Empirical Bayes estimation of the Extreme Value Index in an ANOVA setting

Aletta Gertruida Jordaan

Thesis presented in partial fulfillment of the requirements for the degree of Master of Commerce in the Faculty of Economic and Management Sciences at Stellenbosch University

Supervisor: Dr. T.L. Berning

March 2014

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirely or in part submitted it for obtaining any qualification.

Signature: ……………………..... Date: ………………………
A.G. Jordaan

Abstract

Extreme value theory (EVT) involves the development of statistical models and techniques in order to describe and model extreme events. In order to make inferences about extreme quantiles, it is necessary to estimate the extreme value index (EVI). Numerous estimators of the EVI exist in the literature. However, these estimators are only applicable in the single sample setting. The aim of this study is to obtain an improved estimator of the EVI that is applicable to an ANOVA setting.

An ANOVA setting lends itself naturally to empirical Bayes (EB) estimators, which are the main estimators under consideration in this study. EB estimators have not received much attention in the literature.

The study begins with a literature study, covering the areas of application of EVT, Bayesian theory and EB theory. Different estimation methods of the EVI are discussed, focusing also on possible methods of determining the optimal threshold. Specifically, two adaptive methods of threshold selection are considered.

A simulation study is carried out to compare the performance of different estimation methods, applied only in the single sample setting. First order and second order estimation methods are considered. In the case of second order estimation, possible methods of estimating the second order parameter are also explored.

With regards to obtaining an estimator that is applicable to an ANOVA setting, a first order EB estimator and a second order EB estimator of the EVI are derived. A case study of five insurance claims portfolios is used to examine whether the two EB estimators improve the accuracy of estimating the EVI, when compared to viewing the portfolios in isolation.

The results showed that the first order EB estimator performed better than the Hill estimator. However, the second order EB estimator did not perform better than the “benchmark” second order estimator, namely fitting the perturbed Pareto distribution to all observations above a pre-determined threshold by means of maximum likelihood estimation.

Opsomming

Ekstreemwaardeteorie (EWT) behels die ontwikkeling van statistiese modelle en tegnieke wat gebruik word om ekstreme gebeurtenisse te beskryf en te modelleer. Ten einde inferensies aangaande ekstreem kwantiele te maak, is dit nodig om die ekstreem waarde indeks (EWI) te beraam. Daar bestaan talle beramers van die EWI in die literatuur. Hierdie beramers is egter slegs van toepassing in die enkele steekproef geval. Die doel van hierdie studie is om 'n meer akkurate beramer van die EWI te verkry wat van toepassing is in 'n ANOVA opset.

'n ANOVA opset leen homself tot die gebruik van empiriese Bayes (EB) beramers, wat die fokus van hierdie studie sal wees. Hierdie beramers is nog nie in literatuur ondersoek nie.

Die studie begin met 'n literatuurstudie, wat die areas van toepassing vir EWT, Bayes teorie en EB teorie insluit. Verskillende metodes van EWI beraming word bespreek, insluitend 'n bespreking oor hoe die optimale drempel bepaal kan word. Spesifiek word twee aanpasbare metodes van drempelseleksie beskou.

'n Simulasiestudie is uitgevoer om die akkuraatheid van beraming van verskillende beramingsmetodes te vergelyk, in die enkele steekproef geval. Eerste orde en tweede orde beramingsmetodes word beskou. In die geval van tweede orde beraming, word moontlike beramingsmetodes van die tweede orde parameter ook ondersoek.

'n Eerste orde en 'n tweede orde EB beramer van die EWI is afgelei met die doel om 'n beramer te kry wat van toepassing is vir die ANAVA opset. 'n Gevallestudie van vyf versekeringsportefeuljes word gebruik om ondersoek in te stel of die twee EB beramers die akkuraatheid van beraming van die EWI verbeter, in vergelyking met die EWI beramers wat verkry word deur die portefeuljes afsonderlik te ontleed.

Die resultate toon dat die eerste orde EB beramer beter gevaar het as die Hill beramer. Die tweede orde EB beramer het egter slegter gevaar as die tweede orde beramer wat gebruik is as maatstaf, naamlik die passing van die gesteurde Pareto verdeling (PPD) aan alle waarnemings bo 'n gegewe drempel, met behulp van maksimum aanneemlikheidsberaming.

Aan my ouers, Jorrie en Gerda.

Baie dankie vir al julle liefde, wysheid en ondersteuning.

Dankie dat julle my geleer het van Sy liefde, en op Wie ek altyd kan vertrou.

