851	6.7253	8.3444	6.7838	**7.5104**	80%
	(0.2674)	(0.2719)	(0.2798)	(0.2387)	(0.2706)	(0.2874)	(0.1101)	

Table 4.7.2 Simulation results for $n = 50$ and EVI group $\gamma = 0.5$.

Distr.	Burr	Burr	Burr	Fréchet	*t*	logΓ	Mean	*k*
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	8.1259	3.8879	12.3702	5.9964	7.2427	3.4014	**6.8374**	
	(0.2331)	(0.1571)	(0.5180)	(0.2113)	(0.2185)	(0.1525)	(0.1134)	
Hill	7.5753	12.2141	50.4772	9.3195	8.3653	19.3000	**17.8752**	25%
	(0.4056)	(0.6735)	(1.6828)	(0.4427)	(0.4304)	(0.9272)	(0.3610)	
Moment	12.4526	12.7052	36.7290	13.2760	14.5381	14.3672	**17.3447**	40%
	(0.5917)	(0.6125)	(1.3152)	(0.6292)	(0.6626)	(0.6378)	(0.3204)	
$\boldsymbol{\gamma_2}$	7.7320	8.5394	29.2320	8.9409	8.9665	13.6116	**12.8371**	35%
	(0.4082)	(0.5042)	(1.4500)	(0.5458)	(0.5808)	(0.8199)	(0.3262)	
$\boldsymbol{\gamma_3}$	7.7318	9.3235	31.0933	8.8351	8.6965	14.7076	**13.3980**	25%
	(0.3896)	(0.5281)	(1.2409)	(0.4381)	(0.4561)	(0.7744)	(0.2872)	
GJ	8.1286	3.6858	18.6514	6.2978	7.4470	8.9215	**8.8553**	90%
	(0.2629)	(0.1871)	(0.6119)	(0.2612)	(0.2342)	(0.6385)	(0.1674)	
GJ(ρ_3)	11.4240	12.9977	12.8548	12.3795	16.3281	24.0976	**15.0136**	80%
	(0.4301)	(0.4929)	(0.6145)	(0.5228)	(0.5348)	(1.2680)	(0.2874)	
GJ(ρ_4)	7.2671	7.5942	29.4589	8.8431	7.7184	19.9849	**13.4778**	55%
	(0.3567)	(0.4746)	(1.1121)	(0.5211)	(0.4382)	(1.5913)	(0.3569)	
NGJ	16.9058	29.0810	43.2194	17.5715	21.9099	25.3435	**25.6718**	65%
	(0.7274)	(1.1108)	(1.6972)	(0.7391)	(0.7994)	(1.2598)	(0.4539)	
NGJ(ρ_3)	20.2785	33.1098	58.4830	25.8227	27.7360	82.2366	**41.2778**	50%
	(0.8593)	(1.2697)	(2.1044)	(1.0248)	(1.0383)	(14.4923)	(2.4661)	
NGJ(ρ_4)	9.1482	16.1260	34.7419	10.3037	15.3897	55.3531	**23.5104**	90%
	(0.4037)	(0.7288)	(1.4857)	(0.4675)	(0.6288)	(27.2626)	(4.5545)	
Peng	9.7983	6.2829	9.8384	7.8638	11.1418	9.7423	**9.1113**	85%
	(0.3460)	(0.2467)	(0.5159)	(0.3239)	(0.3365)	(0.6493)	(0.1737)	
ML	5.8836	3.0492	27.8191	4.2745	4.7790	8.8115	**9.1028**	95%
	(0.1874)	(0.1609)	(0.7391)	(0.1618)	(0.1579)	(0.6359)	(0.1718)	
ML(ρ_3)	7.9818	5.0049	12.6359	7.2359	8.3581	14.9917	**9.3680**	95%
	(0.2648)	(0.2096)	(0.5183)	(0.2732)	(0.2605)	(2.3539)	(0.4105)	
ML(ρ_4)	7.4798	12.0095	46.3047	9.8096	7.3638	24.7801	**17.9579**	55%
	(0.4920)	(0.8330)	(1.6317)	(0.5923)	(0.3832)	(2.4258)	(0.5265)	
GPD Bayes	12.7556	12.5525	30.0131	13.1925	14.4694	14.6287	**16.2686**	60%
Median	(0.5180)	(0.5623)	(1.2673)	(0.5450)	(0.5650)	(0.6775)	(0.3012)	
GPD MLE	15.4518	15.2726	38.1007	16.1251	17.5866	17.8872	**20.0706**	60%
	(0.6604)	(0.7021)	(1.6841)	(0.6969)	(0.7201)	(0.8593)	(0.3912)	
Zipf	15.5993	17.2777	48.8769	17.1959	16.7965	23.0861	**23.1387**	50%
	(0.6872)	(0.8135)	(1.6350)	(0.7228)	(0.7292)	(1.0043)	(0.4038)	
ERM	12.9262	14.5939	54.5218	13.8062	14.6548	21.2426	**21.9576**	65%
	(0.5823)	(0.7151)	(2.0256)	(0.6120)	(0.6331)	(1.0017)	(0.4324)	

Table 4.7.3 Simulation results for $n = 50$ and EVI group $\gamma = 1$.

Distr.	Burr	Burr	Burr	Fréchet	logΓ	Mean	k
ρ	**-2**	**-1**	**-0.5**	**-1**	**0**		
PPD	0.0521	0.0274	0.1177	0.0350	0.0325	**0.0529**	
	(0.0024)	(0.0017)	(0.0063)	(0.0019)	(0.0024)	(0.0015)	
Hill	0.0687	0.0873	0.2606	0.0648	0.1331	**0.1229**	15%
	(0.0047)	(0.0062)	(0.0158)	(0.0047)	(0.0096)	(0.0041)	
Moment	0.7971	0.8439	0.9218	0.7536	0.7680	**0.8169**	50%
	(0.0289)	(0.0185)	(0.0106)	(0.0213)	(0.0295)	(0.0102)	
γ_2	0.0618	0.0638	0.1745	0.0534	0.1022	**0.0911**	25%
	(0.0042)	(0.0060)	(0.0117)	(0.0044)	(0.0082)	(0.0033)	
γ_3	0.0529	0.0634	0.2062	0.0490	0.1046	**0.0952**	20%
	(0.0034)	(0.0052)	(0.0121)	(0.0036)	(0.0076)	(0.0032)	
GJ	0.0495	0.0264	0.1356	0.0342	0.0691	**0.0630**	80%
	(0.0025)	(0.0022)	(0.0059)	(0.0021)	(0.0055)	(0.0018)	
GJ(ρ_3)	0.0626	0.0682	0.0540	0.0459	0.1111	**0.0684**	80%
	(0.0031)	(0.0037)	(0.0038)	(0.0028)	(0.0074)	(0.0020)	
GJ(ρ_4)	0.0573	0.0575	0.1751	0.0476	0.1103	**0.0896**	40%
	(0.0037)	(0.0055)	(0.0107)	(0.0034)	(0.0091)	(0.0032)	
NGJ	0.0918	0.1357	0.2391	0.1018	0.1723	**0.1481**	55%
	(0.0062)	(0.0081)	(0.0143)	(0.0066)	(0.0128)	(0.0045)	
NGJ(ρ_3)	0.1026	0.1922	0.3709	0.1096	0.2434	**0.2038**	50%
	(0.0069)	(0.0112)	(0.0185)	(0.0073)	(0.0158)	(0.0057)	
NGJ(ρ_4)	0.0435	0.0665	0.1864	0.0461	0.1109	**0.0907**	90%
	(0.0026)	(0.0043)	(0.0122)	(0.0032)	(0.0067)	(0.0030)	
Peng	0.0695	0.0408	0.0605	0.0399	0.0573	**0.0536**	90%
	(0.0028)	(0.0021)	(0.0037)	(0.0020)	(0.0047)	(0.0014)	
ML	0.0387	0.0172	0.2119	0.0242	0.0641	**0.0712**	90%
	(0.0018)	(0.0013)	(0.0070)	(0.0015)	(0.0046)	(0.0018)	
ML(ρ_3)	0.0516	0.0264	0.0920	0.0333	0.0738	**0.0554**	95%
	(0.0023)	(0.0015)	(0.0045)	(0.0019)	(0.0047)	(0.0015)	
ML(ρ_4)	0.0475	0.0686	0.3107	0.0460	0.1293	**0.1204**	45%
	(0.0050)	(0.0056)	(0.0137)	(0.0030)	(0.0106)	(0.0038)	
GPD Bayes	0.1784	0.2163	0.2671	0.1447	0.2742	**0.2161**	75%
Mean	(0.0216)	(0.0066)	(0.0043)	(0.0134)	(0.0285)	(0.0078)	
GPD MLE	0.8431	0.9178	0.9670	0.8258	0.8391	**0.8786**	65%
	(0.0305)	(0.0100)	(0.0043)	(0.0168)	(0.0366)	(0.0103)	
Zipf	0.8317	0.8752	0.9511	0.7933	0.8427	**0.8588**	90%
	(0.0257)	(0.0163)	(0.0095)	(0.0227)	(0.0321)	(0.0101)	
ERM	0.8431	0.9430	0.9834	0.7970	0.8311	**0.8795**	70%
	(0.0340)	(0.0091)	(0.0047)	(0.0211)	(0.0372)	(0.0111)	

Table 4.7.4 Simulation results for $n = 100$ and EVI group $\gamma = 0.1$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	***t***	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	1.4283	0.7448	2.3898	0.9574	0.5890	0.6741	**1.1306**	
	(0.0681)	(0.0411)	(0.1262)	(0.0492)	(0.0391)	(0.0458)	(0.0280)	
Hill	1.8662	1.8923	5.9337	1.8646	2.6733	3.4496	**2.9466**	15%
	(0.1203)	(0.1394)	(0.3370)	(0.1348)	(0.2217)	(0.2392)	(0.0869)	
Moment	3.7192	4.1169	5.5957	3.0957	4.3485	2.8550	**3.9552**	55%
	(0.2419)	(0.2617)	(0.2995)	(0.2050)	(0.2524)	(0.1703)	(0.0988)	
γ_2	1.5416	1.3529	4.5067	1.4944	1.7948	2.6124	**2.2172**	30%
	(0.1042)	(0.1115)	(0.2280)	(0.1151)	(0.1679)	(0.2033)	(0.0663)	
γ_3	1.5370	1.4830	4.5642	1.5087	1.9552	2.6743	**2.2870**	20%
	(0.0936)	(0.1107)	(0.2528)	(0.1046)	(0.1558)	(0.1935)	(0.0662)	
GJ	1.2745	0.5941	3.5981	0.9419	0.5401	1.6469	**1.4326**	85%
	(0.0585)	(0.0381)	(0.1347)	(0.0545)	(0.0395)	(0.1336)	(0.0355)	
GJ(ρ_3)	1.6849	1.7072	1.3364	1.4661	2.0952	2.7478	**1.8396**	75%
	(0.0853)	(0.0888)	(0.1029)	(0.0862)	(0.1254)	(0.1862)	(0.0482)	
GJ(ρ_4)	1.4499	1.2247	4.4497	1.4265	1.5392	2.5838	**2.1123**	45%
	(0.0896)	(0.0991)	(0.2152)	(0.0958)	(0.1440)	(0.1977)	(0.0608)	
NGJ	2.6464	3.7717	5.8297	2.6432	4.6899	3.9419	**3.9205**	55%
	(0.1542)	(0.2322)	(0.3573)	(0.1694)	(0.2601)	(0.2618)	(0.1014)	
NGJ(ρ_3)	3.4954	5.6054	9.1119	3.4839	7.5365	5.8670	**5.8500**	40%
	(0.2101)	(0.3262)	(0.5616)	(0.2127)	(0.3703)	(0.3265)	(0.1448)	
NGJ(ρ_4)	1.2261	1.6909	4.4684	1.2623	3.1689	2.7988	**2.4359**	90%
	(0.0743)	(0.1217)	(0.2935)	(0.0803)	(0.1950)	(0.1847)	(0.0717)	
Peng	1.7593	1.0738	1.4943	1.1095	0.6534	1.4915	**1.2636**	90%
	(0.0735)	(0.0489)	(0.0921)	(0.0578)	(0.0350)	(0.1189)	(0.0312)	
ML	0.9689	0.5792	4.7528	0.7529	0.7966	1.9102	**1.6268**	85%
	(0.0506)	(0.0452)	(0.1701)	(0.0470)	(0.0535)	(0.1286)	(0.0391)	
ML(ρ_3)	1.3287	0.7052	2.2197	0.9863	0.4065	1.9168	**1.2605**	95%
	(0.0609)	(0.0394)	(0.1034)	(0.0562)	(0.0263)	(0.1310)	(0.0320)	
ML(ρ_4)	1.4089	1.6897	6.5061	1.7768	2.3594	2.8941	**2.7725**	40%
	(0.0878)	(0.1326)	(0.3673)	(0.1873)	(0.1758)	(0.2012)	(0.0861)	
GPD Bayes	3.8625	4.6057	3.1705	3.4118	4.7157	3.5051	**3.8785**	65%
Mean	(0.1948)	(0.2069)	(0.1887)	(0.1783)	(0.2088)	(0.2132)	(0.0812)	
GPD MLE	5.4850	6.9678	3.7436	4.6071	7.0908	3.5930	**5.2479**	65%
	(0.2966)	(0.3294)	(0.2319)	(0.2629)	(0.3226)	(0.2188)	(0.1144)	
Zipf	4.5200	3.9981	7.5759	4.0056	4.1886	4.3027	**4.7651**	75%
	(0.2707)	(0.2311)	(0.3727)	(0.2209)	(0.2584)	(0.2495)	(0.1110)	
ERM	4.7620	5.7071	3.6227	4.1074	5.8796	3.9502	**4.6715**	65%
	(0.2720)	(0.2913)	(0.2236)	(0.2403)	(0.2977)	(0.2424)	(0.1072)	

Table 4.7.5 Simulation results for $n = 100$ and EVI group $\gamma = 0.5$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	***t***	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	5.9039	2.9743	8.1449	4.3039	6.0318	2.4120	**4.9618**	
	(0.2499)	(0.1608)	(0.4228)	(0.2047)	(0.2482)	(0.1645)	(0.1051)	
Hill	5.4683	7.4903	28.5929	4.9829	4.9002	14.3068	**10.9569**	20%
	(0.3317)	(0.5468)	(1.2960)	(0.3523)	(0.3116)	(0.8638)	(0.2914)	
Moment	6.9076	7.2902	25.8581	6.0729	7.4997	10.5233	**10.6920**	35%
	(0.5165)	(0.4579)	(1.1657)	(0.3913)	(0.5114)	(0.6081)	(0.2698)	
γ_2	5.7536	5.7083	19.6532	5.0455	5.5266	10.8769	**8.7607**	30%
	(0.5492)	(0.4890)	(1.0396)	(0.3997)	(0.3813)	(0.7758)	(0.2650)	
γ_3	5.7194	6.1624	19.0306	5.1445	5.5674	11.4328	**8.8429**	20%
	(0.3787)	(0.4757)	(1.0221)	(0.3631)	(0.3457)	(0.7576)	(0.2494)	
GJ	5.0917	2.3226	14.7469	3.6126	5.2816	7.0629	**6.3531**	85%
	(0.2451)	(0.1538)	(0.5670)	(0.2056)	(0.2120)	(0.5355)	(0.1471)	
GJ(ρ_3)	6.2286	6.6766	5.6291	5.4773	9.5706	11.2403	**7.4704**	75%
	(0.3296)	(0.3785)	(0.4176)	(0.3111)	(0.4047)	(0.7594)	(0.1873)	
GJ(ρ_4)	5.3466	5.2329	19.1291	4.8628	5.3469	11.0499	**8.4947**	45%
	(0.4172)	(0.4585)	(0.9888)	(0.3423)	(0.3548)	(0.8567)	(0.2549)	
NGJ	10.6260	13.3268	22.9163	9.5651	11.2554	18.7724	**14.4103**	55%
	(0.6534)	(0.8744)	(1.3463)	(0.5728)	(0.6418)	(1.1989)	(0.3794)	
NGJ(ρ_3)	12.2684	20.0019	34.3950	11.5140	16.2185	25.7340	**20.0220**	45%
	(0.7584)	(1.1466)	(1.6310)	(0.7023)	(0.9156)	(1.7754)	(0.5009)	
NGJ(ρ_4)	4.1641	6.5884	18.1133	4.8203	8.5856	12.4997	**9.1285**	90%
	(0.2595)	(0.4564)	(1.1282)	(0.3000)	(0.4753)	(1.0418)	(0.2862)	
Peng	6.1287	3.7438	7.5578	4.2322	7.6982	7.0189	**6.0633**	85%
	(0.2847)	(0.2010)	(0.4272)	(0.2290)	(0.2913)	(0.5333)	(0.1420)	
ML	3.7534	2.1819	19.5279	2.8023	3.6751	8.3703	**6.7185**	85%
	(0.1920)	(0.1559)	(0.6978)	(0.1768)	(0.1684)	(0.6072)	(0.1647)	
ML(ρ_3)	5.0854	2.7664	9.3191	3.8389	6.4751	7.8359	**5.8868**	95%
	(0.2280)	(0.1619)	(0.4470)	(0.2134)	(0.2470)	(0.5403)	(0.1371)	
ML(ρ_4)	4.2159	7.1764	32.5223	4.4752	3.8894	12.1540	**10.7389**	50%
	(0.2995)	(0.5813)	(1.2623)	(0.2883)	(0.2681)	(0.7848)	(0.2785)	
GPD Bayes	7.8206	7.2241	21.4764	6.7124	8.4943	11.4124	**10.5234**	55%
Median	(0.4699)	(0.4746)	(1.1119)	(0.4228)	(0.4964)	(0.6605)	(0.2658)	
GPD MLE	8.6174	7.9518	23.9130	7.4056	9.3280	12.6255	**11.6402**	55%
	(0.5234)	(0.5168)	(1.2553)	(0.4624)	(0.5535)	(0.7359)	(0.2972)	
Zipf	8.4098	9.4529	34.7646	8.6890	9.6914	14.8023	**14.3017**	50%
	(0.5958)	(0.6188)	(1.4736)	(0.5757)	(0.6214)	(0.8442)	(0.3472)	
ERM	8.2047	8.3010	27.0794	7.1256	8.4891	15.4174	**12.4362**	55%
	(0.5036)	(0.5472)	(1.3340)	(0.4536)	(0.5169)	(0.8600)	(0.3138)	

Table 4.7.6 Simulation results for $n = 100$ and EVI group $\gamma = 1$.

Distr.	Burr	Burr	Burr	Fréchet	logΓ	Mean	*k*
ρ	**-2**	**-1**	**-0.5**	**-1**	**0**		
PPD	0.0405	0.0212	0.0833	0.0259	0.0286	**0.0399**	
	(0.0022)	(0.0011)	(0.0041)	(0.0014)	(0.0022)	(0.0011)	
Hill	0.0411	0.0419	0.2156	0.0358	0.1082	**0.0885**	15%
	(0.0031)	(0.0031)	(0.0102)	(0.0027)	(0.0061)	(0.0026)	
Moment	0.7444	0.7925	0.8289	0.7113	0.6688	**0.7492**	35%
	(0.0335)	(0.0223)	(0.0147)	(0.0257)	(0.0293)	(0.0116)	
γ_2	0.0448	0.0413	0.1581	0.0357	0.0892	**0.0738**	25%
	(0.0040)	(0.0038)	(0.0071)	(0.0028)	(0.0065)	(0.0023)	
γ_3	0.0478	0.0428	0.1471	0.0394	0.0941	**0.0743**	15%
	(0.0039)	(0.0035)	(0.0077)	(0.0028)	(0.0065)	(0.0023)	
GJ	0.0373	0.0191	0.1107	0.0253	0.0677	**0.0520**	70%
	(0.0027)	(0.0012)	(0.0042)	(0.0015)	(0.0049)	(0.0014)	
GJ(ρ_3)	0.0425	0.0420	0.0304	0.0292	0.0768	**0.0442**	75%
	(0.0023)	(0.0020)	(0.0019)	(0.0018)	(0.0045)	(0.0012)	
GJ(ρ_4)	0.0460	0.0394	0.1440	0.0356	0.0841	**0.0698**	35%
	(0.0040)	(0.0036)	(0.0066)	(0.0025)	(0.0063)	(0.0022)	
NGJ	0.0593	0.0758	0.1271	0.0545	0.1070	**0.0847**	55%
	(0.0036)	(0.0045)	(0.0081)	(0.0034)	(0.0067)	(0.0025)	
NGJ(ρ_3)	0.0799	0.1148	0.2185	0.0626	0.1254	**0.1202**	40%
	(0.0051)	(0.0067)	(0.0116)	(0.0039)	(0.0070)	(0.0033)	
NGJ(ρ_4)	0.0229	0.0290	0.1176	0.0244	0.0957	**0.0579**	90%
	(0.0013)	(0.0019)	(0.0064)	(0.0016)	(0.0047)	(0.0017)	
Peng	0.0563	0.0298	0.0546	0.0271	0.0469	**0.0429**	90%
	(0.0020)	(0.0013)	(0.0025)	(0.0013)	(0.0033)	(0.0010)	
ML	0.0271	0.0157	0.1472	0.0200	0.0780	**0.0576**	70%
	(0.0015)	(0.0010)	(0.0052)	(0.0013)	(0.0046)	(0.0015)	
ML(ρ_3)	0.0312	0.0197	0.0742	0.0223	0.0802	**0.0455**	80%
	(0.0015)	(0.0011)	(0.0033)	(0.0014)	(0.0043)	(0.0012)	
ML(ρ_4)	0.0390	0.0459	0.1930	0.0373	0.0975	**0.0826**	30%
	(0.0028)	(0.0036)	(0.0091)	(0.0026)	(0.0054)	(0.0024)	
GPD Bayes	0.2599	0.2562	0.3466	0.1825	0.2357	**0.2562**	60%
Mean	(0.0295)	(0.0061)	(0.0055)	(0.0101)	(0.0241)	(0.0081)	
GPD MLE	0.7621	0.8330	0.9330	0.7385	0.6957	**0.7925**	50%
	(0.0328)	(0.0141)	(0.0079)	(0.0188)	(0.0312)	(0.0103)	
Zipf	0.7713	0.8496	0.9211	0.7354	0.6665	**0.7888**	80%
	(0.0297)	(0.0143)	(0.0101)	(0.0212)	(0.0279)	(0.0098)	
ERM	0.7662	0.8589	0.9584	0.7250	0.6798	**0.7977**	55%
	(0.0379)	(0.0136)	(0.0075)	(0.0193)	(0.0348)	(0.0114)	

Table 4.7.7 Simulation results for $n = 200$ and EVI group $\gamma = 0.1$.

Distr.	Burr	Burr	Burr	Fréchet	*t*	logΓ	Mean	*k*
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	0.9399	0.5805	1.6694	0.7253	0.4038	0.5613	**0.8134**	
	(0.0437)	(0.0329)	(0.0852)	(0.0380)	(0.0267)	(0.0482)	(0.0202)	
Hill	0.7816	1.1691	4.9751	0.9947	1.3456	2.4953	**1.9602**	15%
	(0.0497)	(0.0891)	(0.2335)	(0.0763)	(0.0902)	(0.1544)	(0.0534)	
Moment	1.6761	2.3184	4.2607	1.7625	2.3053	1.9017	**2.3708**	50%
	(0.1187)	(0.1537)	(0.2184)	(0.1174)	(0.1576)	(0.1140)	(0.0617)	
$\boldsymbol{\gamma_2}$	0.8293	1.0606	3.6934	0.9378	1.1396	2.0233	**1.6140**	25%
	(0.0692)	(0.0824)	(0.1746)	(0.0824)	(0.0881)	(0.1598)	(0.0478)	
$\boldsymbol{\gamma_3}$	0.8755	1.1683	3.4595	1.0385	1.1895	2.1148	**1.6410**	15%
	(0.0613)	(0.0851)	(0.1877)	(0.0786)	(0.0877)	(0.1539)	(0.0482)	
GJ	0.7289	0.5084	2.6745	0.6854	0.3810	1.4671	**1.0742**	70%
	(0.0456)	(0.0342)	(0.1083)	(0.0533)	(0.0316)	(0.1312)	(0.0316)	
GJ($\boldsymbol{\rho_3}$)	0.8599	1.0026	0.8396	0.7765	1.2640	1.7117	**1.0757**	70%
	(0.0513)	(0.0516)	(0.0600)	(0.0489)	(0.0563)	(0.1152)	(0.0278)	
GJ($\boldsymbol{\rho_4}$)	0.8337	1.0280	3.3202	0.9558	1.0215	1.9802	**1.5232**	35%
	(0.0603)	(0.0817)	(0.1620)	(0.0744)	(0.0848)	(0.1908)	(0.0488)	
NGJ	1.4211	1.9975	2.8303	1.6111	2.5396	2.6500	**2.1749**	50%
	(0.0856)	(0.1261)	(0.1691)	(0.1039)	(0.1511)	(0.1876)	(0.0579)	
NGJ($\boldsymbol{\rho_3}$)	1.6162	3.0641	4.9660	1.7513	4.2966	3.1121	**3.1344**	40%
	(0.0959)	(0.1769)	(0.2747)	(0.1124)	(0.2344)	(0.1878)	(0.0780)	
NGJ($\boldsymbol{\rho_4}$)	0.5407	0.7346	2.6929	0.6003	1.5289	1.8997	**1.3328**	90%
	(0.0431)	(0.0494)	(0.1574)	(0.0363)	(0.0947)	(0.0986)	(0.0369)	
Peng	0.9921	0.5639	1.6264	0.6637	0.3620	1.0941	**0.8837**	85%
	(0.0458)	(0.0296)	(0.0766)	(0.0366)	(0.0213)	(0.0869)	(0.0225)	
ML	0.5685	0.4538	3.4677	0.5430	0.4533	1.6147	**1.1835**	70%
	(0.0340)	(0.0325)	(0.1275)	(0.0402)	(0.0339)	(0.1053)	(0.0300)	
ML($\boldsymbol{\rho_3}$)	0.7433	0.4623	1.8048	0.5380	0.2868	1.6320	**0.9112**	85%
	(0.0383)	(0.0261)	(0.0824)	(0.0321)	(0.0192)	(0.0914)	(0.0228)	
ML($\boldsymbol{\rho_4}$)	0.7873	1.1146	4.4790	0.9854	1.2656	2.5007	**1.8554**	30%
	(0.0511)	(0.0801)	(0.2185)	(0.0747)	(0.0858)	(0.1723)	(0.0525)	
GPD Bayes	2.2866	3.3731	1.9231	2.3493	3.8621	2.1205	**2.6524**	60%
Mean	(0.1429)	(0.1745)	(0.1229)	(0.1345)	(0.1886)	(0.1385)	(0.0621)	
GPD MLE	2.8404	4.2245	2.0173	2.8162	4.8304	2.0313	**3.1267**	60%
	(0.1747)	(0.2259)	(0.1294)	(0.1636)	(0.2368)	(0.1336)	(0.0743)	
Zipf	2.0795	2.5999	5.5988	2.2682	2.2619	2.7033	**2.9186**	70%
	(0.1338)	(0.1609)	(0.2708)	(0.1358)	(0.1362)	(0.1676)	(0.0711)	
ERM	2.4008	3.6734	2.0508	2.5100	4.1030	2.2563	**2.8324**	60%
	(0.1565)	(0.2086)	(0.1326)	(0.1478)	(0.2193)	(0.1502)	(0.0703)	

Table 4.7.8 Simulation results for $n = 200$ and EVI group $\gamma = 0.5$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	***t***	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	4.3406	2.3573	5.9909	2.9867	7.2427	1.8962	**3.6134**	
	(0.1824)	(0.1260)	(0.3539)	(0.1754)	(0.1772)	(0.1386)	(0.0843)	
Hill	3.6349	4.7419	19.1134	3.9936	3.3797	10.0345	**7.4830**	15%
	(0.2416)	(0.3861)	(1.0070)	(0.3240)	(0.2183)	(0.5732)	(0.2175)	
Moment	4.1818	5.0987	15.2054	4.3899	4.8296	7.4230	**6.8547**	25%
	(0.2853)	(0.3877)	(0.7978)	(0.2841)	(0.3325)	(0.4563)	(0.1877)	
$\boldsymbol{\gamma_2}$	3.7034	3.9738	14.3528	3.8425	3.9880	8.0753	**6.3226**	25%
	(0.2905)	(0.3677)	(0.7827)	(0.4119)	(0.3443)	(0.5987)	(0.2027)	
$\boldsymbol{\gamma_3}$	2.9603	3.8365	16.7721	3.2986	3.0279	8.1090	**6.3341**	20%
	(0.2039)	(0.3069)	(0.7972)	(0.2961)	(0.2038)	(0.5141)	(0.1799)	
GJ	3.2274	1.8869	10.4231	2.6402	3.1184	5.9138	**4.5350**	70%
	(0.1862)	(0.1438)	(0.4532)	(0.2517)	(0.1600)	(0.4446)	(0.1233)	
GJ(ρ_3)	3.7764	3.9332	3.2242	2.9887	5.5377	6.9346	**4.3991**	70%
	(0.2032)	(0.2072)	(0.2741)	(0.2084)	(0.2605)	(0.4193)	(0.1114)	
GJ(ρ_4)	3.2843	3.5365	15.0900	3.4209	3.4368	7.4522	**6.0368**	40%
	(0.2351)	(0.3524)	(0.7082)	(0.3259)	(0.2882)	(0.5404)	(0.1797)	
NGJ	5.7655	7.6132	13.7484	5.5886	6.2583	10.4433	**8.2362**	55%
	(0.3575)	(0.4417)	(0.8154)	(0.3822)	(0.3554)	(0.6774)	(0.2185)	
NGJ(ρ_3)	6.5261	12.7553	25.4236	6.3689	9.0201	12.1441	**12.0397**	45%
	(0.4204)	(0.7360)	(1.2815)	(0.4394)	(0.5377)	(0.6975)	(0.3041)	
NGJ(ρ_4)	2.3243	2.8976	11.2377	2.4126	4.4176	7.8587	**5.1914**	90%
	(0.1394)	(0.1865)	(0.6671)	(0.1500)	(0.2313)	(0.4182)	(0.1443)	
Peng	3.6814	2.1202	7.0538	2.4689	4.4137	5.1719	**4.1516**	80%
	(0.1745)	(0.1139)	(0.3530)	(0.2070)	(0.1875)	(0.3846)	(0.1046)	
ML	2.4659	1.5850	13.4792	2.0262	2.3507	6.6200	**4.7545**	70%
	(0.1348)	(0.1103)	(0.5183)	(0.1392)	(0.1256)	(0.4298)	(0.1201)	
ML(ρ_3)	2.8347	2.0451	6.4186	2.1496	3.7752	7.0136	**4.0395**	75%
	(0.1474)	(0.1122)	(0.3476)	(0.1352)	(0.1722)	(0.4044)	(0.1009)	
ML(ρ_4)	2.9782	4.4128	19.8918	3.4508	2.7003	8.8661	**7.0500**	35%
	(0.2098)	(0.3598)	(0.9409)	(0.2513)	(0.1824)	(0.5044)	(0.1979)	
GPD Bayes	4.6048	4.7804	14.7892	4.6308	4.5203	7.8255	**6.8585**	45%
Median	(0.3074)	(0.3058)	(0.8370)	(0.3110)	(0.2893)	(0.4742)	(0.1896)	
GPD MLE	4.8835	5.0284	15.7310	4.9073	4.8036	8.2798	**7.2723**	45%
	(0.3263)	(0.3170)	(0.8946)	(0.3283)	(0.3059)	(0.5068)	(0.2018)	
Zipf	5.6826	6.3995	20.1298	5.7355	5.9705	10.3883	**9.0510**	40%
	(0.3939)	(0.4299)	(1.0101)	(0.3855)	(0.3482)	(0.6369)	(0.2378)	
ERM	4.2752	4.5071	18.5679	4.4277	4.2702	9.2798	**7.5546**	50%
	(0.2869)	(0.3082)	(0.9491)	(0.3130)	(0.2740)	(0.5327)	(0.2065)	

Table 4.7.9 Simulation results for $n = 200$ and EVI group $\gamma = 1$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**0**		
PPD	0.0247	0.0130	0.0517	0.0168	0.0194	**0.0251**	
	(0.0019)	(0.0011)	(0.0038)	(0.0013)	(0.0023)	(0.0010)	
Hill	0.0201	0.0242	0.1110	0.0201	0.0676	**0.0486**	10%
	(0.0021)	(0.0025)	(0.0080)	(0.0022)	(0.0064)	(0.0022)	
Moment	0.5472	0.5714	0.7827	0.6060	0.5167	**0.6048**	25%
	(0.0381)	(0.0307)	(0.0249)	(0.0376)	(0.0380)	(0.0153)	
$\boldsymbol{\gamma_2}$	0.0258	0.0230	0.0777	0.0277	0.0577	**0.0424**	15%
	(0.0025)	(0.0027)	(0.0063)	(0.0037)	(0.0061)	(0.0020)	
$\boldsymbol{\gamma_3}$	0.0244	0.0237	0.0792	0.0253	0.0597	**0.0425**	10%
	(0.0022)	(0.0025)	(0.0066)	(0.0030)	(0.0061)	(0.0020)	
GJ	0.0182	0.0131	0.0619	0.0179	0.0475	**0.0317**	50%
	(0.0016)	(0.0015)	(0.0041)	(0.0019)	(0.0045)	(0.0014)	
GJ(ρ_3)	0.0199	0.0194	0.0163	0.0138	0.0533	**0.0245**	65%
	(0.0018)	(0.0015)	(0.0017)	(0.0014)	(0.0041)	(0.0010)	
GJ(ρ_4)	0.0227	0.0209	0.0796	0.0226	0.0537	**0.0399**	25%
	(0.0021)	(0.0025)	(0.0058)	(0.0027)	(0.0054)	(0.0018)	
NGJ	0.0240	0.0391	0.0531	0.0235	0.0557	**0.0391**	55%
	(0.0025)	(0.0036)	(0.0045)	(0.0024)	(0.0051)	(0.0017)	
NGJ(ρ_3)	0.0328	0.0474	0.0885	0.0280	0.0681	**0.0530**	35%
	(0.0031)	(0.0049)	(0.0085)	(0.0027)	(0.0058)	(0.0024)	
NGJ(ρ_4)	0.0215	0.0285	0.0310	0.0104	0.0647	**0.0312**	95%
	(0.0016)	(0.0024)	(0.0028)	(0.0011)	(0.0042)	(0.0012)	
Peng	0.0235	0.0108	0.0634	0.0122	0.0336	**0.0287**	80%
	(0.0016)	(0.0008)	(0.0030)	(0.0010)	(0.0029)	(0.0009)	
ML	0.0141	0.0112	0.0817	0.0134	0.0541	**0.0349**	50%
	(0.0013)	(0.0013)	(0.0048)	(0.0012)	(0.0042)	(0.0013)	
ML(ρ_3)	0.0155	0.0125	0.0412	0.0132	0.0585	**0.0282**	55%
	(0.0015)	(0.0011)	(0.0032)	(0.0013)	(0.0042)	(0.0011)	
ML(ρ_4)	0.0197	0.0240	0.1024	0.0199	0.0638	**0.0460**	20%
	(0.0020)	(0.0025)	(0.0076)	(0.0020)	(0.0059)	(0.0021)	
GPD Bayes	0.3102	0.1941	0.3097	0.2119	0.2862	**0.2624**	35%
Mean	(0.0430)	(0.0117)	(0.0106)	(0.0254)	(0.0467)	(0.0140)	
GPD MLE	0.5484	0.7145	0.8836	0.5355	0.5845	**0.6533**	35%
	(0.0416)	(0.0260)	(0.0180)	(0.0310)	(0.0422)	(0.0148)	
Zipf	0.5967	0.6114	0.7987	0.6422	0.5393	**0.6377**	50%
	(0.0370)	(0.0280)	(0.0238)	(0.0377)	(0.0315)	(0.0143)	
ERM	0.5035	0.7284	0.9233	0.5212	0.5499	**0.6453**	40%
	(0.0357)	(0.0262)	(0.0160)	(0.0300)	(0.0384)	(0.0135)	

Table 4.7.10 Simulation results for $n = 500$ and EVI group $\gamma = 0.1$.