Erkennings

Ek wil graag my opregte dank uitspreek teenoor die volgende persone:

- Aan my hemelse Skepper, aan wie alle eer toekom.
- My studieleier, Dr. Tom Berning, dankie vir al jou hulp en leiding met die skryf van die tesis. Dankie ook vir al jou motivering en geduld.
- Prof. Tertius de Wet en Prof. Neil le Roux, dankie dat julle deure altyd oop was.
- Ivona Contardo-Berning, baie dankie vir die taalversorging.
- Hildegard en Stephan, dankie vir twee jaar se saam-werk, saam-swot en saam-kuier.
- Hannes, dankie vir al jou insette, ure van proeflees, ondersteuning en net dat jy altyd daar is.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Problem statement

Extreme value theory (EVT) is a statistical field in which the emphasis is on studying extreme events, i.e. observations in the tails of the distributions. Fitting a distribution to a tail enables one to extrapolate beyond the data. Specifically, it enables the estimation of the probability of an outcome that is larger than the largest observed value in the data set.

The extreme value index (EVI) is the most significant parameter of the relevant limiting tail distribution. The limiting distribution is obtained by considering only the observations above a threshold and letting this threshold tend to infinity.

Numerous estimators of the EVI in the single sample setting exist in the literature. The aim of this study is to develop techniques that are applicable to an ANOVA-type setting.

The main application of the results from this thesis to real-world problems will be in the context of insurance claims data. It is generally known that the underlying distribution of insurance claims data is heavy-tailed. This thesis will be restricted to the right tail (large observations) of distributions where extremely large observations (claims) are likely to occur, i.e. heavy-tailed distributions.

Since an ANOVA setting implies different treatments, each insurance portfolio can be seen as a treatment. If the goal is to estimate the EVI for a specific insurance portfolio, the question can be asked whether the data from the other portfolios can be used to improve the estimate of the EVI of the portfolio in question. If an improved EVI estimate can be obtained, it will prove to be beneficial in the case of a new portfolio where a limited amount of data is available.

The question of simultaneous tail index estimation has received some attention in the literature (Beirlant and Goegebeur, 2004), however, only maximum likelihood type estimators have been used.

In this thesis, Bayesian techniques will be considered as a means of conducting simultaneous tail index estimation. To apply a Bayesian approach to an ANOVA setting, empirical Bayes (EB) techniques will be considered. The thesis will focus on obtaining an

EB estimator of the extreme value index in an ANOVA setting, applied in cases where a heavy-tailed distribution can be assumed.

If an EB estimator can be obtained which does improve the accuracy of the estimates, this would lead to a great improvement in extreme quantile estimation accuracy for ANOVA settings, especially when data are scarce.

1.2 Scope of the study

The objectives of this study are:

1. To provide an introduction to the fields of extreme value theory, Bayesian estimation and EB estimation. This will be done by means of a literature study.

2. To conduct a simulation study to investigate the performance of several EVI estimators proposed by the literature. The simulation study will cover a wide range of distributions and sample sizes.

3. To present the derivation of an EB estimator proposed by De Wet and Berning (2013) using limiting results from EVT and EB methodology.

4. This EB estimator is also generalised to a second order estimator. This is done by applying results from regular variation in EVT.

5. A case study will be presented, in which the performance of the EB estimator will be assessed.

1.3 Contribution of the study

The contribution of this study in the field of EVT can be summarised as follows:

1. Application of Bayesian methodology in EVT in an ANOVA setting has not received significant attention in the literature. The main purpose of this study is to address this problem.

2. Assessing the performance of a first order and a second order EB estimator of the EVI. The performance of these estimators will be assessed by means of insurance claims data.

3. The study will impact not only the field of insurance, but also many other fields of research. ANOVA-type settings are frequently used in research designs. By accurately predicting extreme events in these settings with only limited amounts of data available, will be beneficial as far as the quantification of risk is concerned.

1.4 Chapter outline

Chapter 2 gives a brief overview of extreme value theory. In section 2.2 the historical development of EVT is described and some areas of application are mentioned. Section 2.3 and 2.4 discuss the two approaches to extreme value theory, namely the block maxima approach and the threshold model approach. In section 2.5 the choice of threshold is discussed when considering the threshold model approach. Section 2.6 mentions some of the difficulties that are associated with EVT and also how these difficulties can be addressed.

Chapter 3 focuses specifically on the estimation of the EVI. A discussion of the different methods that can be used to estimate the EVI is presented in section 3.2. Section 3.3 highlights some important aspects regarding the estimation of the EVI when using the threshold model approach. In this section the well-known estimator of the EVI, namely the Hill estimator, is also introduced. Alternative estimators of the EVI are briefly discussed in section 3.4.

Bayesian methodology is introduced in chapter 4. The chapter begins with the historical development of Bayesian theory and in section 4.3 the basic paradigm of Bayesian theory is discussed. Section 4.4 discusses how inferences can be made by applying Bayesian methods and techniques. Section 4.5 mentions the different types of prior distributions and how an appropriate prior distribution can be obtained. Bayesian methodology in an extreme value context is discussed in section 4.6 and an explanation of how Bayesian estimates are obtained is provided in section 4.7. The concept of Gibbs sampling is introduced in section 4.8. The Gibbs sample algorithm is presented, which is used to simulate values from the desired posterior distribution. Various applications of Bayesian methods are discussed in section 4.9. The last section discusses some of the difficulties associated with the Bayesian approach.

The simulation study assessing the performance of various EVI estimators is presented in chapter 5. This includes the methodology, design and results of the simulation study which was conducted.

Empirical Bayes methodology is introduced in chapter 6. In section 6.2 the development of the field is discussed. An overview of EB methodology is also given in this section. In section 6.3 the two categories of EB methods is mentioned. Section 6.4 describes the EB methodology in terms of an ANOVA setting. The section begins with a description of how EB methods can be applied in the general ANOVA-type setting. A first order EB estimate and a second order EB estimate of the EVI are also derived in this section.

In chapter 7 the data of a specific case study are analysed. The chapter illustrates how the techniques described in the previous chapters can be applied in practice.

In the final chapter, the conclusion of this study is presented. Future research opportunities that stem from this study are identified.

CHAPTER 2

Overview of extreme value theory

2.1 Introduction

According to Aragones, Blanco and Dowd (2000), extreme value theory (EVT) can be defined as a specialist branch of statistics that attempts to make the best possible use of what little information there exists concerning the extremes of the distributions of interest.

In classical statistical methods the focus is generally on the entire area of support for the underlying distribution. Classical statistical measures accommodate the mass of central observations and these measures include, for example, obtaining the average value (mean) or measuring how much each observation deviates from the average value (standard deviation). Little attention is paid to the tails of the underlying distribution. In EVT the focus shifts to very large or very small observations, including observations usually referred to as outliers. EVT provides procedures and techniques for tail estimation. It also differs from classical statistical techniques in that extrapolation beyond the data is required.

As opposed to making use of empirical and physical guidelines, the nature of EVT modelling relies heavily on limiting arguments. In the EVT paradigm, the simplest starting point is to consider the maximum of a sequence of observations. The basic idea is that by letting the sample size n tend to infinity, the approximate behaviour of the maximum can be determined (under certain assumptions) and a family of models can be derived. More specifically, the Fisher-Tippet Theorem, together with the generalised extreme value (GEV) family of models, provides a framework for modelling the distribution of block maxima. A more detailed discussion will be given in section 2.3.

The above mentioned approach is a highly data intensive process, since only one maximum value is considered in each block. It will therefore prove difficult to apply the block maxima approach in practice. An alternative approach to extreme value theory is to consider not only the maxima, but all observations that exceed some high threshold t. These observations are regarded as extremes. The distribution of the tail from which these extreme observations are assumed to be generated from can be approximated despite the fact that the underlying distribution function F is unknown. A discussion of threshold models will be provided in section 2.4.

In the course of discussing the two approaches to EVT, namely the block maxima approach and the threshold model approach, the most important parameter in EVT, namely the extreme value index (EVI), will also be introduced.

In section 2.5 the choice of threshold is discussed, when considering the threshold model approach.

Some of the difficulties associated with EVT are mentioned in section 2.6, followed by a discussion of how these difficulties can possibly be addressed.

Before discussing these different aspects regarding EVT, it is necessary to begin with the historical development of the field. This is done in section 2.2, which also includes a discussion of various areas of application of EVT.

2.2 Historical development and areas of application

Extreme value theory has been used from as early as 1709 (Kotz & Nadarajah, 2000). The first area of application was in astronomy where EVT was used to either include or reject extremely large or extremely small observations.

Von Bortkiewicz (1922) considers the normal distribution and examines the distribution of different samples taken from this distribution. This paper serves as an introduction to examine the distribution of maxima, since it introduces the concept of distribution of largest value for the first time.

Frechét (1927) considers the asymptotic distributions of maxima and recognised one possible limiting distribution for the largest order statistic, namely the Frechét distribution.

Fisher and Tippet (1928) indicate that all limiting distributions of maxima fall in one of three classes of distributions. The three classes are referred to as the Gumbel, Frechét and Weibull families of distributions, and will be discussed in more detail in section 2.3.

Von Mises (1936) provides some useful conditions for the weak convergence of the largest order statistic, formulated according to each of the three classes of limiting distributions.

In 1943, Gnedenko provides a more detailed discussion and formulation of the conditions for the weak convergence of the extreme order statistics.

An important contribution to the field was made by Gumbel in 1958. In his book, *Statistics of extremes*, he discusses how EVT can be applied to certain distributions which have previously been treated by only considering empirical methods (Gumbel, 1958).