Distr.	Burr	Burr	Burr	Fréchet	*t*	logΓ	Mean	*k*
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	0.6359	0.3950	1.1387	0.4377	0.2722	0.3691	**0.5414**	
	(0.0476)	(0.0392)	(0.0834)	(0.0333)	(0.0252)	(0.0364)	(0.0196)	
Hill	0.5600	0.6645	2.7525	0.5590	0.6288	1.8252	**1.1650**	10%
	(0.0569)	(0.0657)	(0.1651)	(0.0619)	(0.0676)	(0.1625)	(0.0440)	
Moment	0.7924	1.0760	2.4470	0.6942	0.9251	1.3132	**1.2080**	40%
	(0.0757)	(0.1144)	(0.1714)	(0.0738)	(0.0990)	(0.1105)	(0.0458)	
$\boldsymbol{\gamma_2}$	0.4702	0.5903	2.4487	0.4886	0.4583	1.4962	**0.9920**	20%
	(0.0436)	(0.0572)	(0.1425)	(0.0573)	(0.0506)	(0.1494)	(0.0386)	
$\boldsymbol{\gamma_3}$	0.4139	0.5474	2.6682	0.4122	0.4925	1.5031	**1.0062**	15%
	(0.0378)	(0.0509)	(0.1454)	(0.0453)	(0.0484)	(0.1253)	(0.0354)	
GJ	0.4342	0.3296	1.7393	0.3506	0.1947	1.0867	**0.6892**	55%
	(0.0389)	(0.0329)	(0.1035)	(0.0376)	(0.0196)	(0.1014)	(0.0266)	
GJ($\boldsymbol{\rho_3}$)	0.4824	0.5545	0.4751	0.3361	0.7165	1.2499	**0.6358**	60%
	(0.0406)	(0.0482)	(0.0590)	(0.0329)	(0.0492)	(0.0929)	(0.0233)	
GJ($\boldsymbol{\rho_4}$)	0.5514	0.6266	1.9635	0.5535	0.4440	1.3855	**0.9208**	25%
	(0.0501)	(0.0588)	(0.1328)	(0.0610)	(0.0532)	(0.1454)	(0.0377)	
NGJ	0.7066	0.8503	1.2702	0.5332	1.1707	1.4649	**0.9993**	45%
	(0.0728)	(0.0976)	(0.1267)	(0.0605)	(0.0978)	(0.1324)	(0.0414)	
NGJ($\boldsymbol{\rho_3}$)	0.9482	1.4996	2.1498	0.6287	2.2995	1.5989	**1.5208**	35%
	(0.0845)	(0.1540)	(0.1944)	(0.0639)	(0.2006)	(0.1245)	(0.0597)	
NGJ($\boldsymbol{\rho_4}$)	0.2731	0.2751	2.0862	0.2489	0.7176	1.7866	**0.8979**	90%
	(0.0265)	(0.0258)	(0.1515)	(0.0257)	(0.0715)	(0.0996)	(0.0333)	
Peng	0.4126	0.2716	1.5227	0.3193	0.1828	1.0309	**0.6233**	65%
	(0.0347)	(0.0277)	(0.0849)	(0.0279)	(0.0178)	(0.0858)	(0.0221)	
ML	0.4021	0.3175	2.0336	0.3015	0.2208	1.2422	**0.7529**	50%
	(0.0388)	(0.0376)	(0.1168)	(0.0332)	(0.0217)	(0.0927)	(0.0273)	
ML($\boldsymbol{\rho_3}$)	0.3936	0.3405	1.0655	0.2533	0.2603	1.4436	**0.6261**	60%
	(0.0341)	(0.0326)	(0.0795)	(0.0241)	(0.0227)	(0.0910)	(0.0223)	
ML($\boldsymbol{\rho_4}$)	0.5536	0.6619	2.5082	0.4976	0.5905	1.6338	**1.0742**	20%
	(0.0513)	(0.0673)	(0.1566)	(0.0574)	(0.0652)	(0.1329)	(0.0398)	
GPD Bayes	1.2848	1.8202	0.9643	1.1156	2.2364	1.2360	**1.4429**	45%
Mean	(0.1082)	(0.1749)	(0.1047)	(0.1192)	(0.1813)	(0.1210)	(0.0565)	
GPD MLE	1.4021	2.0946	0.9578	1.2519	2.6220	1.0962	**1.5708**	45%
	(0.1167)	(0.1926)	(0.1057)	(0.1288)	(0.2004)	(0.1127)	(0.0604)	
Zipf	1.0400	1.2982	3.0392	1.0654	0.9966	1.5786	**1.5030**	60%
	(0.0938)	(0.1346)	(0.2055)	(0.1035)	(0.0920)	(0.1533)	(0.0557)	
ERM	1.3020	1.8699	0.9653	1.1753	2.2992	1.2660	**1.4796**	45%
	(0.1092)	(0.1804)	(0.1018)	(0.1230)	(0.1896)	(0.1260)	(0.0582)	

Table 4.7.11 Simulation results for $n = 500$ and EVI group $\gamma = 0.5$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	***t***	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	2.5581	1.5064	4.0734	1.6857	2.6006	1.5700	**2.3324**	
	(0.1631)	(0.1408)	(0.2897)	(0.1251)	(0.1746)	(0.1965)	(0.0773)	
Hill	1.8834	2.6550	11.1059	1.8062	1.9801	7.0791	**4.4183**	10%
	(0.1949)	(0.2563)	(0.6695)	(0.1956)	(0.1911)	(0.6542)	(0.1712)	
Moment	2.5966	3.0352	7.4692	2.6807	2.6457	5.3605	**3.9647**	15%
	(0.2692)	(0.3418)	(0.5447)	(0.3683)	(0.2554)	(0.4958)	(0.1610)	
γ_2	2.4554	2.4821	7.1498	2.2295	2.5239	5.5324	**3.7288**	15%
	(0.3066)	(0.3219)	(0.5287)	(0.2574)	(0.3756)	(0.5362)	(0.1643)	
γ_3	2.2787	2.5850	7.3156	2.1414	2.3449	5.6868	**3.7254**	10%
	(0.2540)	(0.2913)	(0.5302)	(0.2400)	(0.2748)	(0.5457)	(0.1547)	
GJ	1.6617	1.1969	6.9515	1.2838	1.7014	4.5754	**2.8951**	55%
	(0.1707)	(0.1358)	(0.3659)	(0.1315)	(0.1497)	(0.4452)	(0.1079)	
GJ(ρ_3)	1.8492	1.9926	2.0942	1.3077	2.5900	5.1162	**2.4917**	55%
	(0.1712)	(0.1911)	(0.1885)	(0.1268)	(0.2058)	(0.4266)	(0.0974)	
GJ(ρ_4)	2.1964	2.2153	7.7075	1.8554	2.3262	5.2211	**3.5870**	25%
	(0.2639)	(0.2981)	(0.5111)	(0.2035)	(0.3651)	(0.5045)	(0.1536)	
NGJ	2.2403	3.2357	4.6978	2.3453	3.0733	6.1870	**3.6299**	50%
	(0.1895)	(0.2822)	(0.4374)	(0.2271)	(0.2831)	(0.6603)	(0.1559)	
NGJ(ρ_3)	2.6883	5.0118	8.5618	2.3286	4.1177	7.3749	**5.0138**	40%
	(0.2446)	(0.5163)	(0.9269)	(0.2153)	(0.3938)	(0.6877)	(0.2273)	
NGJ(ρ_4)	0.8518	1.0182	9.2549	1.0592	2.8427	6.7942	**3.6369**	90%
	(0.0709)	(0.0884)	(0.6428)	(0.1073)	(0.1940)	(0.3956)	(0.1325)	
Peng	1.7208	1.3190	5.5144	1.3411	1.8240	4.5235	**2.7071**	55%
	(0.1722)	(0.1439)	(0.3374)	(0.1338)	(0.1575)	(0.4450)	(0.1061)	
ML	1.2434	1.0380	9.2831	1.0620	1.3301	5.4462	**3.2338**	55%
	(0.1114)	(0.1008)	(0.4306)	(0.0952)	(0.1088)	(0.4746)	(0.1123)	
ML(ρ_3)	1.3990	1.3610	4.1562	1.1250	1.9599	5.9737	**2.6625**	55%
	(0.1213)	(0.1280)	(0.2765)	(0.0973)	(0.1525)	(0.4317)	(0.0953)	
ML(ρ_4)	1.8997	2.5524	10.1930	1.8380	2.0420	7.4573	**4.3304**	20%
	(0.2153)	(0.2610)	(0.6141)	(0.1927)	(0.1976)	(0.7428)	(0.1763)	
GPD Bayes	2.0826	2.2481	9.4661	2.0285	2.1664	6.0247	**4.0027**	35%
Median	(0.1864)	(0.2274)	(0.5981)	(0.2199)	(0.1843)	(0.5896)	(0.1558)	
GPD MLE	2.1564	2.3004	9.7196	2.0757	2.2500	6.1925	**4.1158**	35%
	(0.1911)	(0.2340)	(0.6126)	(0.2224)	(0.1927)	(0.6107)	(0.1604)	
Zipf	2.7515	3.1666	10.5126	2.9155	2.8037	6.4112	**4.7602**	30%
	(0.2834)	(0.3134)	(0.7388)	(0.3447)	(0.2939)	(0.5514)	(0.1851)	
ERM	2.1687	2.3950	9.5998	2.1452	2.1827	6.6600	**4.1919**	35%
	(0.2004)	(0.2387)	(0.5991)	(0.2412)	(0.1862)	(0.6266)	(0.1617)	

Table 4.7.12 Simulation results for $n = 500$ and EVI group $\gamma = 1$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**0**		
PPD	0.0167	0.0098	0.0372	0.0133	0.0167	**0.0187**	
	(0.0011)	(0.0009)	(0.0035)	(0.0010)	(0.0017)	(0.0009)	
Hill	0.0247	0.0181	0.0558	0.0260	0.0496	**0.0348**	5%
	(0.0026)	(0.0018)	(0.0051)	(0.0024)	(0.0043)	(0.0016)	
Moment	0.3932	0.4769	0.7594	0.4571	0.4349	**0.5043**	20%
	(0.0289)	(0.0290)	(0.0244)	(0.0316)	(0.0302)	(0.0129)	
γ_2	0.0139	0.0120	0.0629	0.0148	0.0475	**0.0302**	15%
	(0.0013)	(0.0015)	(0.0040)	(0.0015)	(0.0040)	(0.0012)	
γ_3	0.0137	0.0117	0.0623	0.0135	0.0484	**0.0299**	10%
	(0.0013)	(0.0014)	(0.0040)	(0.0013)	(0.0038)	(0.0012)	
GJ	0.0124	0.0089	0.0437	0.0122	0.0410	**0.0236**	40%
	(0.0011)	(0.0010)	(0.0028)	(0.0011)	(0.0034)	(0.0010)	
GJ(ρ_3)	0.0116	0.0144	0.0135	0.0077	0.0503	**0.0195**	60%
	(0.0009)	(0.0010)	(0.0012)	(0.0007)	(0.0032)	(0.0007)	
GJ(ρ_4)	0.0155	0.0121	0.0538	0.0160	0.0445	**0.0284**	20%
	(0.0015)	(0.0015)	(0.0037)	(0.0016)	(0.0039)	(0.0012)	
NGJ	0.0138	0.0229	0.0301	0.0129	0.0487	**0.0257**	50%
	(0.0013)	(0.0021)	(0.0033)	(0.0013)	(0.0044)	(0.0012)	
NGJ(ρ_3)	0.0185	0.0292	0.0494	0.0163	0.0548	**0.0336**	30%
	(0.0016)	(0.0029)	(0.0040)	(0.0014)	(0.0040)	(0.0013)	
NGJ(ρ_4)	0.0144	0.0208	0.0140	0.0057	0.0663	**0.0242**	95%
	(0.0009)	(0.0014)	(0.0013)	(0.0006)	(0.0032)	(0.0008)	
Peng	0.0125	0.0094	0.0383	0.0125	0.0407	**0.0227**	40%
	(0.0011)	(0.0010)	(0.0027)	(0.0011)	(0.0034)	(0.0010)	
ML	0.0098	0.0073	0.0584	0.0090	0.0467	**0.0263**	40%
	(0.0009)	(0.0007)	(0.0034)	(0.0009)	(0.0035)	(0.0010)	
ML(ρ_3)	0.0110	0.0099	0.0279	0.0085	0.0537	**0.0222**	40%
	(0.0009)	(0.0009)	(0.0024)	(0.0008)	(0.0036)	(0.0009)	
ML(ρ_4)	0.0159	0.0131	0.0674	0.0143	0.0516	**0.0325**	15%
	(0.0016)	(0.0017)	(0.0045)	(0.0016)	(0.0044)	(0.0014)	
GPD Bayes	0.2778	0.1858	0.3191	0.2408	0.2436	**0.2534**	25%
Mean	(0.0296)	(0.0163)	(0.0118)	(0.0334)	(0.0334)	(0.0119)	
GPD MLE	0.5004	0.4983	0.7704	0.5213	0.4802	**0.5541**	20%
	(0.0383)	(0.0289)	(0.0230)	(0.0393)	(0.0311)	(0.0146)	
Zipf	0.4591	0.4761	0.7559	0.4760	0.4709	**0.5276**	40%
	(0.0288)	(0.0291)	(0.0246)	(0.0305)	(0.0333)	(0.0131)	
ERM	0.4553	0.5111	0.8211	0.4746	0.4504	**0.5425**	25%
	(0.0332)	(0.0299)	(0.0217)	(0.0366)	(0.0339)	(0.0141)	

Table 4.7.13 Simulation results for $n = 1000$ and EVI group $\gamma = 0.1$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	***t***	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	0.3401	0.1955	0.9762	0.2352	0.2468	0.5327	**0.4211**	
	(0.0242)	(0.0194)	(0.0782)	(0.0200)	(0.0295)	(0.0401)	(0.0166)	
Hill	0.5597	0.5473	1.5635	0.4252	0.5088	1.3871	**0.8320**	5%
	(0.0521)	(0.0758)	(0.1503)	(0.0395)	(0.0533)	(0.1219)	(0.0374)	
Moment	0.5208	0.4990	1.5482	0.4521	0.5518	1.0792	**0.7752**	30%
	(0.0473)	(0.0511)	(0.1221)	(0.0433)	(0.0562)	(0.0769)	(0.0292)	
γ_2	0.3626	0.3370	1.7740	0.2788	0.3445	1.1382	**0.7058**	15%
	(0.0349)	(0.0429)	(0.1300)	(0.0293)	(0.0382)	(0.0873)	(0.0288)	
γ_3	0.3583	0.3376	1.6928	0.2620	0.3500	1.1881	**0.6981**	10%
	(0.0331)	(0.0408)	(0.1237)	(0.0250)	(0.0346)	(0.0858)	(0.0275)	
GJ	0.3394	0.2404	1.1369	0.2392	0.2060	0.9577	**0.5200**	40%
	(0.0316)	(0.0266)	(0.0788)	(0.0227)	(0.0239)	(0.0708)	(0.0197)	
GJ(ρ_3)	0.3012	0.3011	0.3649	0.1593	0.4352	1.2107	**0.4621**	55%
	(0.0223)	(0.0229)	(0.0351)	(0.0159)	(0.0331)	(0.0731)	(0.0158)	
GJ(ρ_4)	0.4064	0.3526	1.5594	0.2990	0.3289	1.0754	**0.6703**	20%
	(0.0388)	(0.0444)	(0.1325)	(0.0300)	(0.0365)	(0.0854)	(0.0291)	
NGJ	0.3381	0.5011	0.5900	0.3082	0.7579	1.0899	**0.5975**	45%
	(0.0326)	(0.0432)	(0.0566)	(0.0311)	(0.0681)	(0.0842)	(0.0229)	
NGJ(ρ_3)	0.5442	0.8402	1.3340	0.3906	1.1897	1.4709	**0.9616**	30%
	(0.0527)	(0.0806)	(0.1222)	(0.0408)	(0.1099)	(0.0977)	(0.0363)	
NGJ(ρ_4)	0.1517	0.1170	2.0730	0.1466	0.3728	1.8497	**0.7851**	90%
	(0.0118)	(0.0128)	(0.1077)	(0.0139)	(0.0323)	(0.0778)	(0.0231)	
Peng	0.2755	0.1855	1.1619	0.1997	0.1637	0.9577	**0.4906**	50%
	(0.0226)	(0.0191)	(0.0714)	(0.0192)	(0.0176)	(0.0679)	(0.0177)	
ML	0.2570	0.1994	1.4206	0.1892	0.1995	1.1621	**0.5713**	40%
	(0.0260)	(0.0213)	(0.0834)	(0.0183)	(0.0254)	(0.0798)	(0.0207)	
ML(ρ_3)	0.2385	0.2017	0.7572	0.1486	0.2012	1.4252	**0.4954**	50%
	(0.0188)	(0.0179)	(0.0558)	(0.0151)	(0.0198)	(0.0832)	(0.0177)	
ML(ρ_4)	0.3934	0.3762	1.7203	0.3179	0.3819	1.3708	**0.7601**	15%
	(0.0391)	(0.0453)	(0.1317)	(0.0295)	(0.0363)	(0.0986)	(0.0302)	
GPD Bayes	0.7310	1.0607	0.5917	0.7194	1.4759	0.9602	**0.9231**	40%
Mean	(0.0674)	(0.0813)	(0.0615)	(0.0680)	(0.1168)	(0.0822)	(0.0333)	
GPD MLE	0.7934	1.2027	0.6024	0.8127	1.6687	0.8691	**0.9915**	40%
	(0.0718)	(0.0898)	(0.0590)	(0.0742)	(0.1260)	(0.0768)	(0.0350)	
Zipf	0.7162	0.6196	2.0337	0.6086	0.5956	1.1376	**0.9519**	50%
	(0.0581)	(0.0622)	(0.1570)	(0.0661)	(0.0590)	(0.0969)	(0.0369)	
ERM	0.7682	1.0733	0.6154	0.7345	1.4841	0.9288	**0.9341**	40%
	(0.0691)	(0.0846)	(0.0618)	(0.0694)	(0.1197)	(0.0772)	(0.0337)	

Table 4.7.14 Simulation results for $n = 1000$ and EVI group $\gamma = 0.5$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	***t***	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	1.5654	0.7953	3.7100	0.9030	7.2427	2.0135	**1.7479**	
	(0.1246)	(0.0747)	(0.2776)	(0.0820)	(0.1215)	(0.1671)	(0.0640)	
Hill	1.2545	1.4613	8.7193	1.2390	1.1177	6.0917	**3.3139**	10%
	(0.1192)	(0.1392)	(0.4903)	(0.1164)	(0.1062)	(0.4048)	(0.1134)	
Moment	1.7133	1.5458	6.5345	1.2978	1.6721	5.1031	**2.9778**	15%
	(0.1721)	(0.1458)	(0.4264)	(0.1335)	(0.2159)	(0.3740)	(0.1102)	
$\boldsymbol{\gamma_2}$	1.5700	1.4882	6.2561	1.2191	1.5913	5.3989	**2.9206**	15%
	(0.1567)	(0.1660)	(0.3963)	(0.1465)	(0.1825)	(0.4420)	(0.1129)	
$\boldsymbol{\gamma_3}$	1.5097	1.4617	6.0806	1.2383	1.4737	5.5190	**2.8805**	10%
	(0.1460)	(0.1478)	(0.4107)	(0.1333)	(0.1611)	(0.4287)	(0.1105)	
GJ	1.3441	0.7917	5.1192	0.9223	1.2503	4.3858	**2.3023**	45%
	(0.1342)	(0.0791)	(0.2678)	(0.1041)	(0.1254)	(0.3389)	(0.0812)	
GJ(ρ_3)	1.3826	1.1565	1.4385	0.7196	1.9022	5.0781	**1.9462**	55%
	(0.1318)	(0.0902)	(0.1303)	(0.0722)	(0.1605)	(0.3128)	(0.0690)	
GJ(ρ_4)	1.7294	1.5490	5.3883	1.3144	1.8190	5.0758	**2.8127**	20%
	(0.1708)	(0.1741)	(0.3694)	(0.1549)	(0.2054)	(0.4220)	(0.1106)	
NGJ	1.6630	1.7745	2.6762	1.3782	1.6768	5.1953	**2.3940**	45%
	(0.1702)	(0.1604)	(0.2689)	(0.1423)	(0.1581)	(0.4078)	(0.0970)	
NGJ(ρ_3)	2.3411	2.8171	4.5966	1.5767	2.2961	5.7305	**3.2263**	35%
	(0.2481)	(0.2594)	(0.3932)	(0.1688)	(0.2525)	(0.3586)	(0.1184)	
NGJ(ρ_4)	1.2149	1.5030	8.1041	1.2582	1.0996	5.5382	**3.1197**	25%
	(0.1129)	(0.1572)	(0.4690)	(0.1259)	(0.1060)	(0.3742)	(0.1086)	
Peng	1.2199	0.6102	5.0788	0.7301	1.2286	4.0652	**2.1555**	55%
	(0.1182)	(0.0547)	(0.2462)	(0.0719)	(0.1126)	(0.2929)	(0.0710)	
ML	1.0412	0.6331	6.7817	0.8184	0.9573	4.6690	**2.4834**	45%
	(0.1119)	(0.0672)	(0.3115)	(0.0849)	(0.0924)	(0.3063)	(0.0788)	
ML(ρ_3)	1.1873	0.8636	2.9959	0.8217	1.3407	5.4686	**2.1130**	45%
	(0.1247)	(0.0758)	(0.2092)	(0.0831)	(0.1287)	(0.3184)	(0.0726)	
ML(ρ_4)	1.6274	1.5641	6.2632	1.5119	1.6544	5.6995	**3.0534**	15%
	(0.1461)	(0.1565)	(0.4371)	(0.1372)	(0.1751)	(0.4202)	(0.1134)	
GPD Bayes	1.9865	1.7296	5.3851	1.7247	1.7800	5.0358	**2.9403**	25%
Median	(0.1940)	(0.1589)	(0.3780)	(0.1687)	(0.2053)	(0.3711)	(0.1073)	
GPD MLE	2.0533	1.7488	5.3849	1.7563	1.8030	4.9906	**2.9562**	25%
	(0.2004)	(0.1620)	(0.3859)	(0.1731)	(0.2084)	(0.3629)	(0.1081)	
Zipf	2.2223	2.2380	7.3444	1.7531	2.4900	6.0941	**3.6903**	25%
	(0.2204)	(0.2031)	(0.5058)	(0.1935)	(0.3096)	(0.4868)	(0.1410)	
ERM	1.9164	1.3797	6.1885	1.3649	1.5595	5.6271	**3.0060**	30%
	(0.1901)	(0.1294)	(0.3928)	(0.1480)	(0.1890)	(0.4012)	(0.1087)	

Table 4.7.15 Simulation results for $n = 1000$ and EVI group $\gamma = 1$.

Distr.	Burr	Burr	Burr	Fréchet	logΓ	Mean	k
ρ	**-2**	**-1**	**-0.5**	**-1**	**0**		
PPD	0.0109	0.0085	0.0263	0.0078	0.0173	**0.0141**	
	(0.0008)	(0.0008)	(0.0025)	(0.0006)	(0.0014)	(0.0006)	
Hill	0.0090	0.0119	0.0404	0.0086	0.0464	**0.0233**	5%
	(0.0008)	(0.0014)	(0.0028)	(0.0008)	(0.0036)	(0.0010)	
Moment	0.2754	0.3595	0.6779	0.3495	0.3176	**0.3960**	15%
	(0.0220)	(0.0258)	(0.0266)	(0.0251)	(0.0277)	(0.0114)	
γ_2	0.0085	0.0098	0.0360	0.0081	0.0460	**0.0217**	10%
	(0.0009)	(0.0009)	(0.0022)	(0.0008)	(0.0045)	(0.0010)	
γ_3	0.0107	0.0122	0.0295	0.0104	0.0455	**0.0217**	5%
	(0.0011)	(0.0012)	(0.0023)	(0.0010)	(0.0047)	(0.0011)	
GJ	0.0085	0.0091	0.0245	0.0079	0.0414	**0.0183**	25%
	(0.0009)	(0.0009)	(0.0018)	(0.0007)	(0.0042)	(0.0010)	
GJ(ρ_3)	0.0045	0.0073	0.0128	0.0038	0.0433	**0.0143**	45%
	(0.0004)	(0.0007)	(0.0011)	(0.0004)	(0.0025)	(0.0006)	
GJ(ρ_4)	0.0082	0.0099	0.0325	0.0083	0.0425	**0.0203**	15%
	(0.0008)	(0.0009)	(0.0021)	(0.0008)	(0.0040)	(0.0009)	
NGJ	0.0083	0.0141	0.0142	0.0075	0.0390	**0.0166**	50%
	(0.0008)	(0.0012)	(0.0013)	(0.0007)	(0.0023)	(0.0006)	
NGJ(ρ_3)	0.0058	0.0159	0.0267	0.0068	0.0471	**0.0205**	35%
	(0.0006)	(0.0016)	(0.0023)	(0.0007)	(0.0028)	(0.0008)	
NGJ(ρ_4)	0.0044	0.0063	0.0203	0.0034	0.0621	**0.0193**	90%
	(0.0004)	(0.0006)	(0.0016)	(0.0003)	(0.0024)	(0.0006)	
Peng	0.0057	0.0056	0.0343	0.0044	0.0377	**0.0175**	40%
	(0.0006)	(0.0005)	(0.0018)	(0.0004)	(0.0026)	(0.0007)	
ML	0.0069	0.0077	0.0336	0.0067	0.0438	**0.0197**	25%
	(0.0008)	(0.0008)	(0.0022)	(0.0006)	(0.0033)	(0.0008)	
ML(ρ_3)	0.0041	0.0059	0.0238	0.0035	0.0457	**0.0166**	40%
	(0.0004)	(0.0006)	(0.0016)	(0.0003)	(0.0023)	(0.0006)	
ML(ρ_4)	0.0098	0.0121	0.0343	0.0089	0.0462	**0.0223**	10%
	(0.0010)	(0.0014)	(0.0025)	(0.0009)	(0.0036)	(0.0010)	
GPD Bayes	0.2190	0.1737	0.2623	0.2243	0.2075	**0.2173**	15%
Mean	(0.0218)	(0.0163)	(0.0124)	(0.0243)	(0.0236)	(0.0090)	
GPD MLE	0.3314	0.3959	0.6894	0.4004	0.3499	**0.4334**	15%
	(0.0253)	(0.0265)	(0.0254)	(0.0278)	(0.0257)	(0.0117)	
Zipf	0.3100	0.4546	0.7279	0.3651	0.3217	**0.4359**	35%
	(0.0238)	(0.0273)	(0.0246)	(0.0249)	(0.0246)	(0.0112)	
ERM	0.3304	0.3758	0.6683	0.3902	0.3357	**0.4201**	15%
	(0.0260)	(0.0271)	(0.0268)	(0.0286)	(0.0266)	(0.0121)	

Table 4.7.16 Simulation results for $n = 2000$ and EVI group $\gamma = 0.1$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	***t***	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	0.2247	0.1601	0.7235	0.1467	0.1578	0.3982	**0.3018**	
	(0.0162)	(0.0152)	(0.0647)	(0.0116)	(0.0163)	(0.0234)	(0.0125)	
Hill	0.2144	0.2814	1.1282	0.3490	0.2479	1.1251	**0.5577**	5%
	(0.0193)	(0.0296)	(0.0955)	(0.0365)	(0.0300)	(0.0765)	(0.0226)	
Moment	0.2383	0.3030	1.0122	0.3359	0.2865	0.8630	**0.5065**	25%
	(0.0228)	(0.0308)	(0.0720)	(0.0343)	(0.0279)	(0.0537)	(0.0179)	
γ_2	0.2001	0.2785	0.9545	0.2934	0.2232	0.9462	**0.4827**	10%
	(0.0176)	(0.0354)	(0.0682)	(0.0289)	(0.0252)	(0.0644)	(0.0181)	
γ_3	0.2518	0.3425	0.7865	0.3802	0.2602	0.9841	**0.5009**	5%
	(0.0220)	(0.0405)	(0.0763)	(0.0368)	(0.0275)	(0.0727)	(0.0206)	
GJ	0.1507	0.1473	0.8958	0.1920	0.1095	0.7843	**0.3799**	35%
	(0.0131)	(0.0156)	(0.0499)	(0.0198)	(0.0117)	(0.0452)	(0.0123)	
GJ(ρ_3)	0.1232	0.1796	0.3532	0.1593	0.2451	0.9143	**0.3291**	40%
	(0.0105)	(0.0162)	(0.0338)	(0.0155)	(0.0194)	(0.0468)	(0.0110)	
GJ(ρ_4)	0.2002	0.2761	0.8815	0.2992	0.2153	0.8933	**0.4609**	15%
	(0.0172)	(0.0357)	(0.0668)	(0.0295)	(0.0247)	(0.0636)	(0.0179)	
NGJ	0.1807	0.2647	0.4615	0.1869	0.4090	0.8766	**0.3966**	40%
	(0.0154)	(0.0238)	(0.0460)	(0.0177)	(0.0320)	(0.0570)	(0.0144)	
NGJ(ρ_3)	0.2197	0.3854	0.6264	0.2850	0.6470	0.9633	**0.5211**	25%
	(0.0233)	(0.0377)	(0.0596)	(0.0286)	(0.0627)	(0.0644)	(0.0200)	
NGJ(ρ_4)	0.1840	0.2555	1.2127	0.2771	0.2758	1.0173	**0.5371**	15%
	(0.0157)	(0.0233)	(0.0901)	(0.0289)	(0.0309)	(0.0708)	(0.0209)	
Peng	0.1529	0.1541	0.7928	0.1953	0.1181	0.7767	**0.3650**	35%
	(0.0134)	(0.0160)	(0.0486)	(0.0200)	(0.0122)	(0.0453)	(0.0122)	
ML	0.1809	0.1679	0.8769	0.2580	0.1578	0.8823	**0.4206**	25%
	(0.0171)	(0.0174)	(0.0556)	(0.0294)	(0.0203)	(0.0567)	(0.0151)	
ML(ρ_3)	0.1204	0.1505	0.5708	0.1401	0.1650	0.9929	**0.3566**	35%
	(0.0113)	(0.0138)	(0.0420)	(0.0133)	(0.0143)	(0.0486)	(0.0116)	
ML(ρ_4)	0.2148	0.2947	0.9865	0.3228	0.2615	1.0531	**0.5222**	10%
	(0.0172)	(0.0277)	(0.0829)	(0.0346)	(0.0312)	(0.0767)	(0.0211)	
GPD Bayes	0.3561	0.6559	0.3194	0.4272	0.9579	0.5935	**0.5517**	35%
Mean	(0.0351)	(0.0547)	(0.0332)	(0.0439)	(0.0638)	(0.0445)	(0.0192)	
GPD MLE	0.3986	0.7493	0.3128	0.4611	1.0835	0.5285	**0.5890**	35%
	(0.0380)	(0.0607)	(0.0324)	(0.0468)	(0.0686)	(0.0411)	(0.0203)	
Zipf	0.3435	0.4110	1.0452	0.4855	0.3402	0.9266	**0.5920**	40%
	(0.0331)	(0.0393)	(0.0808)	(0.0544)	(0.0309)	(0.0646)	(0.0219)	
ERM	0.3707	0.6672	0.3189	0.4558	0.9470	0.5837	**0.5572**	35%
	(0.0355)	(0.0573)	(0.0336)	(0.0464)	(0.0644)	(0.0438)	(0.0197)	

Table 4.7.17 Simulation results for $n = 2000$ and EVI group $\gamma = 0.5$.

Distr.	Burr	Burr	Burr	Fréchet	*t*	logΓ	Mean	*k*
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	1.1343	0.5337	2.9970	0.7780	1.1532	1.5714	**1.3613**	
	(0.0675)	(0.0511)	(0.2597)	(0.0624)	(0.0742)	(0.1208)	(0.0523)	
Hill	0.9969	0.9984	5.0085	1.3636	1.0236	4.1736	**2.2608**	5%
	(0.0965)	(0.0962)	(0.3642)	(0.1671)	(0.1082)	(0.3278)	(0.0910)	
Moment	1.1316	0.8773	4.8986	1.2454	0.8630	3.5689	**2.0975**	10%
	(0.1146)	(0.0861)	(0.3171)	(0.1237)	(0.0939)	(0.2699)	(0.0778)	
γ_2	1.0714	0.8065	4.7376	1.2413	0.8628	3.6337	**2.0589**	10%
	(0.1229)	(0.0958)	(0.3103)	(0.1587)	(0.1218)	(0.2847)	(0.0819)	
γ_3	1.2978	1.0322	4.0275	1.5363	1.1170	3.5883	**2.0999**	5%
	(0.1392)	(0.1077)	(0.3293)	(0.1872)	(0.1456)	(0.3037)	(0.0894)	
GJ	0.8585	0.5216	3.4217	0.9503	0.7781	3.3409	**1.6452**	30%
	(0.0922)	(0.0558)	(0.2002)	(0.1054)	(0.0907)	(0.2595)	(0.0620)	
GJ(ρ_3)	0.5883	0.4944	1.4861	0.5554	0.8336	3.7781	**1.2893**	45%
	(0.0550)	(0.0444)	(0.1177)	(0.0565)	(0.0630)	(0.2265)	(0.0463)	
GJ(ρ_4)	1.0657	0.7875	4.3336	1.2449	0.8678	3.4773	**1.9628**	15%
	(0.1196)	(0.0920)	(0.2956)	(0.1439)	(0.1098)	(0.2856)	(0.0790)	
NGJ	0.8660	1.0187	1.4517	0.7582	1.1983	3.6532	**1.4910**	45%
	(0.0733)	(0.1029)	(0.1593)	(0.0763)	(0.0909)	(0.2417)	(0.0562)	
NGJ(ρ_3)	0.8206	1.1314	2.3588	0.8502	1.2383	4.2551	**1.7757**	30%
	(0.0857)	(0.1078)	(0.2182)	(0.0896)	(0.1135)	(0.2791)	(0.0678)	
NGJ(ρ_4)	0.5702	0.6687	2.2326	0.3946	3.1345	5.5953	**2.0993**	90%
	(0.0399)	(0.0586)	(0.1717)	(0.0412)	(0.1145)	(0.2265)	(0.0529)	
Peng	0.7416	0.4630	3.4624	0.7743	0.7128	3.3213	**1.5792**	35%
	(0.0736)	(0.0506)	(0.1902)	(0.0790)	(0.0784)	(0.2415)	(0.0565)	
ML	0.6230	0.4666	4.3340	0.7309	0.6586	3.9013	**1.7857**	30%
	(0.0626)	(0.0524)	(0.2210)	(0.0753)	(0.0621)	(0.2711)	(0.0620)	
ML(ρ_3)	0.4989	0.4708	2.3844	0.5229	0.7584	4.2617	**1.4829**	35%
	(0.0463)	(0.0551)	(0.1570)	(0.0470)	(0.0602)	(0.2466)	(0.0518)	
ML(ρ_4)	1.0550	0.9254	4.7476	1.3174	0.9796	4.0793	**2.1840**	10%
	(0.0997)	(0.0928)	(0.3677)	(0.1484)	(0.1031)	(0.3304)	(0.0906)	
GPD Bayes	1.8177	1.1562	3.4071	1.6873	1.7467	2.4561	**2.0452**	25%
Mode	(0.1548)	(0.1037)	(0.2651)	(0.1498)	(0.1357)	(0.2049)	(0.0722)	
GPD MLE	1.1566	0.8231	4.5780	1.3005	0.9608	3.8423	**2.1102**	20%
	(0.1214)	(0.0881)	(0.3049)	(0.1214)	(0.0936)	(0.2878)	(0.0785)	
Zipf	1.3849	1.1107	5.9693	1.3702	1.0180	4.0930	**2.4910**	20%
	(0.1443)	(0.1062)	(0.3691)	(0.1408)	(0.1144)	(0.3105)	(0.0909)	
ERM	0.9271	0.6250	5.2409	1.0514	0.8151	4.0556	**2.1192**	25%
	(0.0975)	(0.0643)	(0.2917)	(0.1012)	(0.0820)	(0.2759)	(0.0730)	

Table 4.7.18 Simulation results for $n = 2000$ and EVI group $\gamma = 1$.

Distr.	Burr	Burr	Burr	Fréchet	logΓ	Mean	*k*
ρ	**-2**	**-1**	**-0.5**	**-1**	**0**		
PPD	0.0065	0.0059	0.0183	0.0056	0.0159	**0.0104**	
	(0.0008)	(0.0007)	(0.0022)	(0.0006)	(0.0015)	(0.0006)	
Hill	0.0041	0.0057	0.0420	0.0043	0.0389	**0.0190**	5%
	(0.0006)	(0.0008)	(0.0031)	(0.0005)	(0.0027)	(0.0008)	
Moment	0.1617	0.2594	0.4992	0.1998	0.1759	**0.2592**	10%
	(0.0233)	(0.0313)	(0.0375)	(0.0262)	(0.0243)	(0.0130)	
γ_2	0.0063	0.0077	0.0232	0.0067	0.0317	**0.0151**	5%
	(0.0008)	(0.0011)	(0.0030)	(0.0009)	(0.0032)	(0.0009)	
γ_3	0.0045	0.0055	0.0303	0.0047	0.0340	**0.0158**	5%
	(0.0006)	(0.0008)	(0.0029)	(0.0007)	(0.0027)	(0.0008)	
GJ	0.0047	0.0048	0.0214	0.0042	0.0299	**0.0130**	20%
	(0.0006)	(0.0007)	(0.0022)	(0.0006)	(0.0025)	(0.0007)	
GJ(ρ_3)	0.0030	0.0074	0.0028	0.0018	0.0366	**0.0103**	75%
	(0.0004)	(0.0007)	(0.0003)	(0.0002)	(0.0022)	(0.0005)	
GJ(ρ_4)	0.0053	0.0067	0.0233	0.0057	0.0312	**0.0144**	10%
	(0.0007)	(0.0010)	(0.0028)	(0.0008)	(0.0029)	(0.0009)	
NGJ	0.0037	0.0068	0.0074	0.0034	0.0324	**0.0108**	45%
	(0.0004)	(0.0008)	(0.0009)	(0.0005)	(0.0023)	(0.0005)	
NGJ(ρ_3)	0.0025	0.0082	0.0100	0.0034	0.0367	**0.0122**	30%
	(0.0003)	(0.0013)	(0.0014)	(0.0005)	(0.0028)	(0.0007)	
NGJ(ρ_4)	0.0027	0.0058	0.0063	0.0017	0.0407	**0.0115**	85%
	(0.0003)	(0.0005)	(0.0008)	(0.0002)	(0.0022)	(0.0005)	
Peng	0.0048	0.0049	0.0203	0.0043	0.0298	**0.0128**	20%
	(0.0006)	(0.0007)	(0.0022)	(0.0006)	(0.0025)	(0.0007)	
ML	0.0059	0.0060	0.0223	0.0048	0.0337	**0.0145**	15%
	(0.0008)	(0.0008)	(0.0022)	(0.0006)	(0.0030)	(0.0008)	
ML(ρ_3)	0.0025	0.0038	0.0155	0.0029	0.0356	**0.0121**	25%
	(0.0003)	(0.0005)	(0.0017)	(0.0005)	(0.0026)	(0.0006)	
ML(ρ_4)	0.0045	0.0068	0.0316	0.0044	0.0373	**0.0169**	10%
	(0.0006)	(0.0009)	(0.0028)	(0.0006)	(0.0031)	(0.0009)	
GPD Bayes	0.1412	0.1600	0.2539	0.1460	0.1547	**0.1711**	10%
Mean	(0.0176)	(0.0186)	(0.0183)	(0.0177)	(0.0175)	(0.0080)	
GPD MLE	0.1854	0.2922	0.5401	0.2335	0.2148	**0.2932**	10%
	(0.0246)	(0.0329)	(0.0378)	(0.0280)	(0.0274)	(0.0136)	
Zipf	0.1568	0.2813	0.5601	0.2310	0.1589	**0.2776**	25%
	(0.0223)	(0.0317)	(0.0379)	(0.0291)	(0.0208)	(0.0130)	
ERM	0.1826	0.2807	0.5297	0.2300	0.2092	**0.2864**	10%
	(0.0239)	(0.0328)	(0.0383)	(0.0281)	(0.0264)	(0.0136)	

Table 4.7.19 Simulation results for $n = 5000$ and EVI group $\gamma = 0.1$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	***t***	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	0.1095	0.1008	0.3770	0.1172	0.1356	0.4582	**0.2164**	
	(0.0114)	(0.0116)	(0.0433)	(0.0120)	(0.0197)	(0.0336)	(0.0103)	
Hill	0.0900	0.1360	0.9951	0.0915	0.1969	0.9695	**0.4132**	5%
	(0.0107)	(0.0245)	(0.0720)	(0.0135)	(0.0253)	(0.0664)	(0.0176)	
Moment	0.1541	0.1940	0.5377	0.1455	0.2691	0.8088	**0.3515**	15%
	(0.0185)	(0.0243)	(0.0553)	(0.0187)	(0.0304)	(0.0650)	(0.0162)	
γ_2	0.1760	0.1888	0.5038	0.1588	0.2469	0.8288	**0.3505**	5%
	(0.0234)	(0.0246)	(0.0556)	(0.0205)	(0.0287)	(0.0748)	(0.0175)	
γ_3	0.1102	0.1405	0.6889	0.1045	0.1930	0.8635	**0.3501**	5%
	(0.0134)	(0.0212)	(0.0578)	(0.0140)	(0.0236)	(0.0630)	(0.0155)	
GJ	0.0743	0.0873	0.5623	0.0785	0.1059	0.7892	**0.2829**	25%
	(0.0083)	(0.0117)	(0.0423)	(0.0118)	(0.0126)	(0.0541)	(0.0120)	
GJ(ρ_3)	0.0491	0.0976	0.2414	0.0648	0.1460	0.8521	**0.2418**	30%
	(0.0052)	(0.0114)	(0.0267)	(0.0092)	(0.0171)	(0.0567)	(0.0111)	
GJ(ρ_4)	0.1341	0.1706	0.5066	0.1329	0.2109	0.8041	**0.3265**	10%
	(0.0161)	(0.0225)	(0.0521)	(0.0175)	(0.0257)	(0.0709)	(0.0162)	
NGJ	0.0662	0.1151	0.3402	0.0902	0.1474	0.8886	**0.2746**	35%
	(0.0084)	(0.0143)	(0.0412)	(0.0150)	(0.0165)	(0.0567)	(0.0126)	
NGJ(ρ_3)	0.0560	0.1832	0.2585	0.0908	0.3107	0.9454	**0.3074**	25%
	(0.0064)	(0.0241)	(0.0361)	(0.0119)	(0.0377)	(0.0698)	(0.0152)	
NGJ(ρ_4)	0.0324	0.0450	0.6457	0.0512	0.2873	1.1871	**0.3748**	80%
	(0.0045)	(0.0060)	(0.0384)	(0.0071)	(0.0280)	(0.0527)	(0.0120)	
Peng	0.0745	0.0899	0.5162	0.0799	0.1093	0.7847	**0.2758**	25%
	(0.0083)	(0.0120)	(0.0414)	(0.0119)	(0.0128)	(0.0543)	(0.0120)	
ML	0.0780	0.1046	0.6391	0.0871	0.1218	0.8735	**0.3174**	20%
	(0.0083)	(0.0138)	(0.0457)	(0.0113)	(0.0160)	(0.0614)	(0.0134)	
ML(ρ_3)	0.0440	0.0904	0.3318	0.0665	0.1166	0.9168	**0.2610**	25%
	(0.0053)	(0.0118)	(0.0336)	(0.0105)	(0.0124)	(0.0603)	(0.0120)	
ML(ρ_4)	0.0926	0.1557	0.7508	0.1056	0.1909	0.9328	**0.3714**	10%
	(0.0107)	(0.0230)	(0.0652)	(0.0155)	(0.0275)	(0.0806)	(0.0186)	
GPD Bayes	0.1281	0.3425	0.1847	0.2214	0.5017	0.6609	**0.3399**	25%
Mean	(0.0138)	(0.0419)	(0.0232)	(0.0316)	(0.0555)	(0.0612)	(0.0169)	
GPD MLE	0.1361	0.3772	0.1811	0.2437	0.5462	0.5888	**0.3455**	25%
	(0.0149)	(0.0452)	(0.0224)	(0.0329)	(0.0591)	(0.0553)	(0.0170)	
Zipf	0.1795	0.2182	0.6240	0.1817	0.3057	0.8499	**0.3932**	30%
	(0.0210)	(0.0281)	(0.0611)	(0.0232)	(0.0354)	(0.0740)	(0.0184)	
ERM	0.1395	0.3429	0.1831	0.2189	0.5121	0.6174	**0.3357**	25%
	(0.0155)	(0.0421)	(0.0227)	(0.0303)	(0.0576)	(0.0564)	(0.0166)	

Table 4.7.20 Simulation results for $n = 5000$ and EVI group $\gamma = 0.5$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	***t***	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	0.7888	0.4069	1.7775	0.5022	7.2427	1.3852	**0.9289**	
	(0.0761)	(0.0541)	(0.2028)	(0.0508)	(0.0724)	(0.1085)	(0.0439)	
Hill	0.5408	0.5570	3.7621	0.3325	0.3768	3.6082	**1.5296**	5%
	(0.0877)	(0.0895)	(0.2765)	(0.0453)	(0.0551)	(0.2792)	(0.0698)	
Moment	1.0840	0.6861	1.8442	0.7510	0.7607	2.7931	**1.3198**	5%
	(0.1377)	(0.0835)	(0.1876)	(0.0963)	(0.1103)	(0.3020)	(0.0695)	
γ_2	0.9716	0.6867	1.8004	0.7145	0.8340	2.8923	**1.3166**	5%
	(0.1171)	(0.0881)	(0.1849)	(0.1055)	(0.1624)	(0.3303)	(0.0750)	
γ_3	0.6576	0.5306	2.5546	0.4167	0.4862	3.0658	**1.2852**	5%
	(0.0899)	(0.0769)	(0.2089)	(0.0599)	(0.0746)	(0.2821)	(0.0638)	
GJ	0.4956	0.3288	2.2343	0.3525	0.4193	2.8193	**1.1083**	25%
	(0.0584)	(0.0474)	(0.1642)	(0.0502)	(0.0610)	(0.2235)	(0.0497)	
GJ(ρ_3)	0.3214	0.3297	0.8829	0.2301	0.3678	3.0297	**0.8603**	35%
	(0.0367)	(0.0496)	(0.0989)	(0.0329)	(0.0519)	(0.2088)	(0.0412)	
GJ(ρ_4)	0.7805	0.6322	1.8997	0.5791	0.6388	2.6853	**1.2026**	10%
	(0.1026)	(0.0870)	(0.1792)	(0.0770)	(0.1078)	(0.3013)	(0.0664)	
NGJ	0.4385	0.4944	0.9421	0.2817	0.5657	3.1761	**0.9831**	40%
	(0.0471)	(0.0742)	(0.1674)	(0.0326)	(0.0726)	(0.2036)	(0.0482)	
NGJ(ρ_3)	0.3482	0.6870	1.2253	0.3392	0.4882	3.3520	**1.0733**	30%
	(0.0471)	(0.1023)	(0.1570)	(0.0475)	(0.0742)	(0.2313)	(0.0523)	
NGJ(ρ_4)	0.3126	0.6404	0.7311	0.1551	1.6028	3.9159	**1.2263**	85%
	(0.0350)	(0.0620)	(0.0838)	(0.0214)	(0.0924)	(0.2154)	(0.0433)	
Peng	0.4999	0.3351	2.0453	0.3585	0.4231	2.7997	**1.0769**	25%
	(0.0591)	(0.0481)	(0.1613)	(0.0506)	(0.0613)	(0.2243)	(0.0496)	
ML	0.4948	0.3491	2.6289	0.3809	0.3964	3.2068	**1.2428**	20%
	(0.0709)	(0.0525)	(0.2245)	(0.0528)	(0.0485)	(0.2290)	(0.0567)	
ML(ρ_3)	0.3042	0.3316	1.3510	0.2610	0.3609	3.2472	**0.9760**	25%
	(0.0399)	(0.0461)	(0.1479)	(0.0360)	(0.0494)	(0.2219)	(0.0467)	
ML(ρ_4)	0.5677	0.6619	2.9679	0.4114	0.4649	2.9563	**1.3383**	10%
	(0.0941)	(0.1066)	(0.2393)	(0.0564)	(0.0716)	(0.2698)	(0.0664)	
GPD Bayes	1.3392	0.8565	1.5176	0.8871	0.9740	2.1494	**1.2873**	15%
Mode	(0.1524)	(0.1301)	(0.1898)	(0.0986)	(0.1320)	(0.2261)	(0.0655)	
GPD MLE	0.7518	0.5507	2.7290	0.4594	0.5489	3.2702	**1.3850**	15%
	(0.1188)	(0.0844)	(0.2550)	(0.0635)	(0.0794)	(0.2453)	(0.0660)	
Zipf	0.9057	0.6474	3.0077	0.6919	0.7632	3.2774	**1.5489**	15%
	(0.1131)	(0.0745)	(0.2359)	(0.0952)	(0.1166)	(0.3005)	(0.0721)	
ERM	0.7723	0.5734	2.5415	0.4599	0.5527	3.2306	**1.3551**	15%
	(0.1211)	(0.0892)	(0.2374)	(0.0636)	(0.0846)	(0.2540)	(0.0656)	

Table 4.7.21 Simulation results for $n = 5000$ and EVI group $\gamma = 1$.

Distr.	Burr	Burr	Burr	Fréchet	logΓ	Mean	*k*
ρ	**-2**	**-1**	**-0.5**	**-1**	**0**		
PPD	0.0042	0.0053	0.0111	0.0040	0.0129	**0.0075**	
	(0.0005)	(0.0006)	(0.0011)	(0.0005)	(0.0010)	(0.0003)	
Hill	0.0021	0.0025	0.0349	0.0017	0.0317	**0.0146**	5%
	(0.0003)	(0.0004)	(0.0019)	(0.0002)	(0.0019)	(0.0006)	
Moment	0.0968	0.1606	0.4340	0.1083	0.1118	**0.1823**	10%
	(0.0138)	(0.0188)	(0.0335)	(0.0148)	(0.0154)	(0.0092)	
γ_2	0.0039	0.0032	0.0194	0.0036	0.0221	**0.0104**	5%
	(0.0005)	(0.0004)	(0.0018)	(0.0005)	(0.0019)	(0.0006)	
γ_3	0.0026	0.0024	0.0257	0.0022	0.0260	**0.0118**	5%
	(0.0003)	(0.0004)	(0.0018)	(0.0003)	(0.0017)	(0.0005)	
GJ	0.0035	0.0026	0.0136	0.0028	0.0217	**0.0088**	15%
	(0.0004)	(0.0003)	(0.0013)	(0.0004)	(0.0018)	(0.0005)	
GJ(ρ_3)	0.0019	0.0027	0.0066	0.0021	0.0232	**0.0073**	20%
	(0.0002)	(0.0003)	(0.0009)	(0.0003)	(0.0017)	(0.0004)	
GJ(ρ_4)	0.0031	0.0029	0.0184	0.0029	0.0216	**0.0098**	10%
	(0.0004)	(0.0004)	(0.0016)	(0.0004)	(0.0018)	(0.0005)	
NGJ	0.0029	0.0054	0.0051	0.0019	0.0292	**0.0089**	45%
	(0.0003)	(0.0005)	(0.0008)	(0.0002)	(0.0013)	(0.0003)	
NGJ(ρ_3)	0.0019	0.0043	0.0068	0.0021	0.0274	**0.0085**	20%
	(0.0003)	(0.0006)	(0.0010)	(0.0003)	(0.0020)	(0.0005)	
NGJ(ρ_4)	0.0012	0.0030	0.0081	0.0011	0.0339	**0.0095**	80%
	(0.0002)	(0.0003)	(0.0007)	(0.0002)	(0.0016)	(0.0004)	
Peng	0.0035	0.0027	0.0130	0.0028	0.0216	**0.0087**	15%
	(0.0004)	(0.0003)	(0.0013)	(0.0004)	(0.0018)	(0.0005)	
ML	0.0028	0.0024	0.0161	0.0023	0.0261	**0.0099**	15%
	(0.0003)	(0.0003)	(0.0013)	(0.0003)	(0.0019)	(0.0005)	
ML(ρ_3)	0.0020	0.0028	0.0078	0.0021	0.0255	**0.0081**	15%
	(0.0002)	(0.0004)	(0.0011)	(0.0003)	(0.0019)	(0.0005)	
ML(ρ_4)	0.0039	0.0046	0.0170	0.0045	0.0214	**0.0103**	5%
	(0.0005)	(0.0007)	(0.0022)	(0.0006)	(0.0021)	(0.0006)	
GPD Bayes	0.1159	0.1260	0.2984	0.0980	0.1031	**0.1483**	10%
Mean	(0.0179)	(0.0134)	(0.0208)	(0.0133)	(0.0118)	(0.0071)	
GPD MLE	0.1180	0.1811	0.4706	0.1302	0.1411	**0.2082**	10%
	(0.0177)	(0.0206)	(0.0342)	(0.0182)	(0.0188)	(0.0102)	
Zipf	0.1166	0.1528	0.4238	0.1248	0.1477	**0.1931**	20%
	(0.0136)	(0.0172)	(0.0339)	(0.0177)	(0.0198)	(0.0097)	
ERM	0.1172	0.1708	0.4618	0.1254	0.1380	**0.2026**	10%
	(0.0174)	(0.0196)	(0.0347)	(0.0180)	(0.0187)	(0.0101)	

Table 4.7.22 Simulation results for $n = 10\,000$ and EVI group $\gamma = 0.1$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	***t***	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	0.0914	0.0793	0.2843	0.0843	0.1036	0.3863	**0.1715**	
	(0.0098)	(0.0098)	(0.0343)	(0.0089)	(0.0128)	(0.0269)	(0.0080)	
Hill	0.0430	0.0669	0.8176	0.0455	0.0864	0.8528	**0.3187**	5%
	(0.0054)	(0.0091)	(0.0438)	(0.0057)	(0.0114)	(0.0453)	(0.0109)	
Moment	0.0993	0.0985	0.3211	0.1126	0.1641	0.6447	**0.2401**	10%
	(0.0154)	(0.0170)	(0.0364)	(0.0155)	(0.0220)	(0.0488)	(0.0117)	
γ_2	0.0770	0.0664	0.4057	0.0948	0.1157	0.6962	**0.2426**	5%
	(0.0105)	(0.0111)	(0.0404)	(0.0154)	(0.0165)	(0.0522)	(0.0119)	
γ_3	0.0511	0.0540	0.5678	0.0583	0.0857	0.7557	**0.2621**	5%
	(0.0069)	(0.0075)	(0.0397)	(0.0088)	(0.0120)	(0.0450)	(0.0104)	
GJ	0.0678	0.0548	0.3021	0.0747	0.0840	0.6381	**0.2036**	15%
	(0.0098)	(0.0090)	(0.0275)	(0.0109)	(0.0101)	(0.0456)	(0.0095)	
GJ(ρ_3)	0.0335	0.0408	0.1283	0.0434	0.0971	0.6970	**0.1733**	25%
	(0.0046)	(0.0076)	(0.0139)	(0.0062)	(0.0112)	(0.0432)	(0.0080)	
GJ(ρ_4)	0.0587	0.0605	0.3900	0.0746	0.0973	0.6519	**0.2222**	10%
	(0.0085)	(0.0096)	(0.0362)	(0.0110)	(0.0133)	(0.0486)	(0.0107)	
NGJ	0.0579	0.0575	0.2848	0.0522	0.0804	0.7301	**0.2105**	25%
	(0.0063)	(0.0082)	(0.0274)	(0.0064)	(0.0107)	(0.0503)	(0.0099)	
NGJ(ρ_3)	0.0507	0.0949	0.1419	0.0623	0.2204	0.7380	**0.2180**	20%
	(0.0060)	(0.0113)	(0.0183)	(0.0090)	(0.0291)	(0.0552)	(0.0111)	
NGJ(ρ_4)	0.0548	0.0886	0.4880	0.0603	0.0950	0.7459	**0.2554**	10%
	(0.0071)	(0.0127)	(0.0434)	(0.0078)	(0.0123)	(0.0523)	(0.0118)	
Peng	0.0537	0.0416	0.3438	0.0561	0.0625	0.6467	**0.2007**	20%
	(0.0073)	(0.0064)	(0.0251)	(0.0079)	(0.0084)	(0.0421)	(0.0085)	
ML	0.0687	0.0869	0.3110	0.0755	0.1074	0.6581	**0.2179**	10%
	(0.0098)	(0.0138)	(0.0324)	(0.0090)	(0.0138)	(0.0484)	(0.0105)	
ML(ρ_3)	0.0364	0.0510	0.1904	0.0453	0.0830	0.7263	**0.1887**	20%
	(0.0044)	(0.0070)	(0.0180)	(0.0051)	(0.0107)	(0.0489)	(0.0090)	
ML(ρ_4)	0.0762	0.0838	0.3854	0.1025	0.1504	0.6721	**0.2451**	5%
	(0.0101)	(0.0138)	(0.0469)	(0.0129)	(0.0236)	(0.0572)	(0.0134)	
GPD Bayes	0.1140	0.2092	0.1086	0.1131	0.3472	0.4976	**0.2316**	20%
Mean	(0.0160)	(0.0250)	(0.0132)	(0.0135)	(0.0387)	(0.0486)	(0.0119)	
GPD MLE	0.1210	0.2290	0.1114	0.1224	0.3818	0.4593	**0.2375**	20%
	(0.0175)	(0.0272)	(0.0138)	(0.0146)	(0.0410)	(0.0463)	(0.0121)	
Zipf	0.1280	0.0964	0.4247	0.1244	0.1600	0.6722	**0.2676**	25%
	(0.0194)	(0.0151)	(0.0442)	(0.0172)	(0.0229)	(0.0515)	(0.0129)	
ERM	0.1167	0.1982	0.1157	0.1198	0.3475	0.4721	**0.2284**	20%
	(0.0173)	(0.0257)	(0.0149)	(0.0146)	(0.0395)	(0.0456)	(0.0118)	

Table 4.7.23 Simulation results for $n = 10\,000$ and EVI group $\gamma = 0.5$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	***t***	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	0.3659	0.3506	1.1300	0.2354	0.4183	1.5004	**0.6668**	
	(0.0376)	(0.0386)	(0.1286)	(0.0257)	(0.0435)	(0.1069)	(0.0305)	
Hill	0.1978	0.2207	3.4522	0.2619	0.2200	3.2270	**1.2633**	5%
	(0.0300)	(0.0349)	(0.1866)	(0.0413)	(0.0394)	(0.1927)	(0.0463)	
Moment	0.3481	0.3513	1.6637	0.4958	0.4226	2.7244	**1.0010**	5%
	(0.0494)	(0.0449)	(0.1538)	(0.0847)	(0.0633)	(0.2258)	(0.0501)	
γ_2	0.3432	0.3465	1.6137	0.5007	0.4107	2.7807	**0.9992**	5%
	(0.0444)	(0.0468)	(0.1489)	(0.0982)	(0.0622)	(0.2455)	(0.0527)	
γ_3	0.2185	0.2350	2.3578	0.3254	0.2594	2.9291	**1.0542**	5%
	(0.0303)	(0.0369)	(0.1584)	(0.0639)	(0.0438)	(0.2059)	(0.0459)	
GJ	0.2849	0.2418	1.2908	0.4012	0.3716	2.5462	**0.8561**	15%
	(0.0380)	(0.0372)	(0.1125)	(0.0707)	(0.0544)	(0.2185)	(0.0445)	
GJ(ρ_3)	0.1383	0.2033	0.6296	0.1958	0.1982	2.7817	**0.6912**	25%
	(0.0163)	(0.0303)	(0.0696)	(0.0298)	(0.0275)	(0.1760)	(0.0328)	
GJ(ρ_4)	0.2591	0.2983	1.6331	0.4171	0.3293	2.5517	**0.9148**	10%
	(0.0334)	(0.0443)	(0.1385)	(0.0790)	(0.0526)	(0.2279)	(0.0481)	
NGJ	0.1785	0.2918	0.8686	0.1288	0.2423	3.1245	**0.8058**	35%
	(0.0213)	(0.0356)	(0.0889)	(0.0190)	(0.0281)	(0.1582)	(0.0315)	
NGJ(ρ_3)	0.2104	0.2891	0.6434	0.2611	0.2600	3.0632	**0.7878**	20%
	(0.0252)	(0.0404)	(0.1020)	(0.0332)	(0.0388)	(0.2028)	(0.0396)	
NGJ(ρ_4)	0.1261	0.3074	0.7106	0.1402	0.7830	3.8155	**0.9805**	80%
	(0.0185)	(0.0324)	(0.0621)	(0.0184)	(0.0454)	(0.1715)	(0.0321)	
Peng	0.2171	0.1817	1.4804	0.2589	0.2626	2.6559	**0.8428**	20%
	(0.0270)	(0.0263)	(0.1060)	(0.0451)	(0.0388)	(0.1817)	(0.0370)	
ML	0.2753	0.1851	1.7236	0.3061	0.3103	2.8054	**0.9343**	15%
	(0.0392)	(0.0302)	(0.1287)	(0.0377)	(0.0429)	(0.2104)	(0.0430)	
ML(ρ_3)	0.1549	0.1692	0.8667	0.1788	0.1705	2.9909	**0.7552**	20%
	(0.0174)	(0.0241)	(0.0792)	(0.0220)	(0.0250)	(0.1807)	(0.0337)	
ML(ρ_4)	0.1929	0.2633	2.4119	0.2953	0.2374	2.7409	**1.0236**	10%
	(0.0292)	(0.0478)	(0.1712)	(0.0373)	(0.0382)	(0.2107)	(0.0471)	
GPD Bayes	0.5235	0.4497	1.5089	0.3744	0.4147	2.0429	**0.8857**	15%
Mode	(0.0686)	(0.0625)	(0.1539)	(0.0425)	(0.0509)	(0.1942)	(0.0455)	
GPD MLE	0.3589	0.3929	1.6281	0.5002	0.3901	2.5975	**0.9779**	10%
	(0.0517)	(0.0605)	(0.1532)	(0.0838)	(0.0651)	(0.2193)	(0.0498)	
Zipf	0.4367	0.4777	1.9165	0.6560	0.5988	2.9602	**1.1743**	10%
	(0.0542)	(0.0603)	(0.1825)	(0.1133)	(0.0855)	(0.2595)	(0.0595)	
ERM	0.3726	0.3818	1.5063	0.5223	0.4049	2.6547	**0.9738**	10%
	(0.0504)	(0.0589)	(0.1429)	(0.0930)	(0.0687)	(0.2269)	(0.0503)	