Another contributor to the field is Laurens de Haan. His PhD. thesis (1970) is still a leading reference in most papers referencing EVT. De Haan refined the result of Fisher and Tippet (1928) and the work done by Gnedenko (1943). Together with his co-authors, they introduce the moments estimator, which is a general estimator of the EVI (Dekkers, Einmahl and de Haan, 1989). His original work also led to a relatively new estimator of the EVI, namely the mixed moment (MM) estimator (Fraga Alves, Gomes, de Haan and Neves, 2009). This estimator is an interesting alternative to the more popular EVI estimators.

The theoretical advances made in the 1920s and 1930s were followed by numerous papers which discussed the practical applications of extreme value theory. These areas of application include flood analysis (Gumbel, 1941 & 1944), seismic analysis (Nordquist, 1945) and rainfall analysis (Potter, 1949).

Zipf (1949) used the Pareto distribution to apply extreme value theory to numerous areas such as economics, commerce, industry, travel, sociology, psychology and music.

Kotz and Nadarajah (2000) highlight some of the areas of application and mention numerous examples including horse racing, network design, queues in supermarkets, earthquakes, floods, ozone concentration and insurance.

Several examples are also discussed by Coles (2001). He considers the use of EVT in predicting the probability of extreme climate changes. In the same way extreme value techniques can be used in finance to model extreme events such as large market moves. Another application is reliability modelling, which involves considering extremely small events. This can be useful in areas such as quality control of systems, where the strength of the system is equal to that of the weakest component.

Beirlant, Goegebeur, Segers and Teugels (2004) mention the application of extreme value theory to numerous areas including hydrology, environmental research, geology and seismic analysis, metallurgy, insurance and finance.

From the above mentioned practical applications it is evident that on-going research in the field of EVT is justified.

2.3 Asymptotic models: the block maxima approach

The following theoretical overview of the block maxima approach is based on the theory as discussed by Coles (2001).

Suppose $X_1, X_2, \ldots X_n$ is a sequence of independent and identically distributed random variables. If X is a continuous random variable with density function $f(x), -\infty < x < \infty$, then the distribution function of X, denoted by $F(x)$, is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du.$$

If X is a discrete random variable with probability mass function $f(x) = P(X = x)$ for each $x \epsilon X$, then the distribution function $F(x)$ is given by

$$F(x) = P(X \leq x) = \sum\nolimits_{u \leq x} f(u).$$

Denote the maximum value of the random variables by $M_n = \max\{X_1, X_2, \ldots X_n\}$.

The distribution function of M_n is

$$\begin{aligned} P(M_n \leq x) &= P(X_1 \leq x, X_2 \leq x, \ldots X_n \leq x) \\ &= P(X_1 \leq x)P(X_2 \leq x) \ldots P(X_n \leq x) \\ &= [F(x)]^n. \end{aligned}$$

In practice the distribution function F is usually unknown. One approach that can be followed is to estimate F from the observed values. The problem with this approach is that small differences in the estimate of F can lead to large differences in the estimate F^n. Another approach is to use a limiting argument.

As n tends to infinity, it follows that for any fixed value of x

$$\lim_{n\to\infty} F_{M_n}(x) = \begin{cases} 1 & \text{if } F(x) = 1. \\ 0 & \text{if } F(x) < 1. \end{cases}$$

Therefore, the distribution of M_n is a degenerate distribution. A normalisation of M_n is required. This is done by choosing sequences of constants $a_n > 0$ and b_n and defining the normalised maximum M_n^* as

$$M_n^* = \frac{M_n - b_n}{a_n}.$$

The normalising constants a_n and b_n stabilise the location and scale of M_n^* as n increases. The limiting distribution of M_n^* is now of interest.

The extremal types theorem (Coles, 2001: 46) gives the entire range of possible limit distributions for M_n^*.

Theorem 2.3.1: Extremal Types Theorem

If sequences of constants $a_n > 0$ and b_n exist, such that

$$P\left(\frac{M_n - a_n}{b_n} \leq x\right) \to G(x)$$

as $n \to \infty$, for some non-degenerate distribution G, then G will belong to one of the three extreme value domains, defined as:

Gumbel:

$$G(x) = exp\left(-exp\left[-\left(\frac{x-b}{a}\right)\right]\right) \qquad \text{for all } x$$

Fréchet:

$$G(x) = \begin{cases} 0 & x \leq b \\ exp\left(-\left(\frac{x-b}{a}\right)^{-\frac{1}{\gamma}}\right) & x > b \end{cases}$$

Weibull:

$$G(x) = \begin{cases} exp\left(-\left[-\left(\frac{x-b}{a}\right)\right]^{\frac{1}{\gamma}}\right) & x < b \\ 1 & x \geq b \end{cases}$$

for $a > 0, -\infty < b < \infty$ and, in the case of Fréchet and Weibull domains, $\gamma > 0$. ■

Theorem 2.3.1 indicates that it is not required for the distribution function F to be known in order to determine F^n, since the limiting argument can be used.