Table 4.7.24 Simulation results for $n = 10\,000$ and EVI group $\gamma = 1$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**0**		
PPD	0.0027	0.0030	0.0098	0.0024	0.0132	**0.0062**	
	(0.0003)	(0.0004)	(0.0012)	(0.0002)	(0.0011)	(0.0004)	
Hill	0.0010	0.0018	0.0339	0.0010	0.0322	**0.0140**	5%
	(0.0001)	(0.0002)	(0.0012)	(0.0001)	(0.0014)	(0.0004)	
Moment	0.0831	0.1211	0.2711	0.0917	0.1076	**0.1349**	5%
	(0.0129)	(0.0164)	(0.0282)	(0.0115)	(0.0135)	(0.0079)	
γ_2	0.0018	0.0022	0.0158	0.0017	0.0228	**0.0088**	5%
	(0.0002)	(0.0003)	(0.0010)	(0.0003)	(0.0015)	(0.0004)	
γ_3	0.0012	0.0017	0.0235	0.0011	0.0269	**0.0109**	5%
	(0.0001)	(0.0002)	(0.0010)	(0.0002)	(0.0013)	(0.0003)	
GJ	0.0021	0.0026	0.0087	0.0020	0.0199	**0.0071**	10%
	(0.0003)	(0.0003)	(0.0008)	(0.0003)	(0.0016)	(0.0004)	
GJ(ρ_3)	0.0012	0.0018	0.0045	0.0013	0.0208	**0.0059**	15%
	(0.0002)	(0.0002)	(0.0005)	(0.0002)	(0.0014)	(0.0003)	
GJ(ρ_4)	0.0026	0.0035	0.0091	0.0029	0.0181	**0.0073**	5%
	(0.0003)	(0.0005)	(0.0010)	(0.0004)	(0.0018)	(0.0004)	
NGJ	0.0019	0.0028	0.0067	0.0008	0.0288	**0.0082**	40%
	(0.0002)	(0.0003)	(0.0006)	(0.0001)	(0.0010)	(0.0002)	
NGJ(ρ_3)	0.0007	0.0018	0.0031	0.0011	0.0259	**0.0065**	25%
	(0.0001)	(0.0003)	(0.0005)	(0.0001)	(0.0015)	(0.0003)	
NGJ(ρ_4)	0.0011	0.0046	0.0014	0.0005	0.0315	**0.0078**	80%
	(0.0001)	(0.0004)	(0.0002)	(0.0001)	(0.0011)	(0.0002)	
Peng	0.0022	0.0026	0.0084	0.0020	0.0198	**0.0070**	10%
	(0.0003)	(0.0003)	(0.0008)	(0.0003)	(0.0016)	(0.0004)	
ML	0.0032	0.0041	0.0081	0.0038	0.0204	**0.0079**	5%
	(0.0004)	(0.0006)	(0.0009)	(0.0006)	(0.0021)	(0.0005)	
ML(ρ_3)	0.0013	0.0023	0.0053	0.0013	0.0224	**0.0065**	10%
	(0.0002)	(0.0003)	(0.0006)	(0.0002)	(0.0016)	(0.0004)	
ML(ρ_4)	0.0017	0.0028	0.0119	0.0023	0.0212	**0.0080**	5%
	(0.0002)	(0.0004)	(0.0010)	(0.0004)	(0.0019)	(0.0004)	
GPD Bayes	0.0983	0.1139	0.1995	0.1021	0.1086	**0.1245**	5%
Mean	(0.0137)	(0.0138)	(0.0179)	(0.0135)	(0.0127)	(0.0065)	
GPD MLE	0.0986	0.1395	0.3111	0.1140	0.1374	**0.1601**	5%
	(0.0144)	(0.0178)	(0.0303)	(0.0148)	(0.0165)	(0.0088)	
Zipf	0.0655	0.1093	0.3144	0.0679	0.0850	**0.1284**	15%
	(0.0087)	(0.0136)	(0.0254)	(0.0082)	(0.0105)	(0.0066)	
ERM	0.1026	0.1334	0.2997	0.1103	0.1324	**0.1557**	5%
	(0.0150)	(0.0174)	(0.0305)	(0.0143)	(0.0157)	(0.0087)	

Table 4.7.25 Simulation results for $n = 20000$ and EVI group $\gamma = 0.1$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	***t***	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	0.1063	0.1383	0.1101	0.1142	0.0970	0.2668	**0.1388**	
	(0.0157)	(0.0311)	(0.0129)	(0.0128)	(0.0119)	(0.0301)	(0.0085)	
Hill	0.0280	0.0379	0.8362	0.0271	0.0654	0.7769	**0.2952**	5%
	(0.0045)	(0.0054)	(0.0331)	(0.0038)	(0.0070)	(0.0312)	(0.0078)	
Moment	0.0633	0.0523	0.2718	0.0572	0.0723	0.5818	**0.1831**	10%
	(0.0083)	(0.0074)	(0.0239)	(0.0097)	(0.0104)	(0.0411)	(0.0085)	
γ_2	0.0478	0.0447	0.3763	0.0486	0.0616	0.6260	**0.2008**	5%
	(0.0068)	(0.0073)	(0.0263)	(0.0101)	(0.0088)	(0.0420)	(0.0087)	
γ_3	0.0321	0.0348	0.5688	0.0309	0.0550	0.6902	**0.2353**	5%
	(0.0048)	(0.0059)	(0.0282)	(0.0059)	(0.0064)	(0.0348)	(0.0077)	
GJ	0.0654	0.0482	0.2100	0.0576	0.0603	0.5487	**0.1650**	10%
	(0.0091)	(0.0077)	(0.0193)	(0.0115)	(0.0086)	(0.0431)	(0.0085)	
GJ(ρ_3)	0.0308	0.0357	0.1105	0.0335	0.0539	0.5580	**0.1371**	15%
	(0.0042)	(0.0048)	(0.0114)	(0.0054)	(0.0080)	(0.0381)	(0.0069)	
GJ(ρ_4)	0.0701	0.0734	0.2213	0.0837	0.0920	0.5548	**0.1826**	5%
	(0.0098)	(0.0102)	(0.0245)	(0.0153)	(0.0138)	(0.0530)	(0.0106)	
NGJ	0.0497	0.0346	0.2637	0.0377	0.0505	0.6366	**0.1788**	15%
	(0.0065)	(0.0041)	(0.0230)	(0.0055)	(0.0065)	(0.0371)	(0.0075)	
NGJ(ρ_3)	0.0320	0.0421	0.1008	0.0332	0.1016	0.6233	**0.1555**	15%
	(0.0041)	(0.0051)	(0.0139)	(0.0047)	(0.0134)	(0.0390)	(0.0074)	
NGJ(ρ_4)	0.0587	0.0754	0.3004	0.0693	0.0925	0.6383	**0.2058**	5%
	(0.0072)	(0.0117)	(0.0310)	(0.0101)	(0.0133)	(0.0505)	(0.0105)	
Peng	0.0652	0.0486	0.2040	0.0577	0.0613	0.5469	**0.1639**	10%
	(0.0090)	(0.0077)	(0.0192)	(0.0114)	(0.0087)	(0.0432)	(0.0085)	
ML	0.0595	0.0364	0.2733	0.0394	0.0474	0.6006	**0.1761**	10%
	(0.0088)	(0.0061)	(0.0219)	(0.0063)	(0.0065)	(0.0387)	(0.0078)	
ML(ρ_3)	0.0245	0.0273	0.1502	0.0240	0.0430	0.6100	**0.1465**	15%
	(0.0034)	(0.0039)	(0.0136)	(0.0034)	(0.0062)	(0.0363)	(0.0066)	
ML(ρ_4)	0.0524	0.0556	0.3023	0.0550	0.0771	0.6310	**0.1956**	5%
	(0.0064)	(0.0078)	(0.0298)	(0.0077)	(0.0126)	(0.0511)	(0.0103)	
GPD Bayes	0.0842	0.0963	0.0828	0.0781	0.1646	0.4325	**0.1564**	15%
Mean	(0.0117)	(0.0111)	(0.0127)	(0.0115)	(0.0209)	(0.0410)	(0.0086)	
GPD MLE	0.0847	0.1079	0.0846	0.0839	0.1808	0.4072	**0.1582**	15%
	(0.0113)	(0.0121)	(0.0134)	(0.0123)	(0.0223)	(0.0394)	(0.0086)	
Zipf	0.0777	0.0736	0.3043	0.0830	0.0826	0.5632	**0.1974**	20%
	(0.0101)	(0.0092)	(0.0267)	(0.0137)	(0.0111)	(0.0452)	(0.0095)	
ERM	0.0830	0.1014	0.0894	0.0814	0.1647	0.4167	**0.1561**	15%
	(0.0113)	(0.0114)	(0.0137)	(0.0114)	(0.0216)	(0.0405)	(0.0086)	

Table 4.7.26 Simulation results for $n = 20\,000$ and EVI group $\gamma = 0.5$.

Distr.	**Burr**	**Burr**	**Burr**	**Fréchet**	***t***	**logΓ**	**Mean**	***k***
ρ	**-2**	**-1**	**-0.5**	**-1**	**-1**	**0**		
PPD	0.3776	0.3830	0.4330	0.5100	7.2427	1.0225	**0.5145**	
	(0.0503)	(0.0473)	(0.0562)	(0.0556)	(0.0450)	(0.0883)	(0.0241)	
Hill	0.0743	0.1852	3.3000	0.1216	0.1015	3.3167	**1.1832**	5%
	(0.0084)	(0.0199)	(0.1388)	(0.0171)	(0.0146)	(0.1182)	(0.0308)	
Moment	0.1978	0.1871	1.5891	0.2378	0.1701	2.5932	**0.8292**	5%
	(0.0281)	(0.0233)	(0.1264)	(0.0318)	(0.0190)	(0.1330)	(0.0318)	
γ_2	0.2042	0.1848	1.5394	0.2289	0.1711	2.5564	**0.8141**	5%
	(0.0296)	(0.0239)	(0.1215)	(0.0313)	(0.0196)	(0.1390)	(0.0320)	
γ_3	0.1083	0.1574	2.2831	0.1508	0.1190	2.8937	**0.9521**	5%
	(0.0137)	(0.0198)	(0.1228)	(0.0223)	(0.0142)	(0.1193)	(0.0292)	
GJ	0.2453	0.1751	0.8622	0.2834	0.2139	2.2497	**0.6716**	10%
	(0.0349)	(0.0217)	(0.0938)	(0.0385)	(0.0243)	(0.1462)	(0.0307)	
GJ(ρ_3)	0.1212	0.1374	0.4039	0.1860	0.1289	2.3387	**0.5527**	15%
	(0.0175)	(0.0183)	(0.0508)	(0.0285)	(0.0143)	(0.1215)	(0.0230)	
GJ(ρ_4)	0.3317	0.2897	0.9169	0.3844	0.2551	2.1549	**0.7221**	5%
	(0.0487)	(0.0358)	(0.1220)	(0.0487)	(0.0304)	(0.1784)	(0.0386)	
NGJ	0.1199	0.1277	0.9947	0.1371	0.1077	2.7782	**0.7109**	20%
	(0.0164)	(0.0193)	(0.0822)	(0.0233)	(0.0148)	(0.1255)	(0.0258)	
NGJ(ρ_3)	0.0799	0.1654	0.2892	0.1330	0.0927	2.7680	**0.5881**	20%
	(0.0110)	(0.0253)	(0.0383)	(0.0231)	(0.0127)	(0.1333)	(0.0240)	
NGJ(ρ_4)	0.0932	0.4111	0.1330	0.0798	0.7559	3.3523	**0.8042**	80%
	(0.0098)	(0.0267)	(0.0159)	(0.0114)	(0.0393)	(0.1134)	(0.0208)	
Peng	0.2450	0.1767	0.8371	0.2852	0.2140	2.2426	**0.6668**	10%
	(0.0347)	(0.0218)	(0.0935)	(0.0387)	(0.0243)	(0.1463)	(0.0307)	
ML	0.1746	0.1516	1.0982	0.2418	0.1846	2.5126	**0.7272**	10%
	(0.0230)	(0.0203)	(0.0915)	(0.0369)	(0.0234)	(0.1380)	(0.0290)	
ML(ρ_3)	0.0859	0.1168	0.5586	0.1512	0.1078	2.5803	**0.6001**	15%
	(0.0124)	(0.0149)	(0.0538)	(0.0254)	(0.0130)	(0.1181)	(0.0224)	
ML(ρ_4)	0.2302	0.2426	1.2435	0.3258	0.2029	2.4261	**0.7785**	5%
	(0.0350)	(0.0273)	(0.1266)	(0.0439)	(0.0280)	(0.1676)	(0.0368)	
GPD Bayes	0.2691	0.2232	0.9593	0.3584	0.3266	1.9438	**0.6801**	10%
Mode	(0.0364)	(0.0273)	(0.1070)	(0.0469)	(0.0513)	(0.1298)	(0.0313)	
GPD MLE	0.1735	0.1584	1.3104	0.2533	0.1837	2.6139	**0.7822**	10%
	(0.0224)	(0.0213)	(0.1026)	(0.0371)	(0.0229)	(0.1327)	(0.0293)	
Zipf	0.2821	0.2596	1.6331	0.2929	0.2179	2.6282	**0.8856**	10%
	(0.0393)	(0.0316)	(0.1367)	(0.0381)	(0.0287)	(0.1530)	(0.0361)	
ERM	0.1806	0.1590	1.2372	0.2582	0.1878	2.5774	**0.7667**	10%
	(0.0243)	(0.0192)	(0.1042)	(0.0379)	(0.0234)	(0.1316)	(0.0294)	

Table 4.7.27 Simulation results for $n = 20\,000$ and EVI group $\gamma = 1$.

Some aspects of the results presented here, will be discussed in the next section.

4.8 Discussion of simulation results

In this section we will draw conclusions from the simulation results shown in Section 4.7. From the tables in Section 4.7 one can conclude that the PPD estimator is the method of choice across all sample sizes. However, the PPD estimator does not yield the lowest MSE across all EVI groups for all sample sizes. Furthermore, there are other estimators which sometimes perform equally well for all practical purposes.

The most important reason for the detailed discussion in this section, is that it is extremely difficult to write a computer program to implement the PPD estimator. Also, the PPD estimator is computationally intensive. This section will provide alternatives to the PPD estimator, which are easy to program, and can be calculated rapidly.

These "simple" alternatives to the PPD estimator sometimes follow an adaptive procedure. Refer for instance to samples of size $n = 50$ in Section 4.8.1 below, where the alternative estimator is described as:

Estimate the EVI using GJ at $k = 90\%$ *of* n*. If the resulting estimate is above* 0.75*, leave it unchanged. Otherwise, redo the estimation by using Peng at* $k = 90\%$ *of* n*.*

The procedure is therefore adaptive, in the sense that a preliminary estimate of the EVI is obtained on which the final EVI estimate is based.

In Sections 4.8.1 and 4.8.2 it is stated which estimators are regarded as optimal, and which estimators could be used as alternatives to the PPD. Some general comments on the observed patterns in the behaviour of the estimators will be made in Section 4.8.3. In Section 4.8.4 we show the simulation study results pertaining to the alternative estimators.

In this section we will use, for the sake of brevity, the notation of the tables of Section 4.7 to refer to the estimators.

4.8.1 Samples of up to size $n = 1000$

For these sample sizes, the PPD estimator is without rival, across all sample sizes and across all EVI groups. (There is only one instance, namely at a sample size of $n = 500$ and EVI group $\gamma = 0.1$, where the PPD is in second place by approximately half a standard error. Refer to Table 4.7.10.)

This result is not much of a surprise. One would expect a Bayesian estimator to perform well for smaller sample sizes.

We will now discuss alternatives to the PPD estimator. Note that none of these estimators perform as well as the PPD estimator. If we use the expression "performs best", it will mean the best performing estimator after we remove the PPD estimator.

Sample size $n = 50$

GJ performs best on average (lowest average MSE over all EVI groups), since it performs quite well for $\gamma = 1$. For smaller EVIs, Peng performs better. We therefore estimate the EVI as follows:

Estimate the EVI using GJ at $k = 90\%$ of n. If the resulting estimate is above 0.75, leave it unchanged. Otherwise, redo the estimation by using Peng at $k = 90\%$ of n.

Sample size $n = 100$

ML(ρ_3) performs best on average (lowest average MSE over all EVI groups), since it performs quite well for $\gamma = 1$. For $\gamma = 0.5$ there is virtually no difference between Peng and ML(ρ_3), and for $\gamma = 0.1$ Peng performs better. We therefore estimate the EVI as follows:

Estimate the EVI using ML(ρ_3) at $k = 95\%$ of n. If the resulting estimate is above 0.75, leave it unchanged. Otherwise, redo the estimation by using Peng at $k = 90\%$ of n.

Sample size $n = 200$

Estimate the EVI using ML(ρ_3) at $k = 75\%$ of n. If the resulting estimate is

- above 0.75, leave the estimate unchanged.
- between 0.3 and 0.75, redo the estimation by using Peng at $k = 85\%$ of n.
- below 0.3, redo the estimation by using Peng at $k = 90\%$ of n.

Sample size $n = 500$

Here GJ(ρ_3) performs best. (Peng is marginally better for the EVI group $\gamma = 0.5$.) The optimal threshold, however, is not the same across all EVI groups. We suggest the following:

Estimate the EVI using GJ(ρ_3) at $k = 55\%$ of n. If the resulting estimate is

- above 0.75, leave the estimate unchanged.
- between 0.3 and 0.75, redo the estimation by using GJ(ρ_3) at $k = 60\%$ of n.
- below 0.3, redo the estimation by using GJ(ρ_3) at $k = 65\%$ of n.

Sample size $n = 1000$

Here GJ(ρ_3) performs best. The optimal threshold, however, is not the same across all EVI groups. We suggest the following:

Estimate the EVI using GJ(ρ_3) at $k = 55\%$ of n. If the resulting estimate is above 0.3, leave the estimate unchanged. Otherwise, redo the estimation by using GJ(ρ_3) at $k = 60\%$ of n.

4.8.2 Samples of size $n = 2000$ and larger

We have seen in Section 4.8.1 that GJ(ρ_3) is consistently the second best estimator (compared to the PPD estimator) for the two largest sample sizes considered there, namely for sample sizes $n = 500$ and $n = 1000$. GJ(ρ_3) seems to perform increasingly well relative to other estimators as the sample size increases. This trend continues for samples of size $n = 2000$ and larger, even to the extent that it outperforms the PPD estimator for some EVI groups of certain sample sizes.

Considering the results from Tables 4.7.16 to 4.7.27 (all sample sizes of size $n = 2000$ and larger, and all EVI groups), we note that the two best performing estimators are always GJ(ρ_3) and the PPD estimator. We therefore consider only these two estimators in this section. The performance of the other estimators can be obtained from Tables 4.7.16 to 4.7.27.

The PPD estimator is still the method of choice throughout. The only case where this might be an exception, is for samples sizes $n = 2000$ and $n = 5000$. For these sample sizes GJ(ρ_3) yields a lower MSE for the EVI group $\gamma = 1$, even though these differences are in both cases less than two standard errors.

We will now discuss the results in more detail. If we state that the estimators perform equally well, or differ marginally, this would mean that the difference between the MSEs is less than 0.5 standard errors.

Sample size $n = 2000$

For the EVI group $\gamma = 1$ the MSE obtained for the PPD estimator is 1.55 standard errors above that of GJ(ρ_3). For the EVI group $\gamma = 0.5$ the MSE of the GJ(ρ_3) is 2.18 standard errors above that of the PPD estimator. For the EVI group $\gamma = 0.1$ the two estimators perform equally well.

We suggest the following method of estimation:

Estimate the EVI using GJ(ρ_3) at $k = 45\%$ of n. If the resulting estimate is above 0.75, leave the estimate unchanged. Otherwise, redo the estimation by using the PPD estimator.

As an alternative to the PPD estimator, estimate the EVI using GJ(ρ_3) at $k = 45\%$ of n. (Technically the lowest MSE for the EVI group $\gamma = 0.5$ was obtained at $k = 40\%$ of n, but the difference is negligible.)

Sample size $n = 5000$

For the EVI group $\gamma = 1$ the MSE obtained from the PPD estimator is 1.67 standard errors above that of GJ(ρ_3). For the EVI group $\gamma = 0.5$ the MSE of the GJ(ρ_3) is 2.48 standard errors above that of the PPD estimator. For the EVI group $\gamma = 0.1$ the two estimators perform equally well.

We suggest the following method of estimation:

Estimate the EVI using GJ(ρ_3) at $k = 35\%$ of n. If the resulting estimate is above 0.75, leave the estimate unchanged. Otherwise, redo the estimation by using the PPD estimator.

As an alternative to the PPD estimator, estimate the EVI using GJ(ρ_3) at $k = 35\%$ of n. (Technically the lowest MSE was obtained at $k = 30\%$ of n for the EVI group $\gamma = 0.5$, and at $k = 75\%$ of n for the EVI group $\gamma = 0.1$, but the differences are negligible.)

Sample size $n = 10\,000$ and $n = 20\,000$

The PPD estimator yields the lowest MSE. (In some cases GJ(ρ_3) yields an MSE which is marginally lower than the MSE yielded by the PPD estimator.)

As an alternative to the PPD estimator, estimate the EVI using GJ(ρ_3) at $k = 25\%$ of n for $n = 10\,000$, and GJ(ρ_3) at $k = 15\%$ of n for $n = 20\,000$. (Technically the lowest MSE for $n = 10\,000$ and EVI group $\gamma = 0.1$ was obtained at $k = 20\%$ of n, but the difference is negligible.)

4.8.3 General comments

For estimators based on log-spacings of order statistics (used mainly as alternative estimators to the PPD estimator), we see that k as percentage of n is quite large. We even have $k = 95\%$ of n when determining the ML(ρ_3) estimate for sample sizes of size $n = 100$. This concurs with the observation by Gomes and Martins (2004) that the "estimators are quite stable and close to the target value of the tail index γ for a wide range of k-values." One can also see this property when considering the graphs of Section 4.9.2.

We also see that k as percentage of n decreases on average as n increases. This is a pattern which we observe from the simulation study, which has an important theoretical basis. When performing limiting operations in EVT to determine the limiting properties of estimators as $n \to \infty$, one also has to consider the limit of k. As $n \to \infty$ it is required that $k \to \infty$ and that $k/n \to 0$ (Beirlant et al., 2004).

The following table shows that the alternative estimators to the PPD estimator fulfill these requirements. We consider the alternative estimators for the case where $\gamma = 0.5$.

Sample size	***k* as % of *n***	***k***
50	90%	45
100	90%	90
200	85%	170
500	60%	300
1000	55%	550
2000	45%	900
5000	35%	1750
10 000	25%	2500
20 000	15%	3000

Table 4.8.3 The effect of the samples size on the optimal number of excesses.

Here we see that as n increases from $n = 50$ to $n = 20\,000$, k increases from 45 to 3000, and k/n decreases from 0.9 to 0.15.

4.8.4 Simulation study results for alternative estimators

We now present the simulation study results for the PPD estimator and the alternative estimator, as discussed in Sections 4.8.1 and 4.8.2.

The tables are in the same format as Tables 3.4.4.2 to 3.4.4.6, except that the RMSE columns have been removed. The columns which show the results for the alternative estimators, are indicated as “Alt”.

It is important to note that the PPD estimator results are not the same as those of Tables 3.4.4.2 to 3.4.4.6. For example, samples of size $n = 2000$ have the following description of the estimator involving the PPD:

Estimate the EVI using GJ(ρ_3) at $k = 45\%$ of n. If the resulting estimate is above 0.75, *leave the estimate unchanged. Otherwise, redo the estimation by using the PPD estimator*.

These difference only occur at samples of sizes $n = 2000$ and $n = 5000$. For the rest of the sample sizes the PPD results correspond to the PPD results in Tables 3.4.4.2 to 3.4.4.6. Refer to Sections 4.8.1 and 4.8.2.

The tables which follow show that the estimator involving the PPD performs consistently better than the alternative for a sample size of up to $n = 5000$. For larger samples the difference becomes negligible (a difference of less than one standard error), except for samples of size $n = 20\,000$ and EVI group $\gamma = 1$, where the PPD estimator still outperforms the alternative.

The use of these estimators in practice, will be illustrated in Chapter 5.

				n = 50		n = 100		n = 200	
Distr.		γ	ρ	**PPD**	**Alt**	**PPD**	**Alt**	**PPD**	**Alt**
1	Burr	0.1	-2	0.0677	0.1048	0.0521	0.0695	0.0405	0.0563
				(0.0023)	(0.0040)	(0.0024)	(0.0028)	(0.0022)	(0.0020)
2	Burr	0.1	-1	0.0411	0.0693	0.0274	0.0408	0.0212	0.0298
				(0.0016)	(0.0022)	(0.0017)	(0.0021)	(0.0011)	(0.0013)
3	Burr	0.1	-0.5	0.1812	0.0823	0.1177	0.0605	0.0833	0.0546
				(0.0064)	(0.0042)	(0.0063)	(0.0037)	(0.0041)	(0.0025)
4	Fréchet	0.1	-1	0.0491	0.0712	0.0350	0.0399	0.0259	0.0271
				(0.0018)	(0.0029)	(0.0019)	(0.0020)	(0.0014)	(0.0013)
5	logΓ	0.1	0	0.0406	0.0930	0.0325	0.0573	0.0286	0.0469
				(0.0019)	(0.0057)	(0.0024)	(0.0047)	(0.0022)	(0.0033)
γ **= 0.1**				**0.0759**	**0.0841**	**0.0529**	**0.0536**	**0.0399**	**0.0429**
				(0.0015)	(0.0018)	(0.0015)	(0.0014)	(0.0011)	(0.0010)
6	Burr	0.5	-2	1.8210	2.6404	1.4283	1.7593	0.9399	0.9959
				(0.0547)	(0.0737)	(0.0681)	(0.0735)	(0.0437)	(0.0465)
7	Burr	0.5	-1	0.9261	1.5461	0.7448	1.0779	0.5805	0.5639
				(0.0390)	(0.0575)	(0.0411)	(0.0497)	(0.0329)	(0.0296)
8	Burr	0.5	-0.5	3.8526	3.4017	2.3898	1.6273	1.6694	1.7229
				(0.1432)	(0.1624)	(0.1262)	(0.1052)	(0.0852)	(0.0897)
9	Fréchet	0.5	-1	1.3757	1.8538	0.9574	1.1116	0.7253	0.6637
				(0.0469)	(0.0678)	(0.0492)	(0.0582)	(0.0380)	(0.0366)
10	*t*	0.5	-1	0.8748	1.2659	0.5890	0.6534	0.4038	0.3620
				(0.0428)	(0.0520)	(0.0391)	(0.0350)	(0.0267)	(0.0213)
11	logΓ	0.5	0	0.9493	2.4944	0.6741	1.6565	0.5613	1.2179
				(0.0437)	(0.1450)	(0.0458)	(0.1330)	(0.0482)	(0.1026)
γ **= 0.5**				**1.6332**	**2.2004**	**1.1306**	**1.3143**	**0.8134**	**0.9211**
				(0.0293)	(0.0420)	(0.0280)	(0.0338)	(0.0202)	(0.0255)
12	Burr	1	-2	8.1259	10.1038	5.9039	5.9901	4.3406	3.0624
				(0.2331)	(0.3255)	(0.2499)	(0.2907)	(0.1824)	(0.1703)
13	Burr	1	-1	3.8879	4.9749	2.9743	3.0721	2.3573	2.0461
				(0.1571)	(0.2575)	(0.1608)	(0.1941)	(0.1260)	(0.1149)
14	Burr	1	-0.5	12.3702	18.6514	8.1449	9.3191	5.9909	6.4186
				(0.5180)	(0.6119)	(0.4228)	(0.4470)	(0.3539)	(0.3476)
15	Fréchet	1	-1	5.9964	7.1776	4.3039	4.0642	2.9867	2.1587
				(0.2113)	(0.2929)	(0.2047)	(0.2337)	(0.1754)	(0.1359)
16	*t*	1	-2	7.2427	11.2277	6.0318	8.3363	4.1087	4.1098
				(0.2185)	(0.3596)	(0.2482)	(0.3547)	(0.1772)	(0.1973)
17	logΓ	1	0	3.4014	9.0840	2.4120	7.2871	1.8962	7.0001
				(0.1525)	(0.6397)	(0.1645)	(0.5062)	(0.1386)	(0.4047)
γ **= 1**				**6.8374**	**10.2032**	**4.9618**	**6.3448**	**3.6134**	**4.1326**
				(0.1134)	(0.1804)	(0.1051)	(0.1452)	(0.0843)	(0.1033)
Mean				**3.0120**	**4.4025**	**2.1658**	**2.7190**	**1.5741**	**1.7963**
				(0.0413)	(0.0654)	(0.0384)	(0.0526)	(0.0306)	(0.0376)

Table 4.8.4.1 PPD and alternative estimator results for small samples.

				n = 500		*n* = 1000		*n* = 2000	
Distr.		γ	ρ	**PPD**	**Alt**	**PPD**	**Alt**	**PPD**	**Alt**
1	Burr	0.1	-2	0.0247	0.0199	0.0167	0.0116	0.0109	0.0045
				(0.0019)	(0.0018)	(0.0011)	(0.0009)	(0.0008)	(0.0004)
2	Burr	0.1	-1	0.0130	0.0194	0.0098	0.0144	0.0085	0.0073
				(0.0011)	(0.0015)	(0.0009)	(0.0010)	(0.0008)	(0.0007)
3	Burr	0.1	-0.5	0.0517	0.0163	0.0372	0.0135	0.0263	0.0128
				(0.0038)	(0.0017)	(0.0035)	(0.0012)	(0.0025)	(0.0011)
4	Fréchet	0.1	-1	0.0168	0.0138	0.0133	0.0077	0.0078	0.0038
				(0.0013)	(0.0014)	(0.0010)	(0.0007)	(0.0006)	(0.0004)
5	logΓ	0.1	0	0.0194	0.0533	0.0167	0.0503	0.0173	0.0433
				(0.0023)	(0.0041)	(0.0017)	(0.0032)	(0.0014)	(0.0025)
γ **= 0.1**				**0.0251**	**0.0245**	**0.0187**	**0.0195**	**0.0141**	**0.0143**
				(0.0010)	(0.0010)	(0.0009)	(0.0007)	(0.0006)	(0.0006)
6	Burr	0.5	-2	0.6359	0.4824	0.3401	0.3012	0.2247	0.1075
				(0.0476)	(0.0406)	(0.0242)	(0.0223)	(0.0162)	(0.0094)
7	Burr	0.5	-1	0.3950	0.5545	0.1955	0.3011	0.1601	0.1792
				(0.0392)	(0.0482)	(0.0194)	(0.0229)	(0.0152)	(0.0153)
8	Burr	0.5	-0.5	1.1387	0.4827	0.9762	0.3649	0.7235	0.3536
				(0.0834)	(0.0627)	(0.0782)	(0.0351)	(0.0647)	(0.0313)
9	Fréchet	0.5	-1	0.4377	0.3361	0.2352	0.1593	0.1467	0.1425
				(0.0333)	(0.0329)	(0.0200)	(0.0159)	(0.0116)	(0.0133)
10	*t*	0.5	-1	0.2722	0.7165	0.2468	0.4352	0.1578	0.2684
				(0.0252)	(0.0492)	(0.0295)	(0.0331)	(0.0163)	(0.0211)
11	logΓ	0.5	0	0.3691	1.2538	0.5327	1.2107	0.3982	0.9469
				(0.0364)	(0.0941)	(0.0401)	(0.0731)	(0.0234)	(0.0459)
γ **= 0.5**				**0.5414**	**0.6377**	**0.4211**	**0.4621**	**0.3018**	**0.3330**
				(0.0196)	(0.0237)	(0.0166)	(0.0158)	(0.0125)	(0.0106)
12	Burr	1	-2	2.5581	1.8152	1.5654	1.3826	0.5883	0.5883
				(0.1631)	(0.1658)	(0.1246)	(0.1318)	(0.0550)	(0.0550)
13	Burr	1	-1	1.5064	1.9388	0.7953	1.1565	0.4944	0.4944
				(0.1408)	(0.1827)	(0.0747)	(0.0902)	(0.0444)	(0.0444)
14	Burr	1	-0.5	4.0734	2.0942	3.7100	1.4385	1.4861	1.4861
				(0.2897)	(0.1885)	(0.2776)	(0.1303)	(0.1177)	(0.1177)
15	Fréchet	1	-1	1.6857	1.3088	0.9030	0.7196	0.5554	0.5554
				(0.1251)	(0.1271)	(0.0820)	(0.0722)	(0.0565)	(0.0565)
16	*t*	1	-2	2.6006	2.6263	1.5004	1.9022	0.8336	0.8336
				(0.1746)	(0.2114)	(0.1215)	(0.1605)	(0.0630)	(0.0630)
17	logΓ	1	0	1.5700	5.1164	2.0135	5.0781	3.7781	3.7781
				(0.1965)	(0.4266)	(0.1671)	(0.3128)	(0.2265)	(0.2265)
γ **= 1**				**2.3324**	**2.4833**	**1.7479**	**1.9462**	**1.2893**	**1.2893**
				(0.0773)	(0.0971)	(0.0640)	(0.0690)	(0.0463)	(0.0463)
Mean				**1.0217**	**1.1087**	**0.7710**	**0.8557**	**0.5657**	**0.5768**
				(0.0281)	(0.0353)	(0.0234)	(0.0250)	(0.0169)	(0.0168)

Table 4.8.4.2 PPD and alternative estimator results for medium sized samples.

				***n* = 5000**		***n* = 10 000**		***n* = 20 000**	
Distr.		γ	ρ	**PPD**	**Alt**	**PPD**	**Alt**	**PPD**	**Alt**
1	Burr	0.1	-2	0.0065	0.0023	0.0042	0.0016	0.0027	0.0012
				(0.0008)	(0.0003)	(0.0005)	(0.0002)	(0.0003)	(0.0002)
2	Burr	0.1	-1	0.0059	0.0037	0.0053	0.0024	0.0030	0.0018
				(0.0007)	(0.0006)	(0.0006)	(0.0003)	(0.0004)	(0.0002)
3	Burr	0.1	-0.5	0.0183	0.0100	0.0111	0.0064	0.0098	0.0045
				(0.0022)	(0.0011)	(0.0011)	(0.0008)	(0.0012)	(0.0005)
4	Fréchet	0.1	-1	0.0056	0.0027	0.0040	0.0017	0.0024	0.0013
				(0.0006)	(0.0004)	(0.0005)	(0.0002)	(0.0002)	(0.0002)
5	logΓ	0.1	0	0.0159	0.0335	0.0129	0.0248	0.0132	0.0208
				(0.0015)	(0.0025)	(0.0010)	(0.0016)	(0.0011)	(0.0014)
γ = 0.1				**0.0104**	**0.0105**	**0.0075**	**0.0074**	**0.0062**	**0.0059**
				(0.0006)	(0.0006)	(0.0003)	(0.0004)	(0.0004)	(0.0003)
6	Burr	0.5	-2	0.1095	0.0436	0.0914	0.0335	0.1063	0.0308
				(0.0114)	(0.0047)	(0.0098)	(0.0046)	(0.0157)	(0.0042)
7	Burr	0.5	-1	0.1008	0.0852	0.0793	0.0408	0.1383	0.0357
				(0.0116)	(0.0102)	(0.0098)	(0.0076)	(0.0311)	(0.0048)
8	Burr	0.5	-0.5	0.3770	0.2380	0.2843	0.1283	0.1101	0.1105
				(0.0433)	(0.0271)	(0.0343)	(0.0139)	(0.0129)	(0.0114)
9	Fréchet	0.5	-1	0.1172	0.0578	0.0843	0.0434	0.1142	0.0335
				(0.0120)	(0.0083)	(0.0089)	(0.0062)	(0.0128)	(0.0054)
10	*t*	0.5	-1	0.1356	0.1503	0.1036	0.0971	0.0970	0.0539
				(0.0197)	(0.0169)	(0.0128)	(0.0112)	(0.0119)	(0.0080)
11	logΓ	0.5	0	0.4582	0.8893	0.3863	0.6970	0.2668	0.5580
				(0.0336)	(0.0568)	(0.0269)	(0.0432)	(0.0301)	(0.0381)
γ = 0.5				**0.2164**	**0.2441**	**0.1715**	**0.1733**	**0.1388**	**0.1371**
				(0.0103)	(0.0111)	(0.0081)	(0.0080)	(0.0085)	(0.0069)
12	Burr	1	-2	0.3214	0.3214	0.3659	0.1383	0.3776	0.1212
				(0.0367)	(0.0367)	(0.0376)	(0.0163)	(0.0503)	(0.0175)
13	Burr	1	-1	0.3297	0.3297	0.3506	0.2033	0.3830	0.1374
				(0.0496)	(0.0496)	(0.0386)	(0.0303)	(0.0473)	(0.0183)
14	Burr	1	-0.5	0.8829	0.8829	1.1300	0.6296	0.4330	0.4039
				(0.0989)	(0.0989)	(0.1286)	(0.0696)	(0.0562)	(0.0508)
15	Fréchet	1	-1	0.2301	0.2301	0.2354	0.1958	0.5100	0.1860
				(0.0329)	(0.0329)	(0.0257)	(0.0298)	(0.0556)	(0.0285)
16	*t*	1	-2	0.3678	0.3678	0.4183	0.1982	0.3610	0.1289
				(0.0519)	(0.0519)	(0.0435)	(0.0275)	(0.0450)	(0.0143)
17	logΓ	1	0	3.0297	3.0297	1.5004	2.7817	1.0225	2.3387
				(0.2088)	(0.2088)	(0.1069)	(0.1760)	(0.0883)	(0.1215)
γ = 1				**0.8603**	**0.8603**	**0.6668**	**0.6912**	**0.5145**	**0.5527**
				(0.0412)	(0.0412)	(0.0305)	(0.0328)	(0.0241)	(0.0230)
Mean				**0.3831**	**0.3928**	**0.2981**	**0.3073**	**0.2324**	**0.2452**
				(0.0150)	(0.0150)	(0.0111)	(0.0119)	(0.0090)	(0.0085)

Table 4.8.4.3 PPD and alternative estimator results for large samples.

The results presented above conclude the main results concerning the PPD estimator and its alternatives. In Section 4.9 additional results will be given, which may be more interesting from a theoretical rather than a practical point of view.

4.9 Further comments on the bias and the second order parameter

In this section we will show some results concerning the bias of estimators, and the second order parameter ρ. From a practical application point of view, this section can be skipped without lack of continuity.

Section 4.9.1 presents tables which contain mean biases of the estimators discussed in Sections 4.8.1 and 4.8.2, as calculated from the simulation results. Biases will be represented graphically in Section 4.9.2, showing biases as a function of k for the PPD estimator and the GJ(ρ_3) estimator, for the EVI group $\gamma = 0.5$. This chapter will be concluded with the justification for choosing $-2 \leq \rho \leq 0$ as the range of the second order parameter for the distributions from which we simulate data, by considering estimates of ρ for real data.

4.9.1 Bias tables

In the tables which follow we present the estimated biases obtained from the simulation study for the estimators mentioned in Section 4.8. Only the estimators mentioned in Section 4.8, and only at the thresholds which are specified there, are considered.

In the tables we give the average bias, below it the 100×MSE, and below that the standard error in brackets. See for example Table 4.9.1.1. We include the 100×MSEs and their standard errors for the sake of completeness and ease of comparison. Also, the tables of Section 4.7 showed only the fixed thresholds at which the minimum MSEs were obtained. These thresholds were not necessarily those used to define an alternative estimator to the PPD. For example, for a sample of size $n = 5000$, we stated:

As an alternative to the PPD estimator, estimate the EVI using GJ(ρ_3) at $k = 35\%$ of n. (Technically the lowest MSE was obtained at $k = 30\%$ of n for the EVI group $\gamma = 0.5$, and at $k = 75\%$ of n for the EVI group $\gamma = 0.1$, but the differences are negligible.)

The tables in Section 4.7 showed for $\gamma = 0.1$ the results at $k = 75\%$ of n, and for $\gamma = 0.5$ the results for $k = 30\%$ of n. In the tables of this section the results for $k = 35\%$ of n can be obtained for these two EVI groups.

The calculation of the bias is straightforward. For example, consider the value of -0.0187 in the sixth row of the fourth column in Table 4.9.1.1. The PPD estimates of the EVI were calculated for the 1000 samples of size $n = 50$ from the *Burr*($\beta = 1, \gamma = 0.1, \rho = -2$) distribution. The average of these estimates was calculated, and the true value of the EVI ($\gamma = 0.1$) subtracted from this average. The value obtained was -0.0187. This means that the PPD estimator estimates the EVI of a *Burr*$(\beta = 1, \gamma = 0.1, \rho = -2)$ distribution for samples of size $n = 50$ too low by an average amount of approximately 0.0187.

Note that no samples were generated for the *t* distribution with $\gamma = 0.1$, hence the open spaces in the tables. Also, no value ρ is specified for the *t* distribution, since it differs for different values of γ. For $\gamma = 0.5$, $\rho = -1$, and for $\gamma = 1$, $\rho = -2$.

The mean of the bias (final column) is calculated as the mean of the absolute values of the biases across the distributions.

			Burr	Burr	Burr	Fréchet	t	logΓ	Mean
		ρ	**-2**	**-1**	**-0.5**	**-1**		**0**	
n	γ	**Estimator**							
50	**0.1**	**PPD**	-0.0187	-0.0044	0.0354	-0.0133		0.0024	0.0148
			0.0677	0.0411	0.1812	0.0491		0.0406	**0.0759**
			(0.0023)	(0.0016)	(0.0064)	(0.0018)		(0.0019)	(0.0015)
		Peng	-0.0220	-0.0159	0.0166	-0.0132		0.0095	0.0154
		90%	0.1048	0.0693	0.0823	0.0712		0.0930	**0.0841**
			(0.0040)	(0.0022)	(0.0042)	(0.0029)		(0.0057)	(0.0018)
	0.5	**PPD**	-0.1059	-0.0305	0.1564	-0.0824	0.0172	-0.0021	0.0658
			1.8210	0.9261	3.8526	1.3757	0.8748	0.9493	**1.6332**
			(0.0547)	(0.0390)	(0.1432)	(0.0469)	(0.0428)	(0.0437)	(0.0293)
		Peng	-0.1151	-0.0753	0.0837	-0.0725	-0.0521	0.0503	0.0748
		90%	2.6398	1.5400	2.1611	1.8527	1.2246	2.4857	**1.9840**
			(0.0737)	(0.0567)	(0.1052)	(0.0677)	(0.0475)	(0.1439)	(0.0362)
	1	**PPD**	-0.2356	-0.0949	0.2725	-0.1799	-0.2114	-0.0392	0.1723
			8.1259	3.8879	12.3702	5.9964	7.2427	3.4014	**6.8374**
			(0.2331)	(0.1571)	(0.5180)	(0.2113)	(0.2185)	(0.1525)	(0.1134)
		GJ	-0.1958	-0.0607	0.3766	-0.1046	-0.2011	0.1038	0.1738
		90%	8.1286	3.6858	18.6514	6.2978	7.4470	8.9215	**8.8553**
			(0.2629)	(0.1871)	(0.6119)	(0.2612)	(0.2342)	(0.6385)	(0.1674)
100	**0.1**	**PPD**	-0.0168	-0.0058	0.0282	-0.0119		0.0038	0.0133
			0.0521	0.0274	0.1177	0.0350		0.0325	**0.0529**
			(0.0024)	(0.0017)	(0.0063)	(0.0019)		(0.0024)	(0.0015)
		Peng	-0.0205	-0.0141	0.0175	-0.0114		0.0103	0.0148
		90%	0.0695	0.0408	0.0605	0.0399		0.0573	**0.0536**
			(0.0028)	(0.0021)	(0.0037)	(0.0020)		(0.0047)	(0.0014)
	0.5	**PPD**	-0.0912	-0.0395	0.1261	-0.0614	0.0021	0.0133	0.0556
			1.4283	0.7448	2.3898	0.9574	0.5890	0.6741	**1.1306**
			(0.0681)	(0.0411)	(0.1262)	(0.0492)	(0.0391)	(0.0458)	(0.0280)
		Peng	-0.1011	-0.0750	0.0868	-0.0486	-0.0434	0.0581	0.0688
		90%	1.7593	1.0738	1.4943	1.1095	0.6534	1.4915	**1.2636**
			(0.0735)	(0.0489)	(0.0921)	(0.0578)	(0.0350)	(0.1189)	(0.0312)
	1	**PPD**	-0.1969	-0.0906	0.2266	-0.1581	-0.2018	0.0139	0.1480
			5.9039	2.9743	8.1449	4.3039	6.0318	2.4120	**4.9618**
			(0.2499)	(0.1608)	(0.4228)	(0.2047)	(0.2482)	(0.1645)	(0.1051)
		ML(ρ_3)	-0.1600	-0.0803	0.2570	-0.0645	-0.2129	0.1259	0.1501
		95%	5.0854	2.7664	9.3191	3.8389	6.4751	7.8359	**5.8868**
			(0.2280)	(0.1619)	(0.4470)	(0.2134)	(0.2470)	(0.5403)	(0.1371)

Table 4.9.1.1 Bias calculations for samples of size $n = 50$ and $n = 100$.

			Burr	Burr	Burr	Fréchet	*t*	logΓ	Mean
		ρ	**-2**	**-1**	**-0.5**	**-1**		**0**	
n	γ	**Estimator**							
200	**0.1**	**PPD**	-0.0138	-0.0061	0.0237	-0.0098		0.0091	0.0125
			0.0405	0.0212	0.0833	0.0259		0.0286	**0.0399**
			(0.0022)	(0.0011)	(0.0041)	(0.0014)		(0.0022)	(0.0011)
		Peng	-0.0198	-0.0134	0.0199	-0.0103		0.0145	0.0156
		90%	0.0563	0.0298	0.0546	0.0271		0.0469	**0.0429**
			(0.0020)	(0.0013)	(0.0025)	(0.0013)		(0.0033)	(0.0010)
	0.5	**PPD**	-0.0775	-0.0373	0.1067	-0.0559	-0.0125	0.0273	0.0529
			0.9399	0.5805	1.6694	0.7253	0.4038	0.5613	**0.8134**
			(0.0437)	(0.0329)	(0.0852)	(0.0380)	(0.0267)	(0.0482)	(0.0202)
		Peng	-0.0770	-0.0446	0.1091	-0.0397	-0.0274	0.0646	0.0604
		85%	0.9921	0.5639	1.6264	0.6637	0.3620	1.0941	**0.8837**
			(0.0458)	(0.0296)	(0.0766)	(0.0366)	(0.0213)	(0.0869)	(0.0225)
	1	**PPD**	-0.1717	-0.0911	0.1920	-0.1216	-0.1667	0.0389	0.1303
			4.3406	2.3573	5.9909	2.9867	4.1087	1.8962	**3.6134**
			(0.1824)	(0.1260)	(0.3539)	(0.1754)	(0.1772)	(0.1386)	(0.0843)
		ML(ρ_3)	-0.1020	-0.0628	0.2095	-0.0081	-0.1500	0.1789	0.1185
		75%	2.8347	2.0451	6.4186	2.1496	3.7752	7.0136	**4.0395**
			(0.1474)	(0.1122)	(0.3476)	(0.1352)	(0.1722)	(0.4044)	(0.1009)
500	**0.1**	**PPD**	-0.0122	-0.0058	0.0199	-0.0085		0.0082	0.0109
			0.0247	0.0130	0.0517	0.0168		0.0194	**0.0251**
			(0.0019)	(0.0011)	(0.0038)	(0.0013)		(0.0023)	(0.0010)
		GJ(ρ_3)	-0.0075	-0.0090	0.0081	-0.0005		0.0181	0.0086
		65%	0.0199	0.0194	0.0163	0.0138		0.0533	**0.0245**
			(0.0018)	(0.0015)	(0.0017)	(0.0014)		(0.0041)	(0.0010)
	0.5	**PPD**	-0.0607	-0.0368	0.0926	-0.0489	-0.0241	0.0377	0.0501
			0.6359	0.3950	1.1387	0.4377	0.2722	0.3691	**0.5414**
			(0.0476)	(0.0392)	(0.0834)	(0.0333)	(0.0252)	(0.0364)	(0.0196)
		GJ(ρ_3)	-0.0302	-0.0433	0.0394	-0.0044	-0.0709	0.0916	0.0466
		60%	0.4824	0.5545	0.4751	0.3361	0.7165	1.2499	**0.6358**
			(0.0406)	(0.0482)	(0.0590)	(0.0329)	(0.0492)	(0.0929)	(0.0233)
	1	**PPD**	-0.1316	-0.0792	0.1702	-0.0971	-0.1326	0.0663	0.1128
			2.5581	1.5064	4.0734	1.6857	2.6006	1.5700	**2.3324**
			(0.1631)	(0.1408)	(0.2897)	(0.1251)	(0.1746)	(0.1965)	(0.0773)
		GJ(ρ_3)	-0.0477	-0.0688	0.0824	-0.0001	-0.0922	0.1721	0.0772
		55%	1.8492	1.9926	2.0942	1.3077	2.5900	5.1162	**2.4917**
			(0.1712)	(0.1911)	(0.1885)	(0.1267)	(0.2058)	(0.4266)	(0.0974)

Table 4.9.1.2 Bias calculations for samples of size $n = 200$ and $n = 500$.

			Burr	Burr	Burr	Fréchet	*t*	logΓ	Mean
		ρ	**-2**	**-1**	**-0.5**	**-1**		**0**	
n	γ	**Estimator**							
1000	**0.1**	**PPD**	-0.0098	-0.0062	0.0145	-0.0082		0.0092	0.0096
			0.0167	0.0098	0.0372	0.0133		0.0167	**0.0187**
			(0.0011)	(0.0009)	(0.0035)	(0.0010)		(0.0017)	(0.0009)
		GJ(ρ_3)	-0.0056	-0.0089	0.0076	-0.0004		0.0195	0.0084
		60%	0.0116	0.0144	0.0135	0.0077		0.0503	**0.0195**
			(0.0009)	(0.0010)	(0.0012)	(0.0007)		(0.0032)	(0.0007)
	0.5	**PPD**	-0.0387	-0.0176	0.0770	-0.0337	-0.0127	0.0612	0.0402
			0.3401	0.1955	0.9762	0.2352	0.2468	0.5327	**0.4211**
			(0.0242)	(0.0194)	(0.0782)	(0.0200)	(0.0295)	(0.0401)	(0.0166)
		GJ(ρ_3)	-0.0240	-0.0342	0.0408	-0.0049	-0.0524	0.0975	0.0423
		60%	0.3012	0.3011	0.3649	0.1593	0.4352	1.2107	**0.4621**
			(0.0223)	(0.0229)	(0.0351)	(0.0159)	(0.0331)	(0.0731)	(0.0158)
	1	**PPD**	-0.0907	-0.0394	0.1492	-0.0586	-0.0869	0.1161	0.0902
			1.5654	0.7953	3.7100	0.9030	1.5004	2.0135	**1.7479**
			(0.1246)	(0.0747)	(0.2776)	(0.0820)	(0.1215)	(0.1671)	(0.0640)
		GJ(ρ_3)	-0.0515	-0.0674	0.0853	0.0024	-0.0935	0.1990	0.0832
		55%	1.3826	1.1565	1.4385	0.7196	1.9022	5.0781	**1.9462**
			(0.1318)	(0.0902)	(0.1303)	(0.0722)	(0.1605)	(0.3128)	(0.0690)
2000	**0.1**	**PPD**	-0.0084	-0.0057	0.0117	-0.0069		0.0108	0.0087
			0.0109	0.0085	0.0263	0.0078		0.0173	**0.0141**
			(0.0008)	(0.0008)	(0.0025)	(0.0006)		(0.0014)	(0.0006)
		GJ(ρ_3)	-0.0022	-0.0040	0.0090	-0.0003		0.0188	0.0068
		45%	0.0045	0.0073	0.0128	0.0038		0.0433	**0.0143**
			(0.0004)	(0.0007)	(0.0011)	(0.0004)		(0.0025)	(0.0006)
	0.5	**PPD**	-0.0381	-0.0221	0.0680	-0.0210	-0.0175	0.0567	0.0372
			0.2247	0.1601	0.7235	0.1467	0.1578	0.3982	**0.3018**
			(0.0162)	(0.0152)	(0.0647)	(0.0116)	(0.0163)	(0.0234)	(0.0125)
		GJ(ρ_3)	-0.0116	-0.0201	0.0476	0.0074	-0.0405	0.0900	0.0362
		45%	0.1075	0.1792	0.3536	0.1425	0.2684	0.9469	**0.3330**
			(0.0094)	(0.0153)	(0.0313)	(0.0133)	(0.0211)	(0.0459)	(0.0106)
	1	**PPD**	-0.0876	-0.0402	0.1422	-0.0593	-0.0871	0.1083	0.0875
			1.1343	0.5337	2.9970	0.7780	1.1532	1.5714	**1.3613**
			(0.0675)	(0.0511)	(0.2597)	(0.0624)	(0.0742)	(0.1208)	(0.0523)
		GJ(ρ_3)	-0.0319	-0.0321	0.0998	0.0026	-0.0559	0.1763	0.0665
		45%	0.5883	0.4944	1.4861	0.5554	0.8336	3.7781	**1.2893**
			(0.0550)	(0.0444)	(0.1177)	(0.0565)	(0.0630)	(0.2265)	(0.0463)

Table 4.9.1.3 Bias calculations for samples of size $n = 1000$ and $n = 2000$.

			Burr	Burr	Burr	Fréchet	t	logΓ	Mean
		ρ	**-2**	**-1**	**-0.5**	**-1**		**0**	
n	γ	**Estimator**							
5000	**0.1**	**PPD**	-0.0064	-0.0049	0.0107	-0.0054		0.0112	0.0077
			0.0065	0.0059	0.0183	0.0056		0.0159	**0.0104**
			(0.0008)	(0.0007)	(0.0022)	(0.0006)		(0.0015)	(0.0006)
		GJ(ρ_3)	-0.0011	-0.0021	0.0084	0.0006		0.0170	0.0058
		35%	0.0023	0.0037	0.0100	0.0027		0.0335	**0.0105**
			(0.0003)	(0.0006)	(0.0011)	(0.0004)		(0.0025)	(0.0006)
	0.5	**PPD**	-0.0250	-0.0196	0.0478	-0.0270	-0.0151	0.0630	0.0329
			0.1095	0.1008	0.3770	0.1172	0.1356	0.4582	**0.2164**
			(0.0114)	(0.0116)	(0.0433)	(0.0120)	(0.0197)	(0.0336)	(0.0103)
		GJ(ρ_3)	0.0007	-0.0093	0.0382	-0.0020	-0.0240	0.0885	0.0271
		35%	0.0436	0.0852	0.2380	0.0578	0.1503	0.8893	**0.2441**
			(0.0047)	(0.0102)	(0.0271)	(0.0083)	(0.0169)	(0.0568)	(0.0111)
	1	**PPD**	-0.0737	-0.0403	0.1042	-0.0558	-0.0661	0.1090	0.0749
			0.7888	0.4069	1.7775	0.5022	0.7132	1.3852	**0.9289**
			(0.0761)	(0.0541)	(0.2028)	(0.0508)	(0.0724)	(0.1085)	(0.0439)
		GJ(ρ_3)	-0.0200	-0.0126	0.0746	0.0002	-0.0219	0.1635	0.0488
		35%	0.3214	0.3297	0.8829	0.2301	0.3678	3.0297	**0.8603**
			(0.0367)	(0.0496)	(0.0989)	(0.0329)	(0.0519)	(0.2088)	(0.0412)
10 000	**0.1**	**PPD**	-0.0052	-0.0057	0.0084	-0.0050		0.0106	0.0070
			0.0042	0.0053	0.0111	0.0040		0.0129	**0.0075**
			(0.0005)	(0.0006)	(0.0011)	(0.0005)		(0.0010)	(0.0003)
		GJ(ρ_3)	-0.0006	-0.0018	0.0063	0.0000		0.0148	0.0047
		25%	0.0016	0.0024	0.0064	0.0017		0.0248	**0.0074**
			(0.0002)	(0.0003)	(0.0008)	(0.0002)		(0.0016)	(0.0004)
	0.5	**PPD**	-0.0250	-0.0218	0.0426	-0.0217	-0.0201	0.0586	0.0316
			0.0914	0.0793	0.2843	0.0843	0.1036	0.3863	**0.1715**
			(0.0098)	(0.0098)	(0.0343)	(0.0089)	(0.0128)	(0.0269)	(0.0080)
		GJ(ρ_3)	-0.0029	-0.0074	0.0280	0.0028	-0.0189	0.0793	0.0232
		25%	0.0335	0.0408	0.1283	0.0434	0.0971	0.6970	**0.1733**
			(0.0046)	(0.0076)	(0.0139)	(0.0062)	(0.0112)	(0.0432)	(0.0080)
	1	**PPD**	-0.0479	-0.0450	0.0857	-0.0345	-0.0510	0.1157	0.0633
			0.3659	0.3506	1.1300	0.2354	0.4183	1.5004	**0.6668**
			(0.0376)	(0.0386)	(0.1286)	(0.0257)	(0.0435)	(0.1069)	(0.0305)
		GJ(ρ_3)	-0.0028	-0.0150	0.0591	0.0138	-0.0064	0.1584	0.0426
		25%	0.1383	0.2033	0.6296	0.1958	0.1982	2.7817	**0.6912**
			(0.0163)	(0.0303)	(0.0696)	(0.0298)	(0.0275)	(0.1760)	(0.0328)

Table 4.9.1.4 Bias calculations for samples of size $n = 5000$ and $n = 10\,000$.

			Burr	Burr	Burr	Fréchet	*t*	logΓ	Mean
		ρ	**-2**	**-1**	**-0.5**	**-1**		**0**	
n	γ	**Estimator**							
20 000	**0.1**	**PPD**	-0.0040	-0.0042	0.0079	-0.0040		0.0108	0.0062
			0.0027	0.0030	0.0098	0.0024		0.0132	**0.0062**
			(0.0003)	(0.0004)	(0.0012)	(0.0002)		(0.0011)	(0.0004)
		GJ(ρ_3)	-0.0003	-0.0005	0.0055	0.0001		0.0136	0.0040
		15%	0.0012	0.0018	0.0045	0.0013		0.0208	**0.0059**
			(0.0002)	(0.0002)	(0.0005)	(0.0002)		(0.0014)	(0.0003)
	0.5	**PPD**	-0.0111	-0.0225	0.0248	-0.0201	-0.0215	0.0439	0.0240
			0.1063	0.1383	0.1101	0.1142	0.0970	0.2668	**0.1388**
			(0.0157)	(0.0311)	(0.0129)	(0.0128)	(0.0119)	(0.0301)	(0.0085)
		GJ(ρ_3)	0.0004	-0.0022	0.0235	0.0000	-0.0082	0.0705	0.0175
		15%	0.0308	0.0357	0.1105	0.0335	0.0539	0.5580	**0.1371**
			(0.0042)	(0.0048)	(0.0114)	(0.0054)	(0.0080)	(0.0381)	(0.0069)
	1	**PPD**	-0.0254	-0.0457	0.0496	-0.0448	-0.0250	0.0920	0.0471
			0.3776	0.3830	0.4330	0.5100	0.3610	1.0225	**0.5145**
			(0.0503)	(0.0473)	(0.0562)	(0.0556)	(0.0450)	(0.0883)	(0.0241)
		GJ(ρ_3)	0.0017	0.0016	0.0467	-0.0003	0.0048	0.1476	0.0338
		15%	0.1212	0.1374	0.4039	0.1860	0.1289	2.3387	**0.5527**
			(0.0175)	(0.0183)	(0.0508)	(0.0285)	(0.0143)	(0.1215)	(0.0230)

Table 4.9.1.5 Bias calculations for samples of size $n = 20\,000$.

The average bias of the PPD and that of the alternative estimators do not differ too greatly. The most significant difference is at a sample size of $n = 5000$ and EVI group $\gamma = 1$, where the average bias of the PPD is 0.0749, and that of GJ(ρ_3) is 0.0488; a difference of just over 50%.

Also note that the average bias of GJ(ρ_3) is consistently lower than that of the PPD estimator for all samples of size $n = 2000$ and larger.

Other patterns can be observed when considering the sign of the bias for individual distributions. The PPD estimator consistently has a negative bias for the Burr distribution with $\rho = -2$, the Burr distribution with $\rho = -1$, and the Frechét distribution. The PPD estimator also consistently has a positive bias for the Burr distribution with $\rho = -0.5$. These same patterns occur for the Peng estimator.

As far as GJ(ρ_3) is concerned, it has consistently a positive bias for the Burr distribution with $\rho = -0.5$, and for the loggamma distribution.

Even though much more information can be extracted from these results, we will not discuss these results any further.

4.9.2 Bias graphs

In this section we represent the estimated average bias of some estimators graphically.

In Section 4.9.1 we calculated the average bias at certain thresholds, which were found to be optimal in the sense described in Section 4.8. We now give graphical representations of the bias of some estimators across all thresholds considered, namely at $k = 5\%, 10\%, \dots, 95\%$ of n.

We only consider distributions from the EVI group $\gamma = 0.5$. (The other EVI groups yield results with similar patterns.)

An excerpt from Table 3.3.2.5 shows the parameters of the distributions from the EVI group $\gamma = 0.5$:

No.	**Distribution**	$\boldsymbol{\gamma}$	$\boldsymbol{\rho}$
6	$Burr(\beta = 1, \gamma = 0.5, \rho = -2)$	0.5	-2
7	$Burr(\beta = 1, \gamma = 0.5, \rho = -1)$	0.5	-1
8	$Burr(\beta = 1, \gamma = 0.5, \rho = -0.5)$	0.5	-0.5
9	*Fréchet*(2)	0.5	-1
10	$\lvert t_2 \rvert$	0.5	-1
11	$\log\Gamma(\lambda = 2, \alpha = 2)$	0.5	0

Table 4.9.2 List of distributions considered for bias graphs.

The graphs below show the following:

- The titles of the individual graphs state the number of the distribution, the family and the value of the ρ.
- The true value of the EVI (0.5) is indicated on each of the graphs.
- The sections of the graph pertaining to the PPD estimator are represented by solid lines. For $n = 50$ and $n = 500$ we used the mode of the Gibbs draws, and for $n = 5000$ we used the median of the Gibbs draws. (Refer to Section 3.4.4.)
- The median of the thresholds obtained for the PPD estimator, as described in Section 3.4.4, is indicated by a solid vertical line.
- The sections of the graph pertaining to the GJ(ρ_3) estimator are represented by dashed lines.
- The optimal (fixed) threshold for the GJ(ρ_3) estimator is indicated by a dashed vertical line. This optimal threshold is 70% for $n = 50$ (from Table 4.7.2), 60% for $n = 500$ (from Table 4.9.1.2), and 35% for $n = 5000$ (from Table 4.9.1.4).
- The mean of the Hill estimates are indicated by a dotted line.

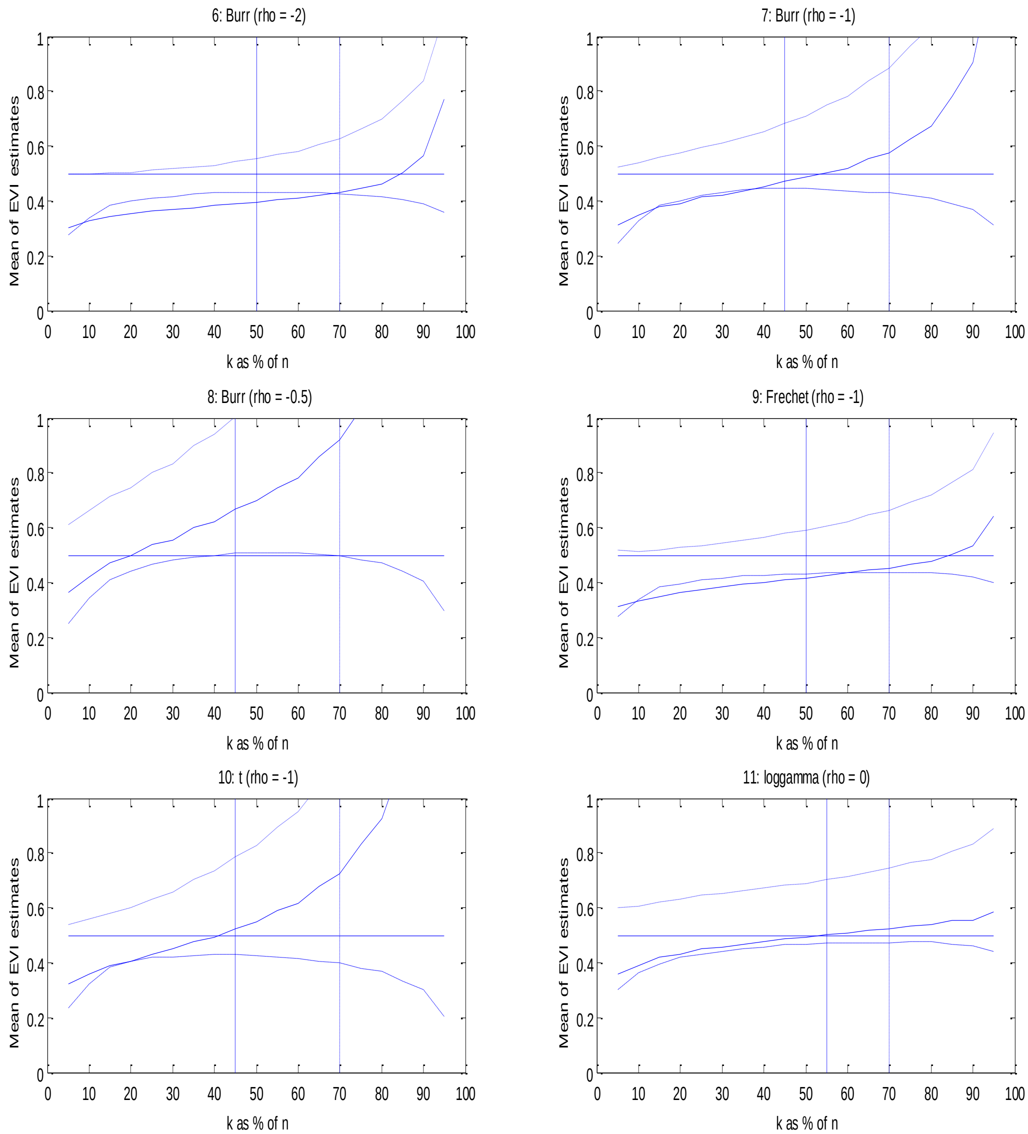

Figure 4.9.2.1 Bias graphs for 1000 samples of size $n = 50$.

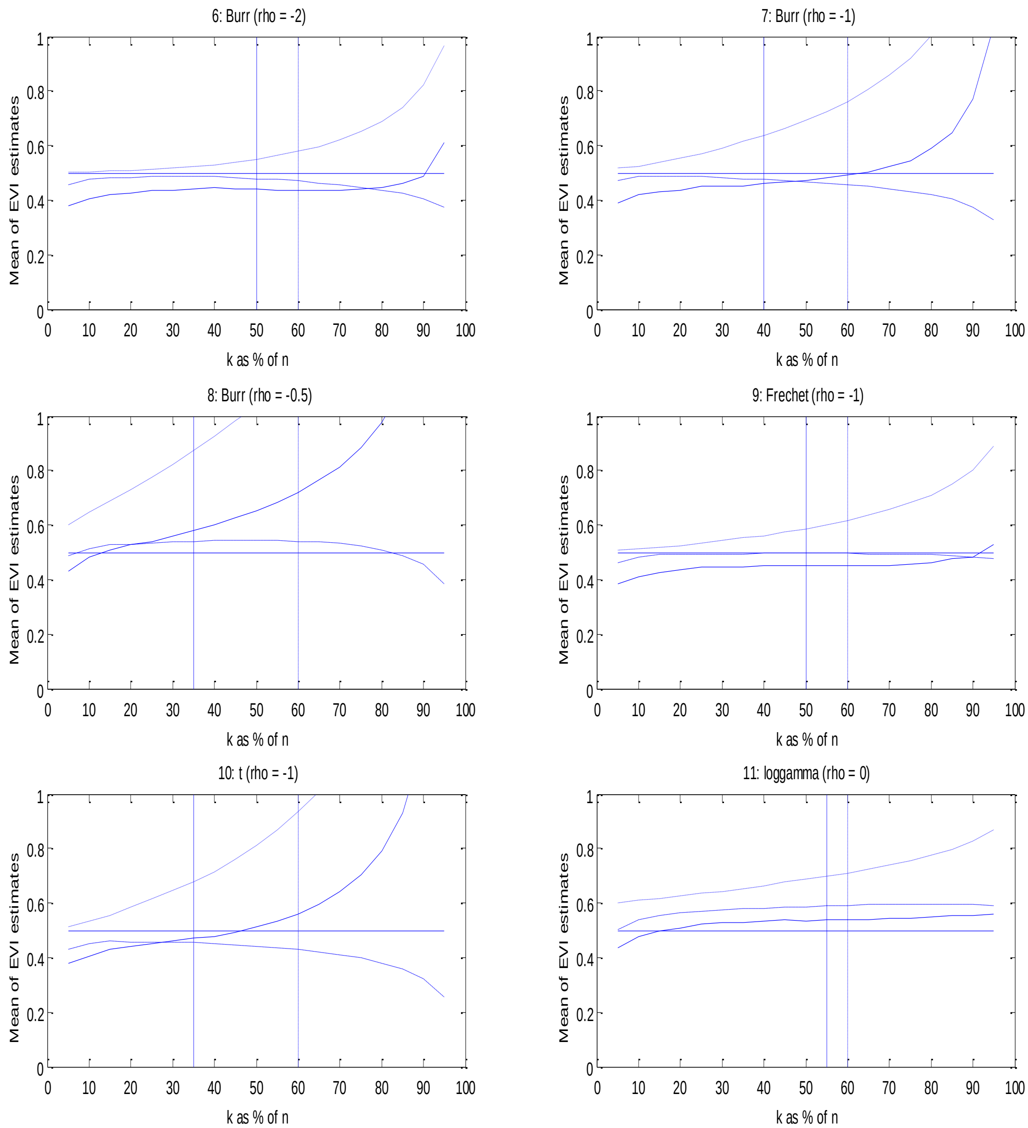

Figure 4.9.2.2 Bias graphs for 200 samples of size $n = 500$.