From a practical point of view it is extremely difficult to know in advance which of the families of limiting distributions to use. Fisher and Tippett (1928) addressed this problem by combining the three families of distributions. The Fisher-Tippet Theorem states if the distribution for the normalised maximum of a sequence of random variables converges, it always converges to the generalised extreme value (GEV) distribution, regardless of the underlying distribution function F.

Theorem 2.3.2: Fisher-Tippett Theorem

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

$$P\left(\frac{M_n - a_n}{b_n} \leq x\right) \to G(x)$$

as $n \to \infty$, where G is some non-degenerate distribution, then G is a member of the GEV family

$$G(x) = exp\left\{-\left(1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\gamma}}\right\}$$

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

(Coles, 2001: 48)

Theorem 2.3.2 states that if the limit distribution exists, the distribution is GEV, regardless of the underlying distribution of F. This is one of the core theoretical principles of EVT and is similar in nature to the central limit theorem.

The parameter γ in Theorem 2.3.2 is known as the EVI. As mentioned previously the aim of this thesis is to find an alternative method of estimating the EVI, in particular using empirical Bayes methodology. A detailed discussion of different EVI estimators will be presented in chapter 3.

Considering Theorem 2.3.2 and the three domains mentioned in Theorem 2.3.1, it is evident that $\gamma = 0$ corresponds to the Gumbel distribution, $\gamma > 0$ corresponds to the Fréchet distribution and $\gamma < 0$ corresponds to the Weibull distribution.

The set of distributions for which the maximum converges in distribution to G_γ (the GEV distribution) is said to fall in the domain of attraction of G_γ.

Definition 2.3.1: Domain of attraction

Suppose G_γ is a non-degenerate limit of the sequence $\{M^*\}$ with EVI γ. The set of distributions of X satisfying the necessary and sufficient conditions for the limit distribution of $\{M^*\}$ to be precisely G_γ, is called the domain of attraction of G_γ, denoted $D(G_\gamma)$.

As an example, consider a sequence of independent standard exponential variables, denoted by $X \sim Exp(1)$. The distribution function of these variables is

$$F(x) = 1 - e^{-x}$$

for $x > 0$.

In order to obtain the limit distribution of the normalised maximum M_n^*, let the normalising constants $a_n = 1$ and $b_n = \log n$. As n tends to infinity, it follows that

$$\begin{aligned} P(M_n^* \leq x) = P\left(\frac{M_n - b_n}{a_n} \leq x\right) &= F^n(x + \log n) \\ &= \left(1 - e^{-(x+\log n)}\right)^n \\ &= (1 - n^{-1}e^{-x})^n \\ &\to \exp(-e^{-x}) \end{aligned}$$

for each fixed $x \in \mathbb{R}$ (Coles, 2001: 52).

Consider again the distribution function of the Gumbel family:

$$G(x) = exp\left(-exp\left[-\left(\frac{x-b}{a}\right)\right]\right) \qquad \text{for all } x.$$

Consequently, the limit distribution of the maximum as $n \to \infty$ is the Gumbel distribution, which corresponds to $\gamma = 0$ in the GEV family.

Distributions in the Gumbel domain have an infinite right endpoint and the right tail of such a distribution decays exponentially. Distributions in the Weibull domain have a finite right endpoint of support. Lastly, distributions in the Fréchet domain have an infinite right endpoint of support and the right tail of such a distribution decays polynomially. The distributions in the Fréchet domain are referred to as heavy-tailed.

The following tables provide a summary of some of the distributions in each domain (Berning, 2010 and Beirlant et al., 2004):

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

Table 2.3.1: Distributions in the Fréchet domain

Distribution	Notation	$\mathbf{1-F(x)}$	Conditions
Normal	$X \sim N(\mu, \sigma^2)$	$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) du$	$x \in \mathbb{R}$ $\mu \in \mathbb{R}, \sigma > 0$
Weibull	$X \sim Weibull(\lambda, \tau)$	$\exp(-\lambda x^{\tau})$	$x > 0$ $\lambda, \tau > 0$
Exponential	$X \sim \mathrm{Exp}(\lambda)$	$\exp(-\lambda x)$	$x > 0$ $\lambda > 0$
Gamma	$X \sim \Gamma(\lambda, m)$	$\frac{\lambda^m}{\Gamma(m)} \int_x^{\infty} u^{m-1} \exp(-\lambda u) du$	$x > 0$ $\lambda, m > 0$
Lognormal	$X \sim LogN(\mu, \sigma)$	$\int_x^{\infty} \frac{1}{\sqrt{2\pi}\sigma u} \exp\left(-\frac{1}{2\sigma^2}(\log(u) - u)^2\right) du$	$x > 0$ $\mu \in \mathcal{R}$ $\sigma > 0$