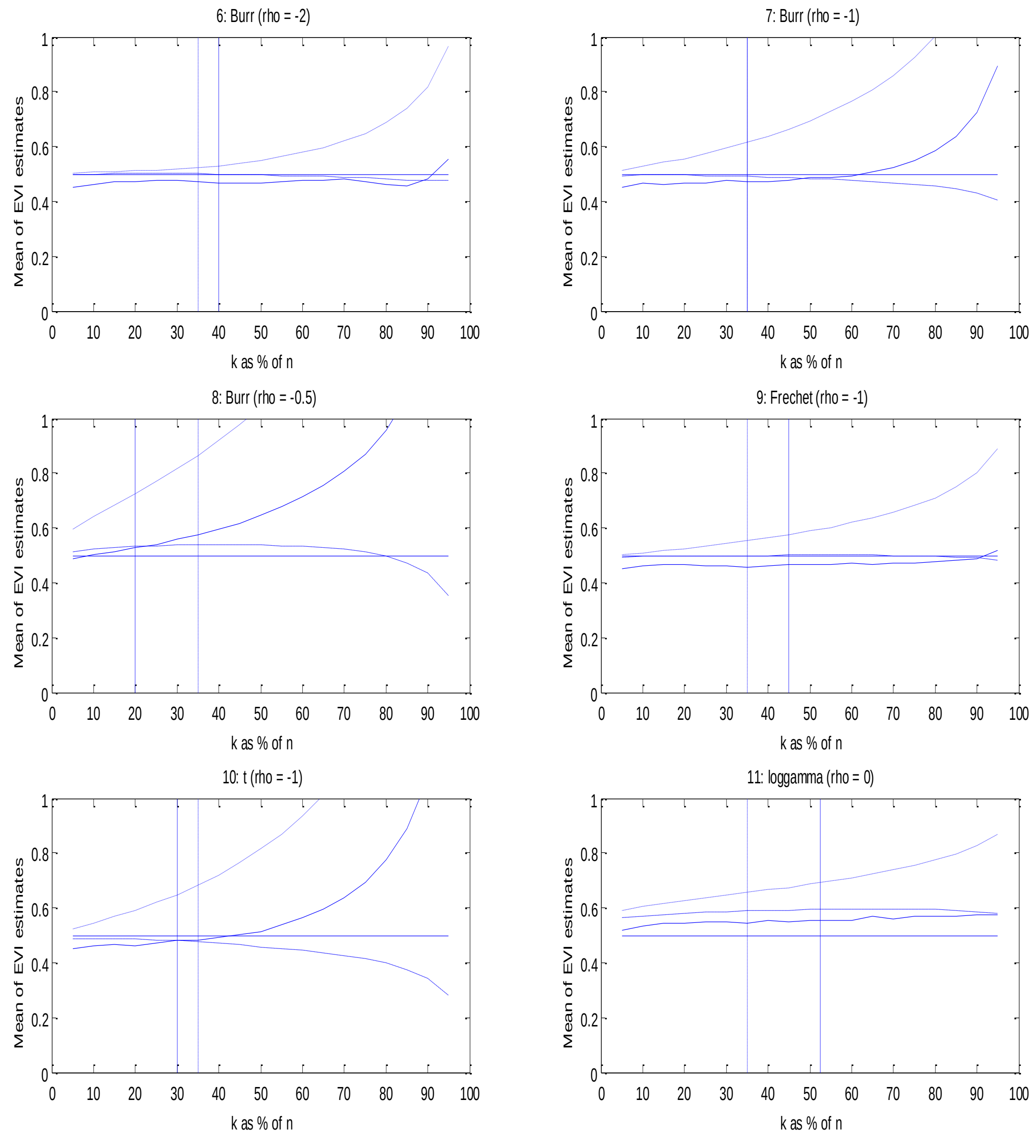

Figure 4.9.2.3 Bias graphs for 100 samples of size $n = 5000$.

We will not discuss these results in great detail, but only make the following remarks:

1. The GJ(ρ_3) estimator seems to be more stable than the PPD estimator, in the sense that the GJ(ρ_3) line is much flatter over a greater interval of thresholds. (One must however keep in mind that the graphs do not give us an indication of the variance of the estimates.)
2. The threshold selection seems to be vital in the performance of the PPD estimator in cases where the mean of the PPD estimates increase rapidly as a function of k, for instance Distributions 8 and 10.

3. From Figure 4.9.2.1 (samples of size $n = 50$) one may conclude that there is not much difference between the GJ(ρ_3) estimator and the PPD estimator, as far as accuracy of estimation is concerned. However, we have seen in Table 4.7.2 that the average 100×MSE for the PPD estimator is 1.6332 (with a standard error of 0.0293), and that the average 100×MSE of the GJ(ρ_3) estimator is more than twice that, namely 3.6623. The variance (as pointed out in Remark 1 above), might not be the only reason. One can see that the GJ(ρ_3) estimator has a relatively constant bias as a function of k. The PPD estimator usually estimates too low on average for small k and too high on average for large k. The benchmark PPD yields an estimate of the EVI which is obtained by averaging over a large range of k. Therefore, even though the PPD estimator is not on average as stable as the GJ(ρ_3) estimator, the fact that it estimates too low for some k and too high for other k, may be to its advantage, because of the averaging which occurs.
4. The PPD fares very poorly relative to the GJ(ρ_3) estimator for Distribution 8 (Burr with $\rho = -0.5$), if one considers the entire range of k.

4.9.3 Range of the second order parameter

In this section we provide a justification for choosing $-2 \leq \rho \leq 0$ as the range of the second order parameter for the distributions from which we simulate data.

We know that $\rho \leq 0$ according to the definition of ρ (Definition 2.4.4.2). We wish to choose the lower limit of ρ in such a way that it is a realistic or reasonable lower limit as far as real data are concerned.

With respect to the estimation of ρ, we reviewed the following papers:

- Beirlant, Joossens and Segers (2009)
- Fraga Alves (2002)
- Fraga Alves, Gomes and De Haan (2003)
- Goegebeur, Beirlant and De Wet (2008)
- Gomes, Martins and Neves (2000)
- Gomes and Martins (2002)
- Gomes, De Haan and Peng (2002)
- Gomes, Caeiro and Figueiredo (2004)
- Gomes and Martins (2004)
- Gomes, Martins and Neves (2007)

Only three of the papers mentioned above show estimates of ρ for real (case study) data. Fraga Alves (2002) calculates estimates of -1.9973, -0.4989 and -1.0609 of ρ for fire insurance claims data. Fraga Alves (2002) also shows various estimators over a wide range of values of k. From these graphs it seems as though -2 is the lower limit for these estimates for that specific data set. Gomes, Caeiro and Figueiredo (2004), estimated the value of ρ as -0.69 for exchange rate data. Gomes, Martins and Neves (2007) obtained an estimate of -0.65, also for exchange rate data.

In Chapter 5, we consider data from five insurance portfolios. Table 4.9.3 below shows four different estimates of ρ for each of the portfolios. In Table 4.9.3 $\hat{\rho}_1$ denotes the estimator by Gomes and Martins (2001), which was discussed in detail in Section 4.1. Estimators $\hat{\rho}_3$ and $\hat{\rho}_4$ were defined in Section 4.4. We also include the estimates of ρ which were obtained by fitting the PPD (yielding $\hat{\gamma}$, $\hat{\rho}$ and $\hat{c}$). We denote this estimator by $\hat{\rho}_{\mathrm{PPD}}$. The details of how $\hat{\rho}_{\mathrm{PPD}}$ was obtained for the respective portfolios, are discussed in Section 5.2.

	Estimator			
Portfolio	$\hat{\boldsymbol{\rho}}_{\mathbf{1}}$	$\hat{\boldsymbol{\rho}}_{\mathbf{3}}$	$\hat{\boldsymbol{\rho}}_{\mathbf{4}}$	$\hat{\boldsymbol{\rho}}_{\mathbf{PPD}}$
1	-0.1686	-0.6898	-1.4340	-0.0277
2	-0.0100	-0.7776	-1.7838	-1.8539
3	-0.0100	-0.8231	-1.6600	-0.1262
4	-1.0360	-0.8842	-1.8605	-0.1764
5	-0.7716	-0.7963	-1.6854	-0.0109

Table 4.9.3 Estimates of the second order parameter for the portfolio data.

For all the cases mentioned above where estimates of ρ were obtained for real data, the estimates were greater than -2. This does not mean that it is impossible that $\rho < -2$ for real data, but it seems reasonable to restrict the range of ρ to $-2 \le \rho \le 0$ for distributions from which we simulate data.

Note that distributions from which data are simulated, include distributions for which $\rho = 0$. The reason is that for some heavy-tailed distributions $\rho = 0$, for instance the loggamma distribution. Omitting these distributions from the simulation study would bias the results in favour of estimators which perform well for data from distributions with $\rho < 0$, and badly when data are from distributions with $\rho = 0$. An estimator of the EVI (for a positive EVI) should produce good results over as wide a spectrum of heavy-tailed distributions as possible.

As far as parameter estimation is concerned, we have to assume $\rho < 0$. Section 3.1.4 presented the derivation of the PPD, which followed from the assumption that $\rho < 0$. From a mathematical point of view ρ must differ from zero, since the constant c of the PPD is defined as $c = \gamma dA(t)/\rho$. From a numerical point of view, we had to apply the restriction $\rho \le 0.01$ to avoid overflow when applying the restriction $\rho^{-1} \le c \le 1$ to c.

Therefore, distributions with $\rho = 0$ need to be included in the simulation study, and when data are from these distributions, this will be indicated by the estimate of ρ being close to zero (as also mentioned in Section 3.1.4).

Conclusion

In this chapter we compared the performance of the PPD estimator to that of various other estimators, and we proposed a slight modification of the PPD estimator for samples of sizes $n = 2000$ and $n = 5000$.

In Section 4.2, where other methods of estimating PPD parameters were investigated, it was found that external estimation of the second order parameter ρ might lead to more accurate estimation of the EVI for very large samples. For samples of sizes $n = 10\,000$ and $n = 20\,000$ it appeared as though it was marginally better to first estimate ρ using Gomes and Martins (2001), and then determining the Bayesian estimates of the two remaining parameters. This idea might be explored further in future research.

Apart from the estimators in Section 4.2, the PPD estimator is the method of choice across all sample sizes when compared to all the other estimators considered in Sections 4.3, 4.4, 4.5 and 4.6. However, the PPD estimator does not yield the lowest MSE across all EVI groups for all sample sizes. The GJ(ρ_3) estimator outperforms the PPD estimator for some EVI groups for samples of sizes $n = 2000$ and $n = 5000$, which is why we suggested a slight modification of the PPD estimator for those sample sizes in Section 4.8.

Since the PPD estimator is computationally intensive, and since it is difficult to write a computer program to implement the PPD estimator, alternatives to the PPD estimator were proposed, which are easy to program, and can be calculated rapidly.

The estimators considered in Section 4.4 provided the basis for these alternative methods of estimation. For small sample sizes (up to $n = 200$), we made use of the GJ, Peng and ML(ρ_3) estimators as alternatives. For sample sizes $n = 500$ and greater, the GJ(ρ_3) estimator is the only estimator we propose as an alternative to the PPD estimator.

Section 4.8 included simulation study results, showing the performance of the (modified) PPD estimator, and the alternative method of estimation.

In Section 4.9 we presented results concerning the bias of some of the estimators considered in Section 4.8. We concluded this chapter by giving a justification for choosing $-2 \leq \rho \leq 0$ as the range of the second order parameter for the distributions from which we simulate data.

Chapter 5

Case study: Insurance claims data

In this chapter case study data are analyzed as an illustration of how the techniques discussed so far can be used in practice. The data are sizes of claims from five insurance portfolios. Section 5.1 gives a general description of the data. Section 5.2 details the estimation of the EVI for each of the portfolios. In Section 5.3 we focus on one of these portfolios and illustrate how inferences can be made. This includes hpd regions for the EVI, exceedance probability estimation, and quantile estimation.

5.1 Description of the data

We analyze claims data from a South African short term insurer. We have at our disposal data from five portfolios. The risks of the portfolios (whether fire, automobile, etc.) are not known, and may differ from one portfolio to the next. The name of the insurer and risks of the portfolios are not disclosed for confidentiality reasons.

For Portfolios 1, 2 and 3, claims date from 1 July 2004 until 21 July 2006. For portfolios 4 and 5, claims date from 31 December 2001 until 24 February 2006. The dates which were used, were the dates the claims occurred, and not the dates the claims were registered.

The claim amounts were the total claim amounts and ignored any excesses paid by the client. The claim amounts were adjusted for inflation to July 2006 as base month. Finally, any negative or zero claim amounts were deleted from the data sets. Negative amounts occur if, for instance, the value of the items salvaged from the wreck exceeds the claim amount in value.

The sample sizes of the portfolios, and the five largest claims in each portoflio, are given in the following table:

		Largest claims (R million)				
Portfolio	**No. of claims**	**1**	**2**	**3**	**4**	**5**
1	16 197	0.6028	0.3218	0.2992	0.2970	0.2872
2	16 104	0.8356	0.8197	0.7323	0.4114	0.3782
3	15 990	1.4353	0.9090	0.4118	0.3941	0.3852
4	18 198	12.6914	6.9382	5.4887	4.3463	4.0587
5	14 917	9.7704	6.1922	5.1499	4.1441	3.4430

Table 5.1 Number of observations and five largest claims per portfolio.

The heavy-tail effect can be seen clearly from the fact the largest claims differ greatly in magnitude.

5.2 Estimation of the EVI

We will now discuss general aspects concerning the analysis of the data, and then examine each of the five data sets separately.

Histograms of the data do not provide useful information, since the underlying distribution of the data is heavy-tailed, causing virtually all observations to fall in the first cell.

For all portfolios, we start our analysis by showing the Hill plot and the Pareto plot. The Hill plot is constructed only up to $k = 40\%$ of n. These two graphs will not be discussed in detail, since they are mainly included for the sake of completeness.

The Hill and Pareto plots will be followed by a graph showing the PPD and GJ(ρ_3) estimates as a function of k.

The sections of the graph pertaining to the GJ(ρ_3) estimator are indicated by dashed lines. The vertical dashed line indicates the threshold at $k = 20\%$ of n. In Section 4.8.2 we saw that, over all EVI groups, a threshold of $k = 45\%$ of n was recommended for samples of size $n = 2000$, a threshold of $k = 35\%$ for samples of size $n = 5000$, a threshold of $k = 25\%$ for samples of size $n = 10\,000$, and a threshold of $k = 15\%$ for samples of size $n = 20\,000$. Since all portfolios have between $n = 10\,000$ and $n = 20\,000$ observations, a recommended threshold of $k = 20\%$ of n throughout seemed a logical choice.

The dashed horizontal line indicates the GJ(ρ_3) estimate at $k = 20\%$ of n, and therefore our GJ(ρ_3) estimate of the EVI for that portfolio.

The part of the graph pertaining to the PPD estimator is indicated by solid lines. The vertical solid line indicates the optimal threshold, and the solid horizontal line indicates the PPD estimate at the optimal threshold, and therefore our PPD estimate of the EVI for that portfolio.

In Section 3.4.4 we described the PPD benchmark estimator. For samples of size $n = 10\,000$ this entailed applying Method 2 to the set of median estimates. For samples of size $n = 20\,000$ this entailed calculating the PPD estimates at $k = 1\%, 2\%, \dots, 30\%$ of n, and applying Method 1. Refer to the end of Section 3.4.4, where estimation procedures for very large samples are discussed.

All portfolios have between $n = 10\,000$ and $n = 20\,000$ observations. We estimate the EVI by applying both methods (for $n = 10\,000$ and $n = 20\,000$), and taking the average of these estimates as our PPD estimate. This estimate we call the original PPD estimate. (Note that in none of the cases we obtained an EVI estimate below 0.3, which would have necessitated a second step in the estimation process for samples of size $n = 10\,000$. Refer to Section 3.4.4.)

We then apply the implied threshold principle as described in Section 3.4.5. Simply stated, we find the EVI estimate (at thresholds $k = 5\%, 10\%, \ldots, 95\%$ of n) which is closest to our original PPD estimate. The EVI estimate at that threshold, which differs slightly from our original estimate, is then taken as our (final) PPD estimate. In addition, if EVI estimates at different thresholds are for all practical purposes equally close to the original PPD estimate, the larger threshold will be selected for our final PPD estimate. The reason for this is to include more observations, leading to narrower intervals when doing inference. It will be indicated where this approach has been applied.

All PPD estimates were obtained by generating $10\,000$ Gibbs draws and 100 burn-in draws.

All plots of the PPD and GJ(ρ_3) estimates have been scaled to a range of 0.4 on the y-axis (for instance 0.3 to 0.7), and 0 to 100 on the x-axis (number of excesses k as percentage of n) in order to ease comparison, especially as far as the stability of the estimates is concerned.

Finally, recall that for the set of PPD parameters to define a distribution, the restriction $\rho^{-1} \leq c \leq 1$ has to hold. All sets of PPD parameter estimates (containing values of $\hat{\gamma}$, $\hat{\rho}$ and $\hat{c}$) calculated for the portfolios below, met the requirement that $\hat{c} \geq \hat{\rho}^{-1}$.

Portfolio 1

The Hill plot and the Pareto plot of the Portfolio 1 data are shown below.

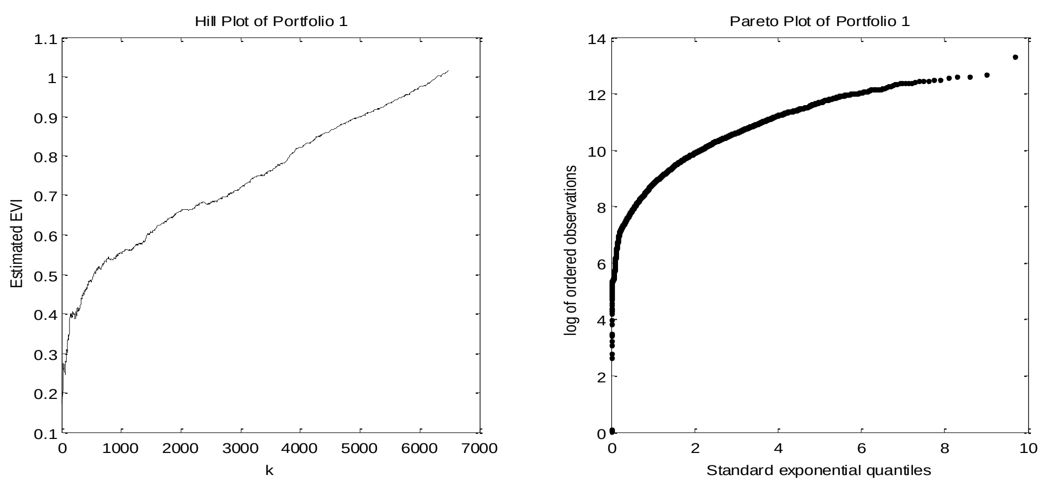

Figure 5.2.1 Hill plot and Pareto plot for Portfolio 1.

The fact that the Hill estimates consistently increase as a function of k, and the fact that it is difficult to see any point from which the Pareto plot becomes linear (because of the curvature of the graph), indicate that the threshold selection will be crucial.

This observation is confirmed in the following graph:

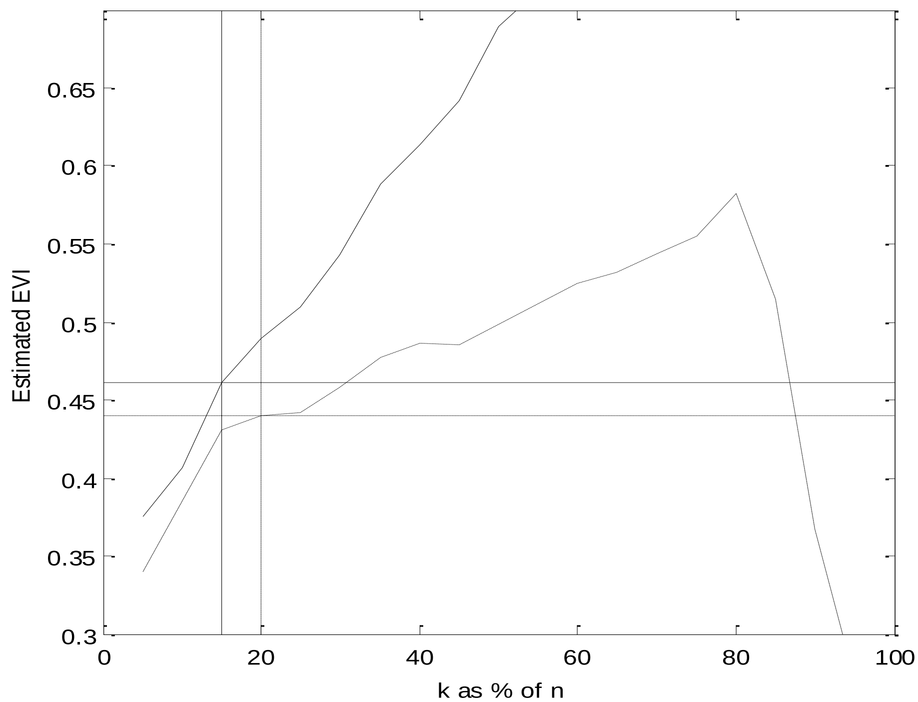

Figure 5.2.2 PPD (solid) and GJ(ρ_3) (dashed) estimates for Portfolio 1.

As indicated on the graph, the GJ(ρ_3) estimate of the EVI is 0.4402.

The PPD estimate for samples of size $n = 10\,000$ is 0.4867. The estimate for samples of size $n = 20\,000$ is 0.4528, which yields an average of 0.4697.

We choose our optimal threshold at the EVI estimate closest to 0.4697, which is 0.4613 at $k = 15\%$ of n.

The PPD and GJ(ρ_3) estimates do not differ by much. We regard the PPD estimate as our final estimate of the EVI.

At our choice of threshold ($k = 15\%$ of n), the estimates of the PPD parameters are $\hat{\gamma} = 0.4613$, $\hat{\rho} = -0.0277$ and $\hat{c} = -17.8846$.

Portfolio 2

The Hill plot and the Pareto plot of the Portfolio 2 data are shown below.

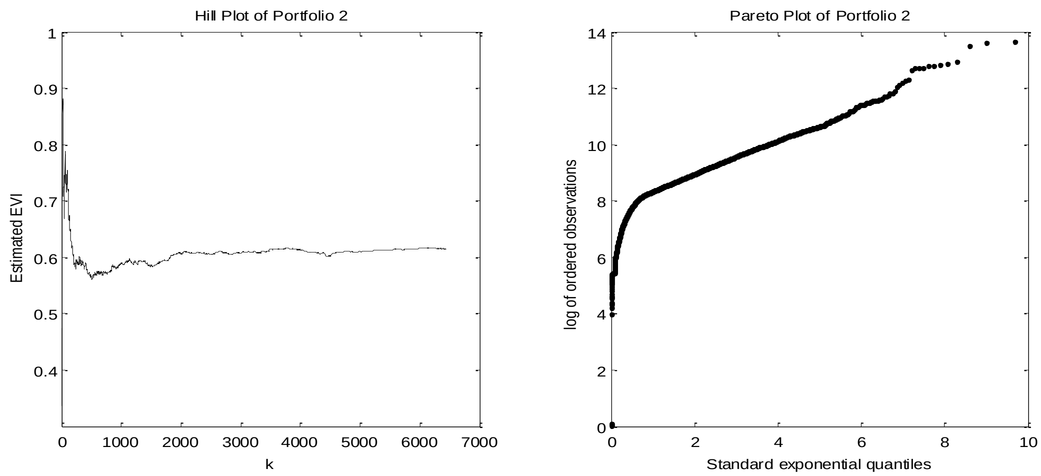

Figure 5.2.3 Hill plot and Pareto plot for Portfolio 2.

The Hill plot is remarkably stable at around 0.6, and the Pareto plot seems almost perfectly linear over a large region. This behaviour is not at all common for real data, even to the extent that one may draw the validity of the data into question. Assuming the data are valid, one might conclude that very reliable estimates of the EVI will be possible.

We now consider the other estimators:

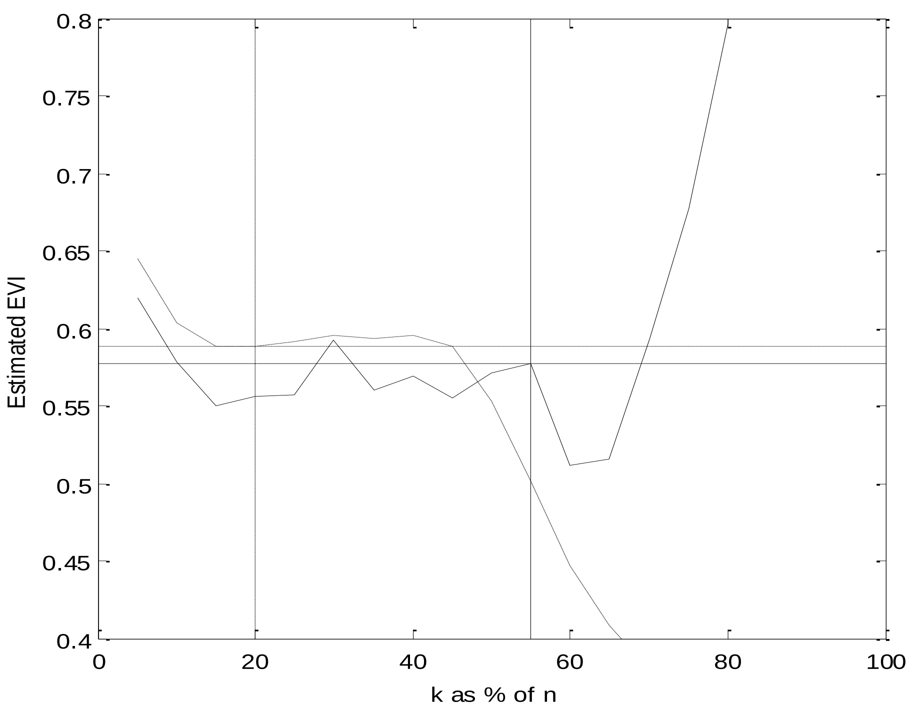

Figure 5.2.4 PPD (solid) and GJ(ρ_3) (dashed) estimates for Portfolio 2.

As indicated on the graph, the GJ(ρ_3) estimate of the EVI is 0.5885.

The PPD estimate for samples of size $n = 10\,000$ is 0.5664. The estimate for samples of size $n = 20\,000$ is 0.5811, which yields an average of 0.5737.

The EVI estimate closest to 0.5737 is 0.5708 at $k = 50\%$ of n. We choose, however, the EVI esimate at $k = 55\%$ of n, which yields 0.5772 as our final PPD estimate, since it is at a larger threshold, and only 0.0006 further away from 0.5737 than 0.5708.

The PPD and GJ(ρ_3) estimates do not differ by much, which confirms that our estimates may be reliable. We regard the PPD estimate as our final estimate of the EVI.

At our choice of threshold ($k = 55\%$ of n), the estimates of the PPD parameters are $\hat{\gamma} = 0.5772$, $\hat{\rho} = -1.8539$ and $\hat{c} = -0.397$.

Portfolio 3

The Hill plot and the Pareto plot of the Portfolio 3 data are shown below.

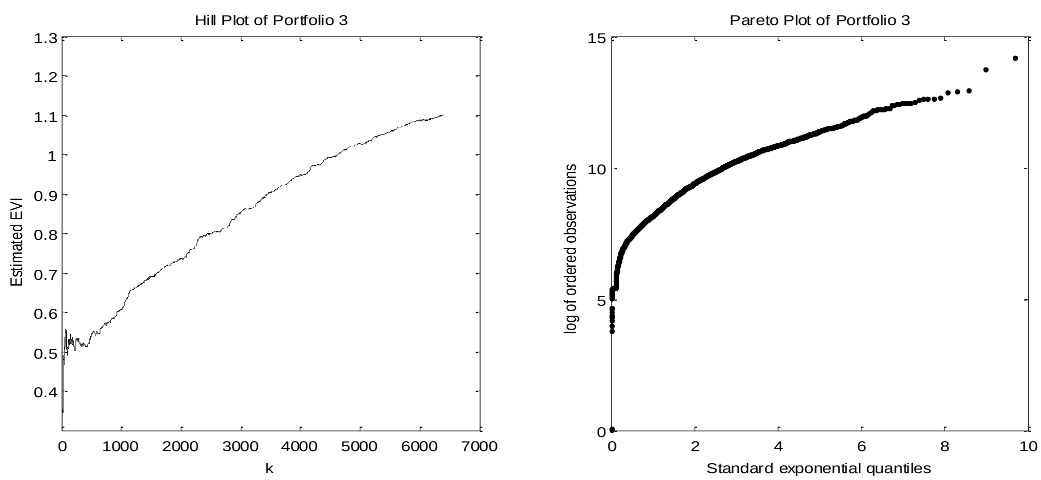

Figure 5.2.5 Hill plot and Pareto plot for Portfolio 3.

Before the Hill estimates start to increase sharply (due to the bias), the estimates seem to be relatively constant around 0.52. As far as the Hill estimates are concerned, this seems to be a quite reliable estimate of the EVI.

We now consider the other estimators:

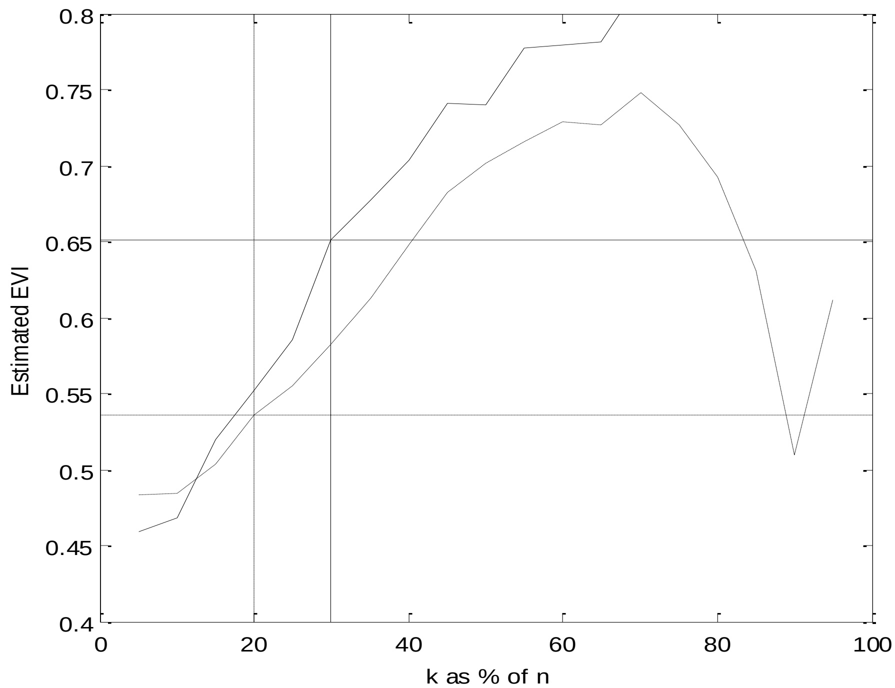

Figure 5.2.6 PPD (solid) and GJ(ρ_3) (dashed) estimates for Portfolio 3.

As indicated on the graph, the GJ(ρ_3) estimate of the EVI is 0.5357.

The PPD estimate for samples of size $n = 10\ 000$ is 0.7792. The estimate for samples of size $n = 20\,000$ is 0.5247, which yields an average of 0.6520.

We cannot continue blindly with the same procedure we followed up to now. For the previous two portfolios the estimate for samples of size $n = 10\ 000$, the estimate for samples of size $n = 20\,000$, and the GJ(ρ_3) estimate were all relatively close to one another. We do not have the same situation here, even after the average of 0.6520 was obtained, as can be seen in Figure 5.2.6.

We will disregard the estimate for samples of size $n = 10\ 000$ completely, on the following grounds:

- As we have seen, after taking the average of the estimate for samples of size $n = 10\,000$ and the estimate for samples of size $n = 20\ 000$, the PPD and GJ(ρ_3) estimates still differ too much.
- The sample size of Portfolio 3 is $n = 15\ 990$, which is closer to $n = 20\ 000$ than $n = 10\,000$, if a choice has to be made purely on the basis of sample size.
- The method for samples of size $n = 20\,000$, is superior to that of $n = 10\,000$, because of the additional results obtained for very large samples. Refer to the end of Section 3.4.4.