Table 2.3.2: Distributions in the Gumbel domain

Distribution	Notation	$1 - F\left(x_+ - \frac{1}{x}\right)$	Conditions
Uniform	$X \sim U(0,1)$ for $0 \leq x \leq 1$	$\frac{1}{x}$	$x > 1$
Beta	$X \sim Beta(p,q)$	$\int_{1-\frac{1}{x}}^{1} \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} u^{p-1}(1-u)^{q-1} du$	$x > 1$ $p,q > 0$
Reversed Burr	$X \sim Reversed\ Burr(\beta,\tau,\lambda)$	$\left(\frac{\beta}{\beta + x^{\tau}}\right)^{\lambda}$	$x > 0$ $\lambda,\beta,\tau > 0$
Extreme Value Weibull	$X \sim EV\ Weibull(\alpha)$	$1 - \exp(-x^{-\alpha})$	$x > 0$ $\alpha > 0$

Table 2.3.3: Distributions in the Weibull domain

Note that the second last column in table 2.3.3 does not give the survival function, but $1 - F\left(x_+ - \frac{1}{x}\right)$, where x_+ is the finite upper limit of the distribution. This function is mathematically more convenient in the extreme value context (refer to Beirlant, 2004: 65 - 69).

One of the main motivations for developing the field of EVT is to make inferences regarding the tails of heavy-tailed distributions. Only distributions in the Fréchet domain will be considered in this thesis. Heavy-tailed distributions (that is distributions from the Fréchet domain) frequently occur in practice, for example insurance claim data, sizes of the files transferred from a web-server and earthquake magnitudes (Berning, 2010: 18).

2.4 Threshold models

The block maxima (one observation per block) approach can lead to the loss of information if other extreme values are also present in the blocks.

The main idea behind the threshold model approach is to approximate the distribution of observations that exceed a pre-defined high threshold t. More specifically, the amount by which the threshold is exceeded is of importance.

The extreme observations which exceed the threshold are referred to as excesses and can be defined in two ways. The first is to define Z as the additive excess where $Z = X - t$, given $X \geq t$. The second is to define Z as the multiplicative excess where $Z = X/t$, given $X \geq t$.

The threshold model known as the generalised Pareto distribution will now be considered as an introduction to threshold models.

The generalised Pareto distribution (GPD) is defined by Coles (2001: 75) as follows:

Theorem 2.4.1: The generalised Pareto distribution

Let $X_1, X_2, \ldots$ *be a sequence of independent random variables with common distribution function* F, *and let*

$$M_n = max(X_1, \ldots X_n).$$

Suppose that, for large n, *the distribution function satisfies*

$$P(M_n \leq x) \approx G(x)$$

where

$$G(x) = exp\left\{-\left(1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\gamma}}\right\}$$

for some $\sigma > 0, \mu$ *and* γ.

Then, for large enough t, *the distribution of* $Z = X - t$ *conditional on* $X \geq t$ *is approximately*

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

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

Theorem 2.4.1 states that, if the Fisher-Tippett Theorem holds, the additive excesses will be approximately distributed generalised Pareto. Note that the parameter γ in Theorem 2.4.1 is the same parameter γ specified in the Fisher-Tippett Theorem, namely the EVI.

In the case of multiplicative excesses, the limiting distribution of the excesses is the Pareto distribution.

The Pareto survival function is defined as

$$\bar{F}(z) = 1 - F(z) = z^{-\alpha}$$

where $z \geq 1$ and $\alpha > 0$.

For the Pareto distribution the EVI is $\gamma = \frac{1}{\alpha}$.

Note that the Pareto distribution implies a heavy right-tail, since $\alpha > 0$. As discussed in section 2.3, heavy-tailed distributions frequently occur in practice. This serves as motivation for only considering positive values of the EVI and multiplicative excesses in this thesis.

The threshold t can also be defined in terms of the number of observations that exceed the threshold. If the number of excesses is denoted by k, the excesses $Z_1 \dots Z_k$ can be determined and an appropriate model can be fitted to these excesses. Defining the number of excesses is analogous to defining the threshold t. In particular the relationship can be stated as $t = X_{n-k}$.

In this thesis k will be specified and not t (unless stated otherwise). Note also that small k is associated with large t.

2.5 Threshold selection

Before the parameters of the limiting distribution can be estimated, the excesses have to be determined. In order to obtain the excesses, the choice of an appropriate threshold is required.

The advantage of a high threshold is that the limiting assumptions hold. This will lead to a model that reflects the true distribution of the underlying tail more accurately and in turn will lead to low bias in parameter estimation. However, a high threshold will result in few excesses. If the number of excesses is small, it will lead to a high variance in parameter estimation. Conversely too low a threshold will result in small variance, but also high bias (Coles, 2001: 78). Optimal threshold selection techniques are frequently based on this trade-off.

Two approaches of choosing the optimal choice of k will be considered. The first approach is to choose k as a fixed percentage of the total number of observations. The second approach involves using specific (usually adaptive) methods of threshold selection.

An example of two methods of threshold selection pertaining to the Hill estimator is the method of Guillou and Hall (2001) and the method of Drees and Kaufmann (1998). The Hill estimator will be defined in section 3.3.2 and a further discussion of the two methods will be presented in that section.

2.6 Limitations of EVT and handling thereof

In order to estimate beyond or at the limit of available data, certain assumptions have to be made about the distribution of the tail. However, it proves difficult to validate these assumptions in practice.

As mentioned in section 2.5, the choice of threshold is critical. There are various methods that can be used to determine a threshold and each method will produce different results. Deciding which method is “best” can prove difficult (Embrechts, 2000).

It is also necessary to choose appropriate methods to estimate the unknown parameters of the model. Various approaches have been proposed by Coles (2001). The most common method of estimation is maximum likelihood estimation. Wackerly, Mendenhall and Scheaffer (2008: 477) defines maximum likelihood estimation as the method of selecting as estimates the values of the parameters that maximise the joint probability function or the joint density function, referred to as the likelihood function. Let $x_1, x_2, \dots, x_n$ denote a sample of n observations and let the likelihood function be denoted by

$$L\left(x_1, x_2, \dots, x_n \middle| \theta_1, \theta_2, \dots, \theta_p\right),$$

which depends on the parameters $\theta_1, \theta_2, \dots, \theta_p$. The method of maximum likelihood is used to obtain those values of the parameters that maximise the likelihood function.

It is important not to underestimate the effects of different sources of uncertainty. Not only can model assumptions, threshold selection techniques and different methods of parameter estimation have a significant impact on the variability in estimation, but also the extrapolation beyond the data associated with EVT.

Due to the above mentioned limitations and the fact that EVT is associated with high uncertainty, it is important to include all possible information available when developing a model or estimating a specific parameter. This can be done by using covariate information or constructing multivariate models (Beirlant et. al., 2004). This approach will not be explored in this thesis.

Another way of incorporating additional prior knowledge is to use methods developed in the Bayesian paradigm. These methods use additional information together with the information provided by the data to obtain a specific posterior distribution. A detailed discussion of Bayesian statistics and methods will be given chapter 4.

2.7 Conclusion

Section 2.2 provided an overview of the historical development of EVT and presented a number of areas where EVT are applied. The numerous areas of application justify on-going research in the field.

In section 2.3 a detailed discussion is given of the classical approach to EVT, namely using asymptotic models for the maximum of a series. In this section the Extremal Types Theorem was introduced, which states that the limit distribution of the normalised maximum can belong to one of three domains of attraction. The three domains are the Weibull domain, the Gumbel domain and the Fréchet domain.

The Fisher-Tippett Theorem was also stated. This theorem combines the three possible domains of attraction and states that the GEV distribution can be used as the limiting distribution of the normalised maximum.

Section 2.4 presented another approach to EVT. This approach uses threshold models, where not only the maximum of a series is considered, but all the observations above a pre-defined threshold. Specifically the GPD was provided as an example of such a threshold model. In section 2.5 the importance of selecting the optimal threshold was mentioned.

The last section mentioned some of the limitations of EVT. Specifically, EVT is associated with a high level of uncertainty. Due to this high level of uncertainty, it is necessary to include all available information when estimating a parameter.

In summary, the following assumptions will be made in this thesis:

- only heavy-tailed distributions (that is distributions from the Fréchet domain) will be considered;
- the threshold model approach will be used;
- only multiplicative excesses will be considered.

CHAPTER 3

Estimating the extreme value index

3.1 Introduction

The chapter begins with a brief description of the different methods that can be used to estimate the EVI and mentions the three groups in which these methods can be categorised.

In section 3.3 certain aspects regarding the estimation of the EVI when considering threshold models are presented. The most well-known estimator of the EVI, namely the Hill estimator, is introduced. Two methods that will be used to determine the optimal threshold of the Hill estimator are also discussed. The perturbed Pareto distribution (PPD), with the EVI being one of its parameters, is also defined in this section.

In the last section five alternative estimates of the EVI are defined and briefly discussed.

3.2 Approaches to parameter estimation

The EVI is the most important parameter when doing inference with regards to the extreme quantiles and small exceedance probabilities of a population. This section focuses on the estimation of these parameters.

Numerous methods exist which can be used to estimate the EVI parameter. The choice of method depends on whether the EVI is positive, negative or zero. According to Beirlant et al. (2004: 131-132), the available methods used to estimate the EVI for all domains of attraction can generally be divided into three groups, namely:

- the method of block maxima;
- the quantile view; and
- the tail probability view.

The method of block maxima is based on the results given in section 2.4. Since the Fisher-Tippett Theorem provides a model for the distribution of block maxima, the EVI can be estimated by fitting the generalised extreme value (GEV) distribution to the maxima of the subsamples (Gumbel, 1958). Maximum likelihood (ML) estimation or the method of probability-weighted moments can then be used to estimate the EVI.