- The estimate for samples of size $n = 20\ 000$ is 0.5247, which corresponds closely to the GJ(ρ_3) estimate of 0.5357, and even the approximate value of the EVI of 0.52, as suggested by the Hill plot.

The graph is adjusted to incorporate only the method for samples of size $n = 20\,000$ for the PPD estimate.

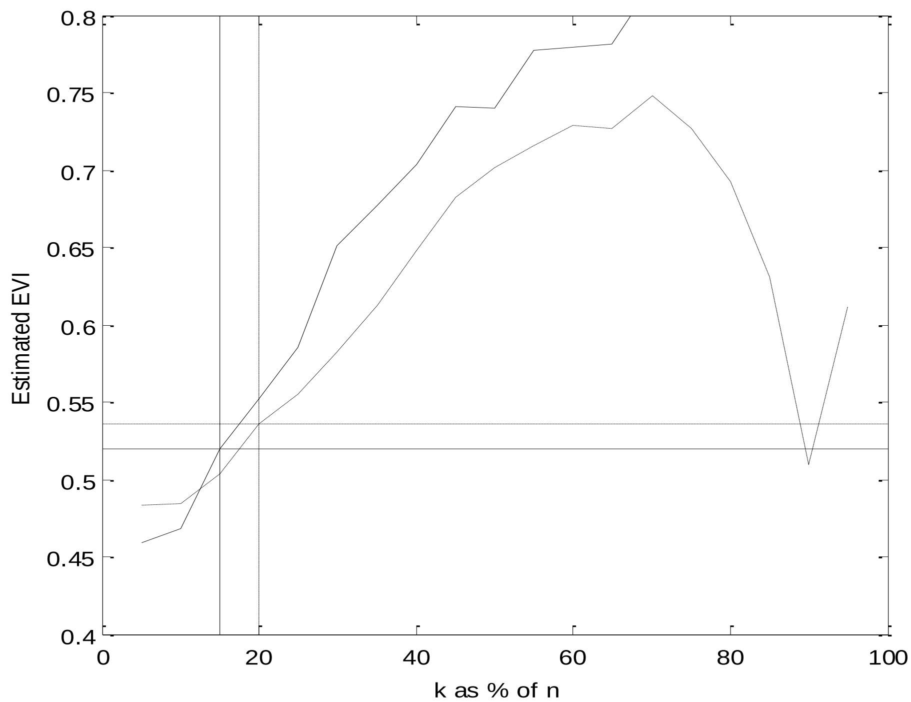

Figure 5.2.7 Ajusted PPD (solid) and original GJ(ρ_3) (dashed) estimates for Portfolio 3.

The EVI estimate closest to 0.5247 is 0.5202 at $k = 15\%$ of n, which we regard as our final estimate of the EVI.

At our choice of threshold ($k = 15\%$ of n), the estimates of the PPD parameters are $\hat{\gamma} = 0.5202$, $\hat{\rho} = -0.1262$ and $\hat{c} = -4.6826$.

Portfolio 4

The Hill plot and the Pareto plot of the Portfolio 4 data are shown below.

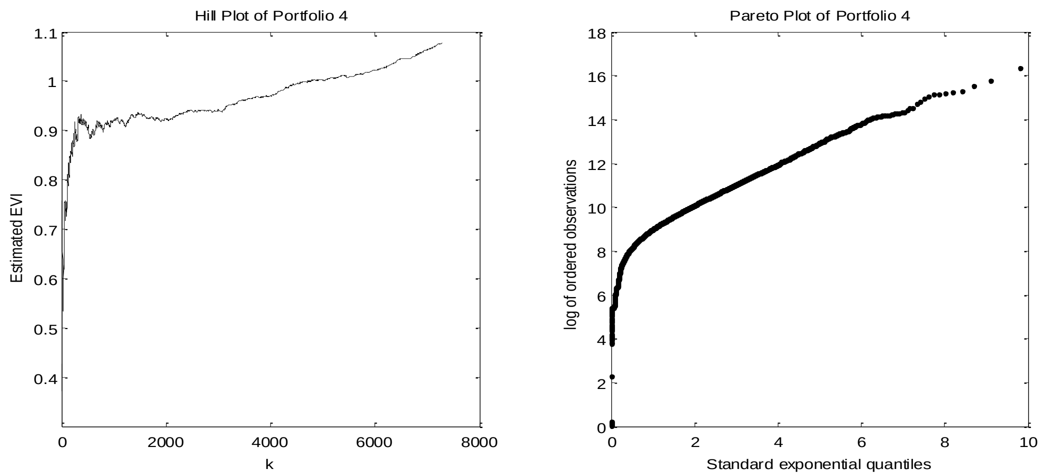

Figure 5.2.8 Hill plot and Pareto plot for Portfolio 4.

Considering the Hill plot, our guess would be that the EVI is in the region of 0.91. The fact that the Pareto plot is linear over a wide range, suggests that a reliable estimate of the EVI may be obtained for these data.

We now consider the other estimators:

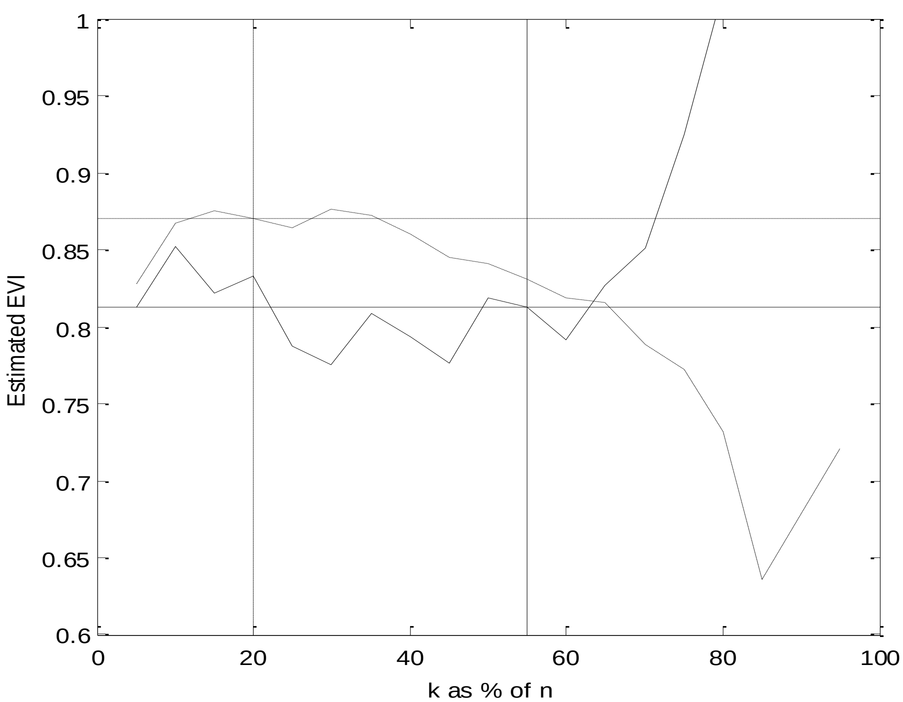

Figure 5.2.9 PPD (solid) and GJ(ρ_3) (dashed) estimates for Portfolio 4.

As indicated on the graph, the GJ(ρ_3) estimate of the EVI is 0.8707.

The PPD estimate for samples of size $n = 10\,000$ is 0.8084. The estimate for samples of size $n = 20\,000$ is 0.8159, which yields an average of 0.8121. At $k = 55\%$ of n, the estimated EVI is 0.8122, as indicated by the solid horizontal line.

We cannot continue blindly with the same procedure we followed for Portfolios 1 and 2, where the estimate for samples of size $n = 10\,000$, the estimate for samples of size $n = 20\,000$, and the GJ(ρ_3) estimate were all relatively close to one another.

The solution we had for Portfolio 3 is also not applicable here. For Portfolio 3 the difference between the estimate for samples of size $n = 10\,000$ and the estimate for samples of size $n = 20\,000$ was significant, and one estimate was clearly preferable to the other. For Portfolio 4, the estimate for samples of size $n = 10\,000$ and the estimate for samples of size $n = 20\,000$ differ by only 0.0075. The PPD estimate does not vary much depending on the technique used.

For Portfolio 4, we rather have to choose between the PPD estimate and the GJ(ρ_3) estimate. If we decide on the PPD estimate, we will choose our threshold at $k = 55\%$ of n, which yields PPD parameter estimates $\hat{\gamma} = 0.8122$, $\hat{\rho} = -0.1764$ and $\hat{c} = -3.4672$. If we decide on the GJ(ρ_3) estimate, we choose a threshold yielding a PPD estimate as close as possible to the GJ(ρ_3) estimate of 0.8707. (Note that we only do inferences with respect to the PPD in this thesis.) The threshold which yields a PPD estimate closest to 0.8707, is at $k = 70\%$ of n, which yields PPD parameter estimates $\hat{\gamma} = 0.8505$, $\hat{\rho} = -0.0682$ and $\hat{c} = -12.015$, indicated in Figure 5.2.10 below.

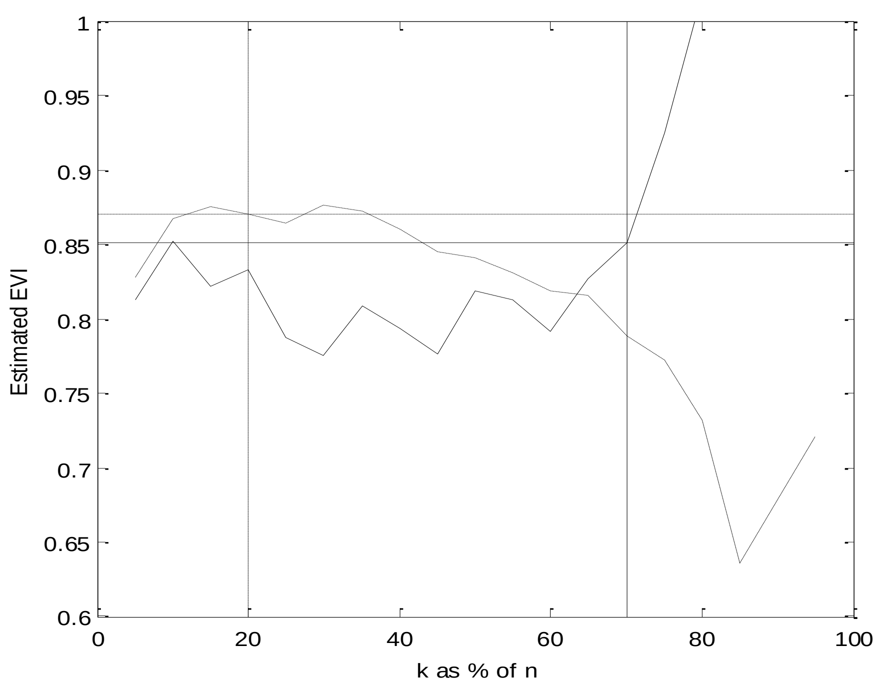

Figure 5.2.10 The PPD estimate (solid) as close as possible to the GJ(ρ_3) estimate (dashed).

We will now describe a technique based on resampling, which can be used to decide between two sets of PPD parameters.

In the case of Portfolio 4, these sets are

Set A: $\hat{\gamma} = 0.8122$, $\hat{\rho} = -0.1764$ and $\hat{c} = -3.4672$, and

Set B: $\hat{\gamma} = 0.8505$, $\hat{\rho} = -0.0682$ and $\hat{c} = -12.015$.

The purpose of extreme value analysis is the accurate estimation of small exceedance probabilities and/or large quantiles. We use as measure to differentiate between the parameter sets the accuracy of estimation of exceedance probabilities rather than the accuracy of estimation of large quantiles, since it is simpler to justify an error measure for probability estimation, and the scale (all values between 0 and 1) makes the interpretation of individual estimates easier.

The basic idea of the method is the following: We choose a large quantile from the Portfolio 4 data, which has a known exceedance probability p_1. We regard this as the "true exceedance probability". We then repeatedly draw subsamples (of size $n = 1000$, say) from Portfolio 4, and estimate p_1 by using both parameter sets A and B. The set which estimates p_1 "best" on average, is then regarded as the set which is preferable.

We will now give the details of the procedure. We will use the same exceedance probability estimation technique we discussed in Section 3.2.6. Since we have two sets of data (the entire Portfolio 4 data set and the subsample), we need to define some notation to avoid confusion.

X The random variable X denotes the size of a claim relating to Portfolio 4.

n The size of the subsamples is denoted by n. We have chosen $n = 1000$. Other samples sizes are also valid.

N The total number of observations in a portfolio is denoted by N. For Portfolio 4 we have $N = 18\,198$ observations.

k The number of excesses pertaining to the subsamples is denoted by k. The selection of k will be discussed after we have described the procedure.

K The number of excesses pertaining to the total portfolio is denoted by K. The selection of K will be discussed after we have described the procedure.

The steps of the procedure are the following:

1. Choose K. Note that this is equivalent to choosing a large quantile, as pointed out in Step 2. We choose K and not the quantile, simply because it is more convenient.
2. Calculate x, the large quantile implied by K, which we take as the average of the K-th largest and $(K + 1)$-th largest observation of the portfolio.

3. Calculate the quantity we wish to estimate, namely $p_1 = P(X > x) \approx K/N$.
4. Draw a random subsample of size n with replacement from the total number of observations N.
5. Choose k.
6. Obtain the threshold implied by k, namely $t = x_{n-k,n}$, where $x_{1,n} \leq x_{2,n} \leq \cdots \leq x_{n,n}$ denote the ordered observations from the subsample.
7. Calculate $z = x/t$.
8. Calculate $\hat{p}_{2A} = (1 - \hat{c})z^{-1/\hat{\gamma}} + \hat{c}z^{-(1-\hat{\rho})/\hat{\gamma}}$, using parameter set A.
9. Calculate $\hat{p}_{1A} = k\hat{p}_{2A}/n$.
10. Calculate $\hat{p}_{2B} = (1 - \hat{c})z^{-1/\hat{\gamma}} + \hat{c}z^{-(1-\hat{\rho})/\hat{\gamma}}$, using parameter set B.
11. Calculate $\hat{p}_{1B} = k\hat{p}_{2B}/n$.

We repeat Steps 4 to 11 a large number of times m, obtaining m $\hat{p}_{1A}$ estimates and m $\hat{p}_{1B}$ estimates. (For Portfolio 4 we chose $m = 500$.) An error measure is then calculated, and the set of estimates yielding the lowest error is regarded as the preferable set of the two.

As a measure of error, we propose calculating the absolute values of the relative errors

$$r = \left|\frac{\hat{p}_1 - p_1}{p_1}\right|.$$

A relative error makes sense. One can see this by considering an example of an estimation error. Suppose $p_1 = 0.04$ and $\hat{p}_1 = 0.06$. The interpretation of the error is that we estimate the number of exceedances above the (given) large quantile as one and a half times (50%) greater than it actually is. The MSE does not have an intuitive interpretation in the same sense (even though there is, from a practical point of view, not much difference in the final conclusions). We then take the absolute value, since we do not wish positive and negative error values to cancel each other out.

The median of the errors r_1, r_2, …, r_m is taken as the overall error measure for a specific parameter set.

We will now discuss the choice of K and k.

K must be chosen small compared to N, so that the exceedance probability $p_1 = K/N$ is small. Recall that $N = 18\,198$ for Portfolio 4. We conducted the simulation study at five values of K, namely $K_1 = 50$, $K_2 = 100$, $K_3 = 150$, $K_4 = 200$, $K_5 = 250$.

The obvious choice of k is the implied threshold (number of excesses) as obtained as a result of the PPD estimation procedure. Refer to Section 3.4.5. Aside from this choice of k, we also considered five fixed values of k, namely $k_1 = 5\%$ of n, $k_2 = 10\%$ of n, $k_3 = 15\%$ of n, $k_4 = 20\%$ of n, $k_5 = 25\%$ of n. Since $n = 1000$, this translates to $k_1 = 50$, $k_2 = 100$, $k_3 = 150$, $k_4 = 200$, $k_5 = 250$.

The results obtained for parameter set A ($\hat{\gamma} = 0.8122$, $\hat{\rho} = -0.1764$ and $\hat{c} = -3.4672$) for Portfolio 4, are given in Table 5.2.1 below. The PPD column indicates that the value of k used, was the value of k obtained from the threshold selection procedure of the benchmark PPD estimator.

		k					
		PPD	**50**	**100**	**150**	**200**	**250**
	50	0.3209	0.6728	0.6546	0.6153	0.5368	0.4528
	100	0.2816	0.5295	0.5490	0.5284	0.4645	0.3915
K	**150**	0.3399	0.5525	0.6087	0.6033	0.5463	0.4761
	200	0.3426	0.5110	0.5948	0.6020	0.5528	0.4875
	250	0.3922	0.5169	0.6336	0.6547	0.6123	0.5503

Table 5.2.1 Exceedance probability estimation errors for parameter set A for Portfolio 4.

Parameter set B ($\hat{\gamma} = 0.8505$, $\hat{\rho} = -0.0682$ and $\hat{c} = -12.015$) yielded the results in Table 5.2.2 below.

		k					
		PPD	**50**	**100**	**150**	**200**	**250**
	50	1.4206	1.5532	1.7300	1.7875	1.7446	1.6674
	100	1.1793	1.1710	1.3853	1.4658	1.4478	1.3932
K	**150**	1.2061	1.0845	1.3515	1.4592	1.4595	1.4177
	200	1.1506	0.9539	1.2517	1.3762	1.3903	1.3594
	250	1.1760	0.8928	1.2330	1.3793	1.4080	1.3869

Table 5.2.2 Exceedance probability estimation errors for parameter set B for Portfolio 4.

From the two tables above, it is clear that parameter set A leads to more accurate exceedance probability estimation.

Additional information can be obtained by constructing histograms of the estimates themselves. Figure 5.2.11 below shows a histogram of 500 estimates of p_1 using parameter set A, and a histogram of 500 estimates of p_1 using parameter set B.

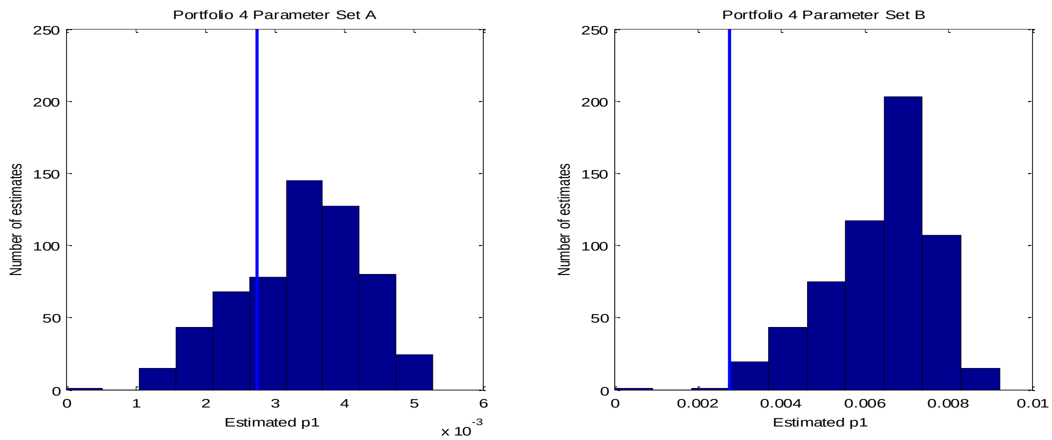

Figure 5.2.11 Estimated exceedance probabilities for both parameter sets for Portfolio 4.

The histograms in Figure 5.2.11 pertain to the case where $K = 50$, and k is the PPD estimator threshold. The true value of p_1 is indicated by the solid vertical line: $p_1 = 50/18\,198 = 0.0027$.

Here we can see that both parameter sets yield estimates of p_1 which are too large on average. The estimate of the EVI as yielded by the PPD method ($\hat{\gamma} = 0.8122$) is preferable to the GJ(ρ_3) estimate of $\hat{\gamma} = 0.8505$, because it is smaller than the GJ(ρ_3) estimate (and even $\hat{\gamma} = 0.8122$ might be too large).

We therefore choose as our final estimate of the EVI the PPD estimate (0.8122), obtained at $k = 55\%$ of n, which yielded estimated PPD parameters $\hat{\gamma} = 0.8122$, $\hat{\rho} = -0.1764$ and $\hat{c} = -3.4672$.

Portfolio 5

The Hill plot and the Pareto plot of the Portfolio 5 data are shown below.

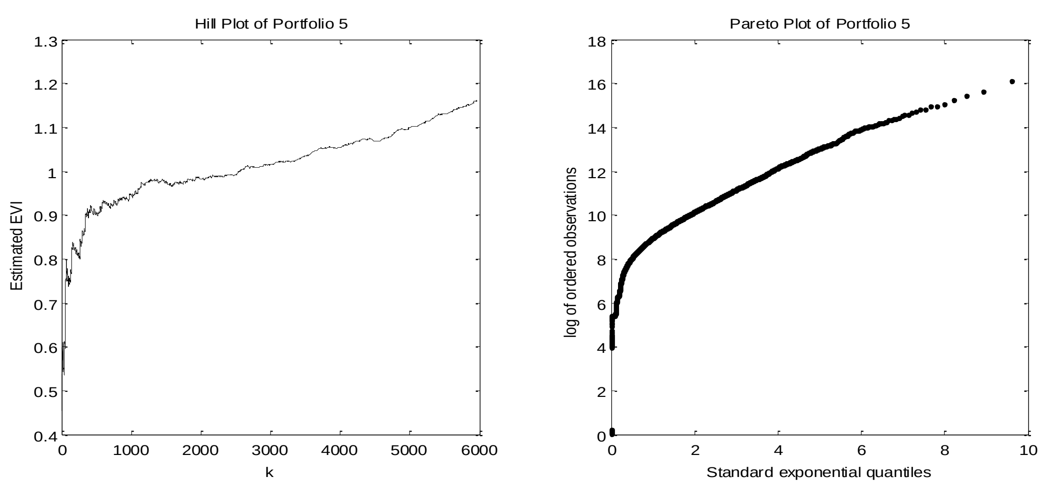

Figure 5.2.12 Hill plot and Pareto plot for Portfolio 5.

The Hill plot is not very helpful. There is no apparent region of stability, which causes the estimate of the EVI to depend too heavily on the choice of threshold. The Pareto plot looks a bit more promising, with an almost perfectly straight line from approximately 5.5 (on the *x*-axis) onwards.

We now consider the other estimators:

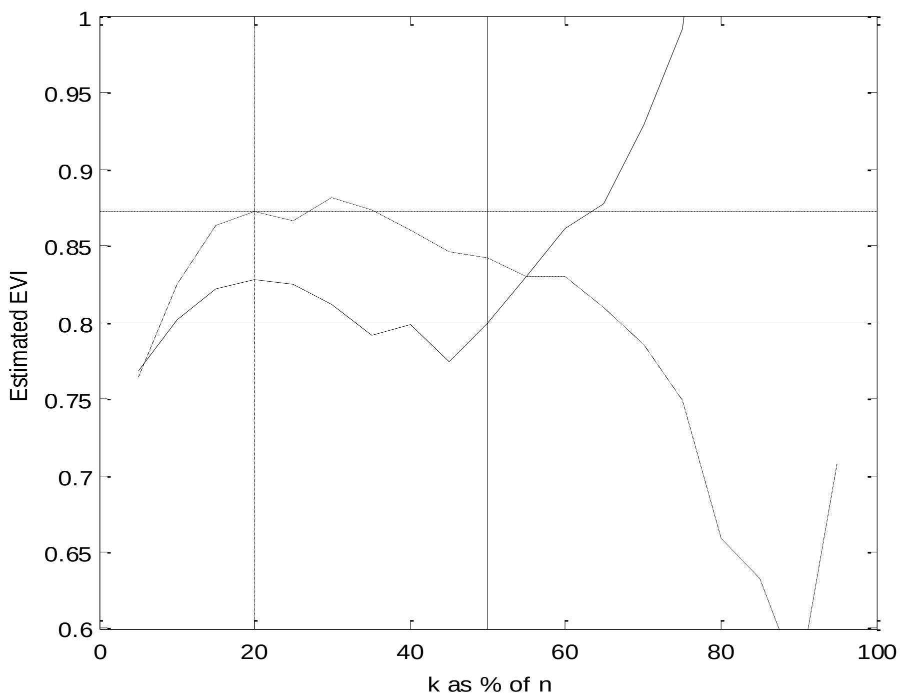

Figure 5.2.13 PPD (solid) and GJ(ρ_3) (dashed) estimates for Portfolio 5.

As indicated on the graph, the GJ(ρ_3) estimate of the EVI is 0.8726.

The PPD estimate for samples of size $n = 10\,000$ is 0.8057. The estimate for samples of size $n = 20\,000$ is 0.799, which yields an average of 0.8024. At $k = 50\%$ of n, the estimated EVI is 0.7994, as indicated by the solid horizontal line.

We have here the same situation as we had for Portfolio 4, namely that we have to choose between the PPD estimate and the GJ(ρ_3) estimate, and for the same reasons as mentioned there.

If we decide on the PPD estimate, we will choose our threshold at $k = 50\%$ of n, which yields PPD parameter estimates $\hat{\gamma} = 0.7994$, $\hat{\rho} = -0.0109$ and $\hat{c} = -55.5485$. If we decide on the GJ(ρ_3) estimate, we choose a threshold yielding a PPD estimate as close as possible to the GJ(ρ_3) estimate of 0.8726. The threshold which yields a PPD estimate closest to 0.8726, is at $k = 65\%$ of n, which yields PPD parameter estimates $\hat{\gamma} = 0.8774$, $\hat{\rho} = -0.05$ and $\hat{c} = -15.0158$, indicated in Figure 5.2.14 below.

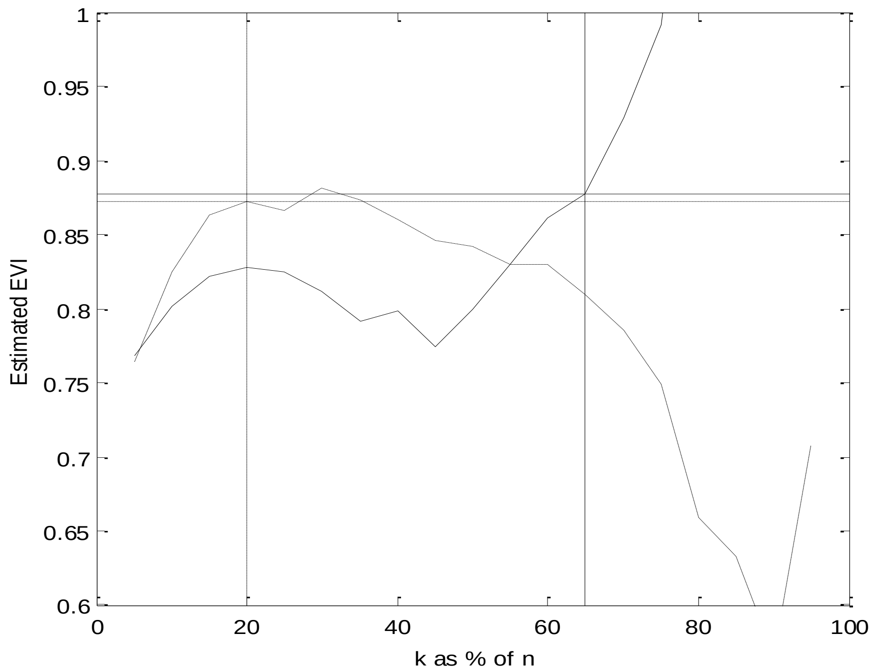

Figure 5.2.14 The PPD estimate (solid) as close as possible to the GJ(ρ_3) estimate (dashed).

We use the same technique of resampling used for Portfolio 4 to choose between the following two sets of parameters:

Set A: $\hat{\gamma} = 0.7994$, $\hat{\rho} = -0.0109$ and $\hat{c} = -55.5485$, and

Set B: $\hat{\gamma} = 0.8774$, $\hat{\rho} = -0.05$ and $\hat{c} = -15.0158$.

The results obtained for parameter set A for Portfolio 5, are given in Table 5.2.3 below. The PPD column indicates that the value of k used, was the value of k obtained from the threshold selection procedure of the benchmark PPD estimator.

		k					
		PPD	**50**	**100**	**150**	**200**	**250**
	50	0.4064	0.8029	0.7671	0.7314	0.6591	0.5594
	100	0.3876	0.6666	0.6828	0.6706	0.6147	0.5276
K	**150**	0.3740	0.5724	0.6251	0.6294	0.5849	0.5066
	200	0.3344	0.4684	0.5462	0.5622	0.5270	0.4567
	250	0.2805	0.3611	0.4547	0.4787	0.4508	0.3879

Table 5.2.3 Exceedance probability estimation errors for parameter set A for Portfolio 5.

Parameter set B yielded the results in Table 5.2.4 below.

		k					
		PPD	**50**	**100**	**150**	**200**	**250**
	50	1.3600	1.5004	1.6304	1.6830	1.6506	1.5564
	100	1.1700	1.1624	1.3503	1.4317	1.4245	1.3547
K	**150**	1.0638	0.9572	1.1832	1.2833	1.2923	1.2377
	200	0.9502	0.7763	1.0233	1.1342	1.1539	1.1109
	250	0.8238	0.6128	0.8686	0.9841	1.0108	0.9766

Table 5.2.4 Exceedance probability estimation errors for parameter set B for Portfolio 5.

From the two tables above, it is clear that parameter set A leads to more accurate exceedance probability estimation.

Figure 5.2.15 below shows a histogram of 500 estimates of p_1 using parameter set A, and a histogram of 500 estimates of p_1 using parameter set B.

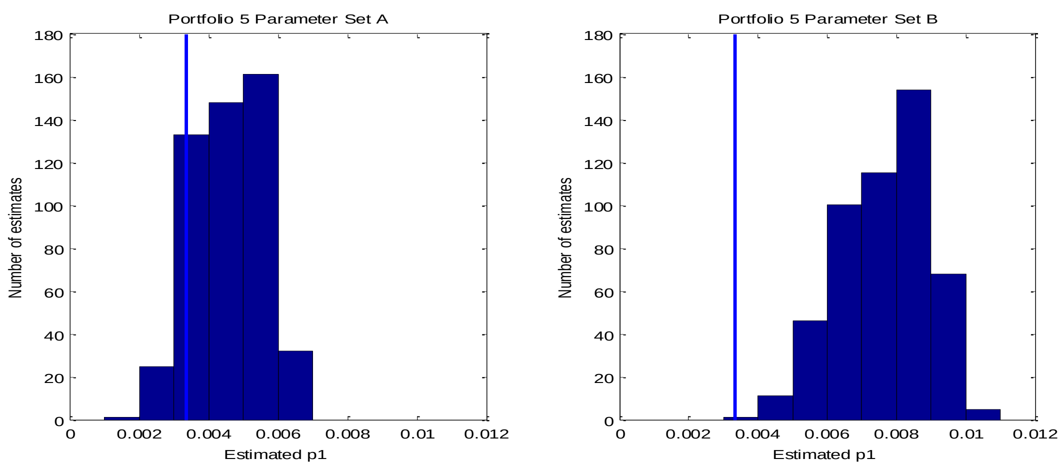

Figure 5.2.15 Estimated exceedance probabilities for both parameter sets for Portfolio 5.

The histograms in Figure 5.2.15 pertain to the case where $K = 50$, and k is the PPD estimator threshold. The true value of p_1 is indicated by the solid vertical line: $p_1 = 50/14917 = 0.0034$.

Here we can see that both parameter sets yield estimates of p_1 which are too large on average. The estimate of the EVI as yielded by the PPD method ($\hat{\gamma} = 0.7994$) is preferable to the GJ(ρ_3) estimate of $\hat{\gamma} = 0.8726$, because it is smaller than the GJ(ρ_3) estimate (and even $\hat{\gamma} = 0.7994$ might be too large).

We therefore choose as our final estimate of the EVI the PPD estimate (0.7994), obtained at $k = 50\%$ of n, which yielded estimated PPD parameters $\hat{\gamma} = 0.7994$, $\hat{\rho} = -0.0109$ and $\hat{c} = -55.5485$.

5.3 Inference for Portfolio 1

In this section we give an illustration of how inference is done in practice, once a final decision has been made on the choice of threshold, and the PPD parameter estimates have been obtained. We will only show inferences for Portfolio 1, since the same techniques are applicable to all portfolios.

In Section 5.2 we decided on a threshold of $k = 15\%$ of n, which yielded PPD parameter estimates $\hat{\gamma} = 0.4613$, $\hat{\rho} = -0.0277$ and $\hat{c} = -17.8846$.

The number of excesses is $k = 15\%$ of n = $0.15(16197) = 2429.55 \approx 2430$.

The threshold is t = $(k+1)$-th largest observation = 18404.93.

To get an idea of how well the model fits, a PPD Q-Q plot is constructed. The data are the 2430 largest claims, divided by the threshold 18404.93. The model quantiles are that of a PPD with parameters $\gamma = 0.4613$, $\rho = -0.0277$ and $c = -17.8846$.

Taking the logs of both the ordered observations and the PPD quantiles yields the following graph:

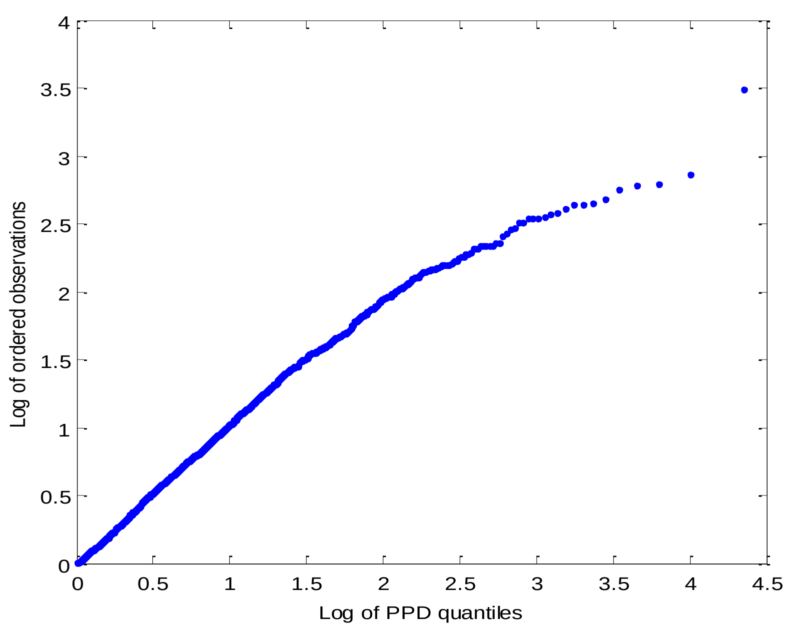

Figure 5.3.1 PPD Q-Q plot for multiplicative excesses of Portfolio 1.

From Figure 5.3.1 we can see that the fit is reasonable, except for the last few observations. Frequently this is the case, due to the sparsity of observations in the extreme tails.

In Section 3.2.4 we described the definition and calculation of the hpd region for parameter estimates obtained using Gibbs sampling. For the EVI of Portfolio 1, the respective hpd regions are given in the following table:

Coverage probability	Lower limit	Upper limit	Length
90%	0.4336	0.4927	0.0591
95%	0.4300	0.5078	0.0778
99%	0.4259	0.5328	0.1069

Table 5.3.1 Hpd regions for the EVI estimate of Portfolio 1.

The hpd regions of the EVI do not have a great deal of practical value. They do, however, give an indication of uncertainty with respect to the estimate, but one can do little with that information. At best one can use the upper limits instead of the estimate $\hat{\gamma} = 0.4613$ for the worst case scenario.

Next we discuss how we do inferences which give us a practical indication of claim sizes and their probabilities over a given time frame.

Table 5.1 showed the number of claims and the size of the five largest claims for the five portfolios. The following table is an excerpt from Table 5.1 pertaining to Portfolio 1:

		Largest claims (R million)				
Portfolio	**No. of claims**	**1**	**2**	**3**	**4**	**5**
1	16 197	0.6028	0.3218	0.2992	0.2970	0.2872

Table 5.3.1 Excerpt from Table 5.1 pertaining to Portfolio 1.

We start with exceedance probability estimation as described in Section 3.2.6.

Suppose we are interested in the exceedance probabiliy of the amounts 0.5, 1, 2, 5 and 10 (all in R million). Note that even though $500\,000$ is within the range of the data, we are still able to estimate its exceedance probability using the PPD model. Empirically (using only the data available without fitting a model) we see that only one value exceeds $500\,000$. Empirically, therefore, the estimated probability is $1/16197 = 6.174 \times 10^{-5}$.

As an example, we show the calculation probabiliy for $x = 500\,000$. Recall that $k = 2430$ and t = 18404.93.

$z = x/t$ = $500000/18404.93$ = 27.1679.

For $\hat{\gamma} = 0.4613$, $\hat{\rho} = -0.0277$ and $\hat{c} = -17.8846$ we have

$\hat{p}_2 = (1 - \hat{c})z^{-1/\hat{\gamma}} + \hat{c}z^{-(1-\hat{\rho})/\hat{\gamma}}$ = $0.003288 \approx P(Z > z)$.

$\hat{p}_1 = k\hat{p}_2/n$ = $0.15(0.003288) = 0.000493$.

We therefore estimate the probability that a single claim (or the next claim) exceeds R500 000, as 0.000493.

Table 5.3.2 below shows the exceedance probability above, as well as the results for the other amounts, up to R10 million.

	Claim amount (R million)				
	0.5	**1**	**2**	**5**	**10**
Next claim	4.9319E-04	1.2534E-04	3.1222E-05	4.8586E-06	1.1736E-06

Table 5.3.2 Probability that a single claim exceeds the given amounts.

Suppose we are also interested in the probability of exceeding those amounts over the next year, for example.

We follow a simple approach to answer this question. Many more complex and reliable ways of modelling the number of claims in a given year are bound to exist, but we are only interested in giving the general idea behind this approach.

For Portfolio 1 we have data from 1 July 2004 to 21 July 2006. We omit the data for July 2006, so that we only have data for 24 full months. We do this only to simplify the example.

During that two year period, there was at least one claim each day. The graph below shows a histogram of the number of claims per day over this period.

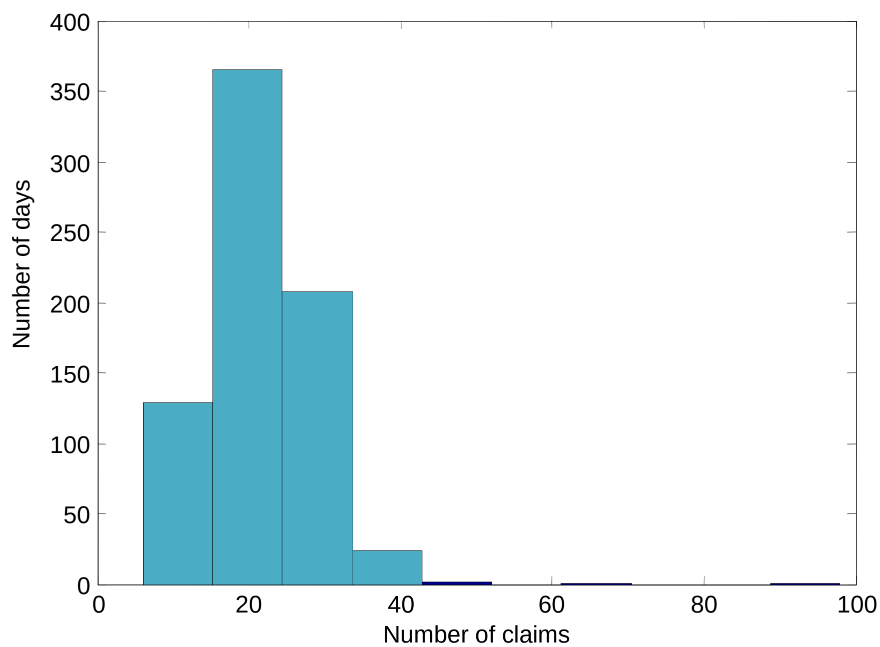

Figure 5.3.2 Histogram of number of claims per day for Portfolio 1.

Figure 5.3.3 Shows the number of claims per day as a time series. Here Day 1 is 1 July 2004 up to Day 730, which is 30 June 2006.

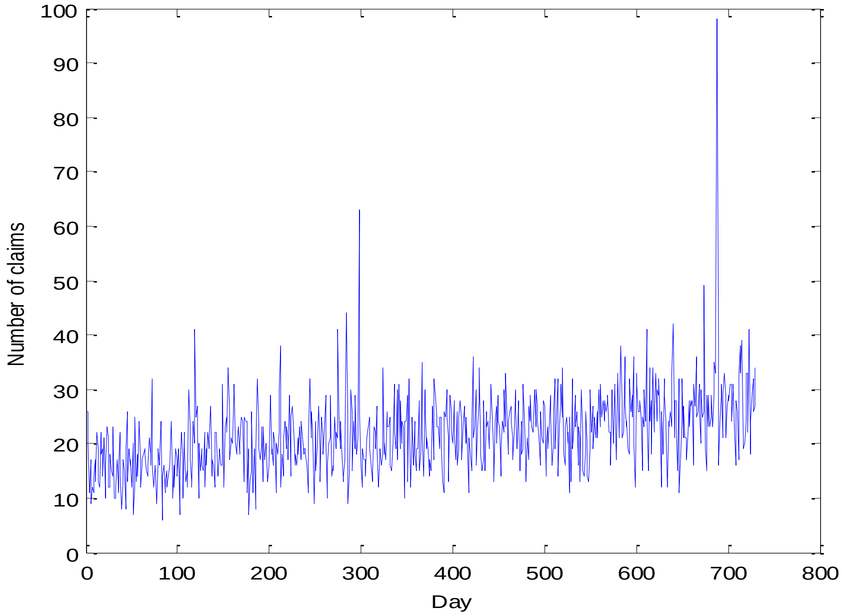

Figure 5.3.3 Time series of number of claims per day for Portfolio 1.

We include the above two graphs for the sake of completeness. The next two graphs show us the histogram and time series of the number of claims per month, respectively, which is what we are interested in.

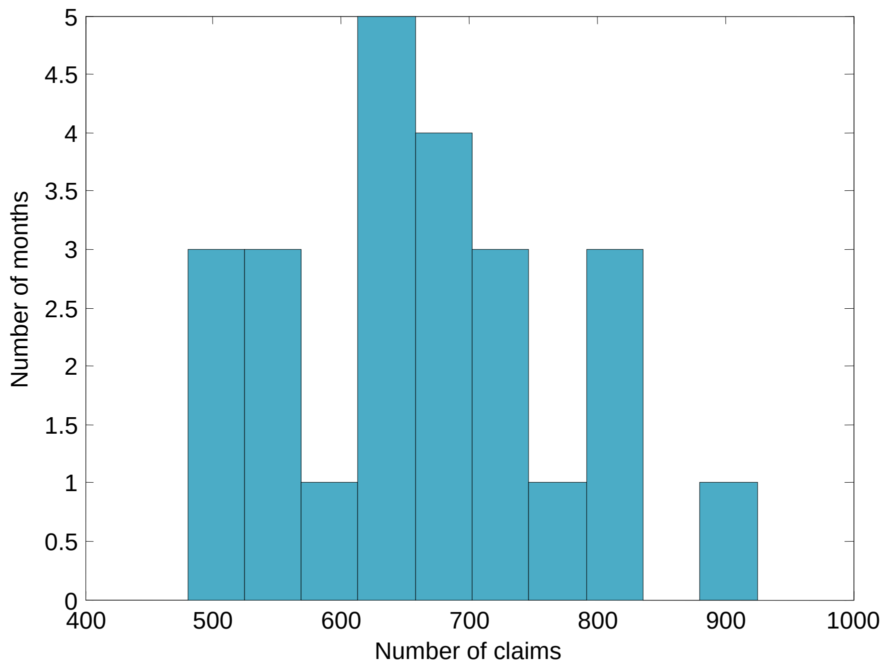

Figure 5.3.4 Histogram of number of claims per month for Porfolio 1.

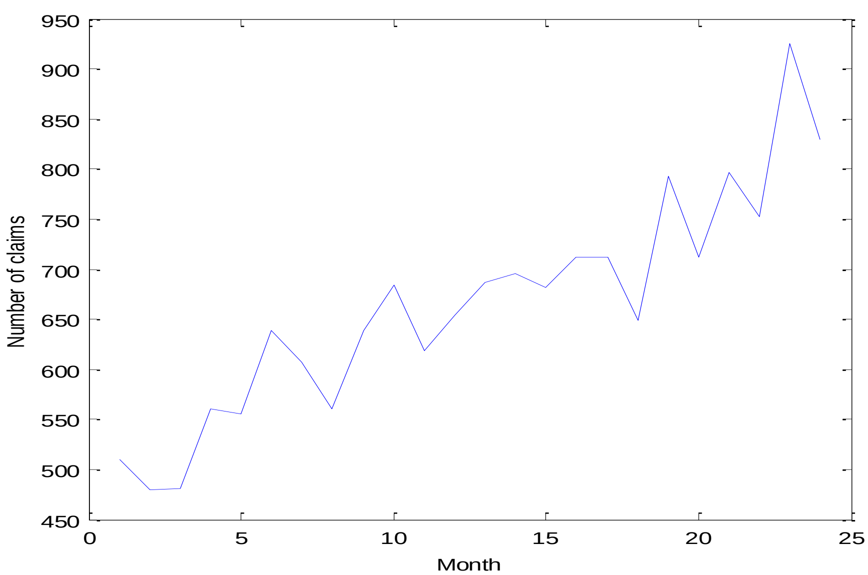

Figure 5.3.5 Time series of number of claims per month for Portfolio 1.

To estimate the probability of a claim exceeding (for example) R500 000 over the next year, we need to estimate the number of claims during the next year.

We determine the least squares regression line through the time series data points, and project the number of claims per month for the next 12 months, and add them up. A more refined time series model can be used to predict the number of claims during the next year more accurately, but we again follow the simplest approach possible for the sake of illustration.

The regression equation is $y = 485.4457 + 14.2843x$. Substituting $x = 25, 26, \ldots, 36$ into the equation, we obtain the following values:

Date	***x***	**Estimated number of claims**
2006 July	25	842.55
2006 August	26	856.84
2006 September	27	871.12
2006 October	28	885.41
2006 November	29	899.69
2006 December	30	913.98
2007 January	31	928.26
2007 February	32	942.54
2007 March	33	956.83
2007 April	34	971.11
2007 May	35	985.40
2007 June	36	999.68
Total		**11053.42**

Table 5.3.3 Predicted number of claims per month over the next year.

We show these predicted values in Figure 5.3.6.

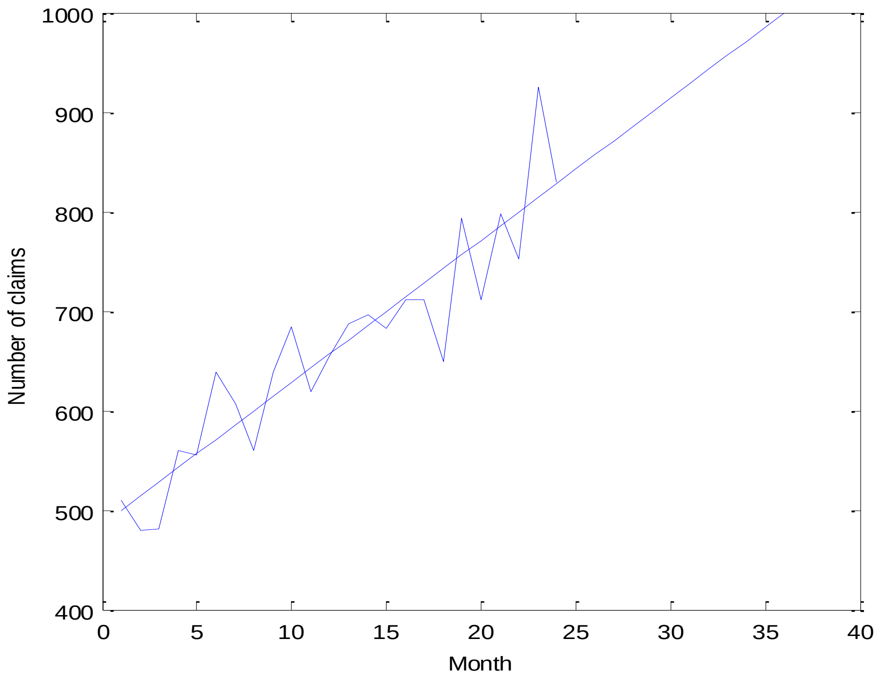

Figure 5.3.6 Fitted regression line and forecast for number of claims per month.

We can see that a linear trend is a good model for the number of claims per month.

From Table 5.3.3 the estimated the number of claims during the next year is $11\,053$.

The simplest way to proceed from here is to assume a binomial model. We calculate the probabillity of getting at least one claim exceeding R$500\,000$ over the next year, $P(X > 0)$, where X follows a binomial distribution with number of repetitions $n = 11\,053$ and probability of success $p = 4.9139 \times 10^{-4}$.

If we substitute these values into the equation

$$P(X > 0) = 1 - P(X = 0) = 1 - (1 - p)^n,$$

we obtain the estimated probability, namely 0.9957.

We do similar calculations for the amounts shown in Table 5.3.2, and for time periods one month, six months, one year (as illustrated for R$500\,000$), two years, five years and ten years. We obtain the following table:

	Claim amount (R million)				
Time period	**0.5**	**1**	**2**	**5**	**10**
Next claim	4.9319E-04	1.2534E-04	3.1222E-05	4.8586E-06	1.1736E-06
1 month	0.340230	0.100278	0.025977	4.0875E-03	9.8882E-04
6 months	0.925709	0.483455	0.151717	0.025280	6.1656E-03
1 year	0.995715	0.749798	0.291850	0.052286	0.012888
2 years	0.999993	0.951633	0.529733	0.110775	0.027960
5 years	1.000000	0.999926	0.906313	0.308205	0.085154
10 years	1.000000	1.000000	0.998238	0.627224	0.212071

Table 5.3.4 Probability the given amount will be exceeded over the specified period.

The above table can be used to estimate the risk of the portfolio.

Note that the claim amounts have not been adjusted for inflation in the calculation of the probabilities in the table above. (The original data of Portfolio 1 have been adjusted for inflation. Refer to Section 5.1.) This property is in fact desirable, since one requires the risk of the portfolios with respect to the current value of money.

As we mentioned at the beginning of this section, we used the simplest method available to construct Table 5.3.4 above. The assumptions made, were the following:

- The number of claims per month grows linearly.
- The value of the trend is a sufficiently accurate estimate of the number of claims per month. (We assumed it is sufficient to substitute time into the least squares regression equation and not to bring any seasonal or cyclical effect into account.)
- Subsequent claims are independent. (For the binomial model to hold.)

In the remaining part of this section, the focus will be on quantile estimation, as described in Section 3.2.5.

As an example we show the quantile associated with an exceedance probability of $p_1 = 10^{-5}$. Therefore, the claim size which will be exceeded only once in $100\,000$ claims on average, is required.

For Portfolio 1, the parameter estimates were $\hat{\gamma} = 0.4613$, $\hat{\rho} = -0.0277$ and $\hat{c} = -17.8846$, at $k/n = 0.15$.

$p_2 = np_1/k$ = $10^{-5}/0.15$ = 6.666710^{-5}.

Solving for z numerically from the equation $(1-\hat{c})z^{-1/\hat{\gamma}} + \hat{c}z^{-(1-\hat{\rho})/\hat{\gamma}} = p_2$, yields $\hat{z} = 190.6442$.

The number of excesses $k = 2430$ corresponds to a threshold of t = 18 404.93, as shown earlier in this section.

The quantile is therefore estimated as $\hat{x} = t\hat{z}$ = 3 508 632.89.

Table 5.3.5 below also includes the quantiles associated with $p_1 = 10^{-3}$, $p_1 = 10^{-4}$ and $p_1 = 10^{-6}$. Also included are the time spans of Table 5.3.4. For instance, the predicted number of claims in the next month is $842.55 \approx 843$. The upper probability associated with a month is therefore $p_1 = 1/843 = 0.001186$. From Table 5.3.5 we see that the claim size exceeded once every month on average, is approximately R318 425.36.

For the sake of comparison, Table 5.3.5 also shows empirical estimates. For example, consider $p_1 = 10^{-3}$. The total number of claims in the portfolio is 16 197. The empirical estimate of the claim size exceeded only once in 1000 claims, is the 0.001(16197) = 16.197-th largest claim. The empirically estimated quantile is calculated as the average of the 16-th and 17-th largest claim, which is 230 542.39.

The smallest value of p_1 for which the empirical estimate exists, is $1/16197 = 6.174 \times 10^{-4}$. For smaller values of p_1, NA (not available) is given as an entry in the table.

Time period	One claim in	Exceedance probability	Empirical quantile (R)	Model quantile (R)
	1 000	1.0000E-03	230 542.39	347 834.17
	10 000	1.0000E-04	462 291.61	1 119 913.72
	100 000	1.0000E-05	NA	3 508 632.89
	1 000 000	1.0000E-06	NA	10 808 934.99
1 month	843	1.1862E-03	216 238.14	318 425.36
6 months	5 270	1.8975E-04	298 088.99	811 700.06
1 year	11 053	9.0473E-05	462 291.61	1 177 469.20
2 years	24 164	4.1384E-05	NA	1 738 992.89
5 years	75 837	1.3186E-05	NA	3 062 049.54
10 years	203 097	4.9238E-06	NA	4 967 511.89

Table 5.3.5 Quantile estimation for Portfolio 1.

Considering Table 5.3.5, one immediately observes the great difference between the empirical quantiles and those obtained from the fitted PPD model. (The main values of interest are those of one claim in 1000 and one in the next month. The other empirical quantiles are based on too few observations to be reliable.)

The estimates obtained by fitting the PPD model are much larger than those obtained empirically. The same holds for the exceedance probabilities of a claim of size R$500\,000$, which were calculated earlier in this section, where the model probability was much larger than the empirical probability.

It may seem at first that the estimated value of the EVI ($\hat{\gamma} = 0.4613$) is too large. This is not the case, as one can see by choosing a threshold yielding a smaller value of $\hat{\gamma}$.

Figure 5.2.2 showed how rapidly $\hat{\gamma}$ (based on the PPD) increases as a function of k for Portfolio 1. The lowest threshold which we considered there was $k = 5\%$ of n. At that threshold the estimated PPD parameters are $\hat{\gamma} = 0.375$, $\hat{\rho} = -0.0593$ and $\hat{c} = -7.9812$. Note that $\hat{\gamma} = 0.375$ is even far below the GJ(ρ_3) estimate of the EVI, namely 0.4402.

If we redo Table 5.3.5 with this parameter set, we obtain the following table:

		Exceedance	**Empirical**	**Model**
Time period	**One claim in**	**probability**	**quantile (R)**	**quantile (R)**
	1 000	1.0000E-03	230 542.39	427 848.75
	10 000	1.0000E-04	462 291.61	1 092 701.22
	100 000	1.0000E-05	NA	2 731 995.74
	1 000 000	1.0000E-06	NA	6 742 629.97
1 month	843	1.1862E-03	216 238.14	398 602.95
6 months	5 270	1.8975E-04	298 088.99	844 062.18
1 year	11 053	9.0473E-05	462 291.61	1 137 526.58
2 years	24 164	4.1384E-05	NA	1 555 358.43
5 years	75 837	1.3186E-05	NA	2 449 185.61
10 years	203 097	4.9238E-06	NA	3 611 603.71

Table 5.3.6 Effect of lower EVI on quantile estimation for Portfolio 1.

From Table 5.3.6 one can see that the model quantiles for one claim in 1000 and once in the next month are even higher than before. However, the final entry in the table (model quantiles once every 10 years) decreased from R4.97 million to R3.61 million. The effect of the heavier tail is seen in extreme quantiles and does not necessarily have the same effect on quantiles within the range of the data.

The correct table to use to quantify the risk of the portfolio is therefore Table 5.3.5, and in that to consider the quantiles obtained from the PPD model.

Conclusion

In this chapter an illustration was given of how inferences can be made when one has real data at one’s disposal. In Section 5.2 we demonstrated the use of both graphical and numerical techniques to obtain an estimate of the EVI. The use of the PPD estimator for two different sample sizes in combination with another estimator was demonstrated. It was also shown that discretion has to be exercised in some cases. In certain instances some of the estimators need to be dismissed. Also, a technique based on resampling was developed to enable the analyst to distinguish between two sets of parameter estimates.

In the final section an example of how inferences can be made once the choice of threshold has been made, the PPD model has been fitted and the EVI estimate has been obtained. It was shown how tables containing exceedance probability estimates, and tables containing (large) quantile estimates can be constructed, for given time spans.

Finally, it was also shown that empirical quantiles may differ greatly from quantile estimates obtained from the model, and that the difference is not due to overestimation of the EVI, but rather that the empirical quantiles are based on too few observations, and that the strength of EVT does not lie in estimating quantiles within the range of the data, but in the estimation of quantiles far into the tail of the distribution.

Chapter 6

Conclusions and further research areas

Even though several estimators of the EVI exist in the literature, there is still room for improvement, since many estimators do not produce adequate results, and in many instances a great deal of subjectivity is involved.

The first aim of the study was therefore to develop an improved method of estimating the EVI in the case where it is assumed to be positive. This was achieved by constructing an estimator based on the estimation of the PPD parameters, of which one is the EVI.

Bayesian methodology for the estimation of PPD parameters received much attention. Specific issues arising from the parameterization of the PPD were addressed from a theoretical as well as a practical point of view. In order to implement the Bayesian methodology for the estimation of the PPD parameters, significant computational and numerical obstacles had to be overcome, but the end result is an extremely reliable Gibbs sampler. A detailed description was given of how to apply the results from the Gibbs sampler for the purpose of inferences concerning the estimation of parameters, extreme quantiles and exceedance probabilities.

A measure was defined which quantifies the instability of estimates over a range of thresholds. By applying this measure to the Bayesian estimates of the EVI, a threshold selection technique was developed, which significantly improves the estimation accuracy. Combining the Bayesian methodology with the threshold selection process resulted in an estimator which we call the PPD estimator.

The second aim of the study was to conduct an extensive simulation study to compare estimators in the literature to each other, as well as to the PPD estimator. The design of the simulation study used was thoroughly justified by incorporating theoretical and practical considerations. The simulation study compared the performance of the PPD estimator to a wide range of estimators in the literature. Comparing the simulation results, the PPD estimator turned out to be the method of choice.

For the case of a positive EVI, we put the PPD estimator forward as a benchmark estimator against which the performance of other estimators can be measured.

The PPD estimator is complex and computationally intensive. As a byproduct of the simulation study, the third aim of the study was achieved, namely providing suggestions as to which estimators can be used as alternatives to the PPD estimator if the use of the PPD estimator is not possible from a practical point of view.

The fourth aim of the study was to illustrate the use of the proposed estimator in practice. A case study of five insurance claims data sets illustrated how data sets can be analyzed in practice. It was shown to what extent discretion can/should be used, as well as how different estimators can be used in a complementary fashion to give more insight into the nature of the data and the extreme tail of the underlying distribution. The analysis was carried out from the point of raw data, to the construction of tables which can be used directly to gauge the risk of the insurance portfolio over a given time frame.

Further research possibilities include the following:

- The simulation study can be extended indefinitely by considering more estimators from the literature.
- It was found that external estimation of the second order parameter ρ might lead to more accurate estimation of the EVI for very large samples. For samples of sizes $n = 10\,000$ and $n = 20\,000$ it appeared as though it was marginally better to first estimate ρ using Gomes and Martins (2001), and then determining the Bayesian estimates of the two remaining parameters. This idea should be pursued further.
- The only estimator which was considered for the external estimator of the second order parameter for the PPD estimator was the estimator by Gomes and Martins (2001). The performance of other estimators of ρ in this respect needs to be investigated.
- Some results were shown for sample sizes of up to $n = 100\,000$. More research needs to be done for samples of sizes exceeding $n = 20\,000$.
- The estimation accuracy was significantly improved by applying the instability measure to obtain a threshold selection procedure in the case of the PPD estimator. In Section 4.9.2 some graphs were considered which compared the mean bias of the PPD estimator over a range of values of k, to that of the GJ(ρ_3) estimator. It was observed that the PPD estimates are in many cases too low on average for some k and too high on average for other k. Threshold selection is therefore vital in the case of the PPD estimator, because of the averaging which occurs. In the case of the GJ(ρ_3) estimator we observed a relatively constant bias as a function of k. The GJ(ρ_3) estimator would most probably not benefit greatly from a threshold selection procedure which leads to averaging over a wide range of values of k. This may also be true for many of the other estimators considered, but whether or not the instability measure leads to a useful threshold selection procedure for these other estimators, still needs to be investigated.
- Some additional analyses within the Bayesian paradigm will be valuable in order to obtain more insight into the results. Two possible additions are: to include prediction (instead of only estimation) of quantiles and exceedance probabilities, and calculating the MCMC error when posterior analysis is performed (the convergence of the relevant posterior distributions). The latter will give a further indication of estimation accuracy.

- An extension of the PPD estimator to other more complex settings is still an open problem. These include, amongst others, conditional distributions with covariates, and multivariate settings.

References

Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J.L. 2004. *Statistics of extremes*. Chichester: John Wiley & Sons.

Beirlant, J., Joossens, E. & Segers, J. 2009. Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference*, vol.139, pp.2800-2815.

Bickel, D.R. & Frühwirth, R. 2006. On a fast, robust estimator of the mode: comparisons to other robust estimators with applications. *Computational Statistics & Data Analysis*, vol.50, no.12, pp.3500-3530.

Bingham, N.H., Goldie, C.M. & Teugels, J.L. 1987. *Regular variation*. Cambridge: University Press.

Bingham, N.H. 2007. Regular variation and probability: the early years. *Journal of Computational and Applied Mathematics*, vol.200, no.1, pp.357-363.

Coles, S.G. & Powell, E.A. 1996. Bayesian methods in extreme value modelling: a review and new developments. *International Statistical Review*, vol.64, no.1, pp.119-136.

Coles, S.G. 2001. *An introduction to statistical modeling of extreme values*. London: Springer.

Csörgő, S. & Viharos, L. 1998. Estimating the tail index. In *Asymptotic methods in probability and statistics*. Edited by B. Szyszkowicz. Amsterdam: Elsevier, pp.833-881.

De Haan, L. 1970. *On regular variation and sample extremes*. PhD thesis. Amsterdam: University of Amsterdam.

De Haan, L. & Ferreira, A. 2006. *Extreme value theory: an introduction*. Boston: Springer.

Dekkers, A.L.M., Einmahl, J.H.J. & De Haan, L. 1989. A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, vol.17, pp.1833-1855.

Dierckx, G. 2000. *Estimation of the extreme value index*. PhD thesis. Leuven: Katholieke Universiteit Leuven.

Drees, H. & Kaufmann, E. 1998. Selecting the optimal sampling fraction in univariate extreme value estimation. *Stochastic Processes and their Applications*, vol.75, pp.149-172.

Embrechts, P., Klüppelberg, C. & Mikosch, T. 1997. *Modelling extremal events for insurance and finance*. Berlin: Springer.

Feuerverger, A. & Hall, P. 1999. Estimating a tail exponent by modelling departure from a Pareto distribution. *Annals of Statistics*, vol.27, no.2, pp.760-781.

Fisher, R.A. & Tippett, L.H.C. 1928. On the estimation of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, vol.24, pp.180-190.

Fraga Alves, M.I. 2002. *Estimation of first and second order parameters in heavy tails*. Technical report. Lisbon: University of Lisbon.

Fraga Alves, M.I., Gomes, M.I. & De Haan, L. 2003. A new class of semi-parametric estimators of the second order parameter. *Portugaliæ Mathematica*, vol.60, no.2, pp.193-213.

Geluk, J., De Haan, L., Resnick, S. & Starica, C. 1997. Second-order regular variation, convolution and the central limit theorem. *Stochastic Processes and their Applications*, vol.69, pp.139-159.

Gnedenko, B.V. 1943. Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, vol.44, no.3, pp.423-453.

Goegebeur, Y., Beirlant, J. & De Wet, T. 2008. Linking Pareto-tail kernel goodness-of-fit statistics with tail index at optimal threshold and second order estimation. *REVSTAT Statistical Journal*, vol.6, pp.51-69.

Gomes, M.I., Martins, M.J. & Neves, M. 2000. Alternatives to a semi-parametric estimator of parameters of rare events – the jackknife methodology. *Extremes*, vol.3, no.3, pp.207-229.

Gomes, M.I. & Martins, M.J. 2001. Generalizations of the Hill estimator - asymptotic versus finite sample behaviour. *Journal of Statistical Planning and Inference*, vol.93, pp.161-180.

Gomes, M.I. & Martins, M.J. 2002. "Asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes*, vol.5, no.1, pp.5-31.

Gomes, M.I., De Haan, L. & Peng, L. 2002. Semi-parametric estimation of the second order parameter in statistics of extremes. *Extremes*, vol.5, no.4, pp.387-414.

Gomes, M.I., Caeiro, F. & Figueiredo, F. 2004. Bias reduction of a tail index estimator through an external estimation of the second-order parameter. *Statistics: A Journal of Theoretical and Applied Statistics*, vol.38, no.6, pp.497-510.

Gomes, M.I. & Martins, M.J. 2004. Bias reduction and explicit semi-parametric estimation of the tail index. *Journal of Statistical Planning and Inference*, vol.124, pp.361-378.

Gomes, M.I., Martins, M.J. & Neves, M. 2007. Improving second order reduced bias extreme value index estimation. *REVSTAT Statistical Journal*, vol.5, no.2, pp.117-207.

Guillou, A. & Hall, P. 2001. A diagnostic for selecting the threshold in extreme value analysis. *Journal of the Royal Statistical Society*, vol.63, pp.293-305.

Gumbel, E.J. 1958. *Statistics of extremes*. Columbia: University Press.

Hill, B.M. 1975. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, vol.3, pp.1163-1174.

Karamata, J. 1930. Sur un mode de croissance régulière des fonctions. *Mathematica (Cluj)*, vol.4, pp.38-53.

Leadbetter, M.R., Lindgren, G. & Rootzén, H. 1983. *Extremes and related properties of random sequences and processes*. New York: Springer.

Matthys, G. & Beirlant, J. 2003. Estimating the extreme value index and high quantiles with exponential regression models. *Statistica Sinica*, vol.13, pp.853-880.

Pareto, V. 1897. *Cours d'économie politique*. Lausanne: Rouge & Cie.

Peng, L. 1998. Asymptotically unbiased estimator for the extreme-value index. *Statistics and Probability Letters*, vol.38, no.2, pp.107-115.

Pickands, J. 1975. Statistical inference using extreme order statistics. *Annals of Statistics*, vol.3, pp.119-131.

Rice, J.A. 2006. *Mathematical statistics and data analysis*. 3rd ed. Belmont: Duxbury Press.

Zellner, A. 1971. *An introduction to Bayesian inference in econometrics*. New York: John Wiley & Sons.

Zipf, G.K. 1949. *Human behavior and the principle of least effort: an introduction to human ecology*. Reading, Massachusetts: Addison-Wesley.