Let $Y_1, Y_2 \dots, Y_m$ be a sample of independent sample maxima, where Y denote the maximum of a subsample $X_1, X_2 \dots, X_n$. The ML method involves maximising the log-likelihood function of the independent and identically distributed GEV random variables for a sample $Y_1, Y_2 \dots, Y_m$. In the case of $\gamma \neq 0$, the log-likelihood function is given by:

$$\log L(\sigma,\gamma,\mu) = -m\log\sigma - \left(\frac{1}{\gamma}+1\right)\sum_{i=1}^{m}\log\left(1+\gamma\frac{Y_i-\mu}{\sigma}\right) - \sum_{i=1}^{m}\left(1+\gamma\frac{Y_i-\mu}{\sigma}\right)^{-\frac{1}{\gamma}}$$

given that $1+\gamma\frac{Y_i-\mu}{\sigma} > 0$ for $i = 1,2,\dots,m$.

In the case of $\gamma = 0$, the log-likelihood function is given by:

$$\log L(\sigma,\gamma,\mu) = -m\log\sigma - \sum_{i=1}^{m}\exp\left(-\frac{Y_i-\mu}{\sigma}\right) - \sum_{i=1}^{m}\frac{Y_i-\mu}{\sigma}.$$

By maximising the log-likelihood function, the ML estimator $(\hat{\sigma},\hat{\gamma},\hat{\mu})$ of (σ,γ,μ) is obtained (Coles, 2001: 55).

The method of probability-weighted moments involves obtaining the probability-weighted moments of a random variable Y with distribution function F (Greenwood, Landwehr, Matalas and Wallis, 1979). In general, these probability-weighted moments are defined as

$$M_{p,r,s} = E(Y^p[F(Y)]^r[1-F(Y)]^s)$$

for real p, r and s.

In order to estimate the parameters for the GEV distribution, the probability-weighted moments are obtained by setting $p = 1$, $r = 0, 1, 2, \dots$ and $s = 0$. The resulting moments are

$$M_{1,r,0} = \frac{1}{r+1}\left(\mu - \frac{\sigma}{\gamma}\{1-(r+1)^{\gamma}\Gamma(1-\gamma)\}\right)$$

for $\gamma < 1$ and $\gamma \neq 0$.

Considering a sample of $Y_1, Y_2 \dots, Y_m$ independent and identically distributed GEV random variables, the probability-weighted moments estimator $(\hat{\sigma}, \hat{\gamma}, \hat{\mu})$ of (σ, γ, μ) is obtained by solving the system of equations, obtained from the above equation with $r = 0, 1, 2$. Specifically, in order to obtain $\hat{\gamma}$, the following equation has to be solved numerically (Hosking, Wallis and Wood,1985):

$$\frac{3M_{1,2,0} - M_{1,0,0}}{2M_{1,1,0} - M_{1,0,0}} = \frac{3^\gamma - 1}{2^\gamma - 1}.$$

There are several drawbacks when using the GEV distribution to estimate the EVI. Firstly, the GEV distribution only considers the maximum. Useful information therefore may be lost. Another drawback of using the GEV distribution is determining the appropriate size of the blocks.

Estimation methods that form part of the quantile view group are based on observations that exceed some high threshold, therefore making use of the generalised Pareto distribution (GPD). The main difficulty with these estimation methods is to determine the optimal threshold. However, if the threshold is determined, the parameters of the GPD can be estimated by using maximum likelihood or Bayesian methods of estimation. Beirlant et al. (2004: 140) argue that estimators based on the extreme order statistics will be more reliable than estimators based on a single largest order statistic. For this reason only estimators based on the extreme order statistics will be considered in this thesis.

Estimation methods that form part of the quantile view group are based on the following general condition:

$$\frac{U(xu) - U(x)}{a(x)} \to \frac{u^\gamma - 1}{\gamma} \text{ for any } u > 0 \text{ as } x \to \infty,$$

for some regularly varying function a with index γ, where U is the tail quantile function, i.e. $U(x) = Q\left(1 - \frac{1}{x}\right)$.

This condition is necessary in order for a non-degenerate limit distribution of a normalized maximum to exist.

The third group of estimation methods are similar to the methods that form part of the quantile view group, since these methods are also based on the largest order statistics. Estimation methods that form part of the probability view group are based on the following general condition:

$$\frac{\bar{F}(t + yb(t))}{\bar{F}(t)} \to (1 + \gamma y)^{-1/\gamma} \text{ for any } y > 0 \text{ as } t \uparrow x_+,$$

for some auxiliary function b, where x_+ is the finite upper limit of the distribution and $\bar{F}$ is the survival function (See Beirlant et al. 2004 for derivations).

Parameter estimation methods that form part of the probability view group include the maximum likelihood method, the probability-weighted moments method and the elemental percentile method (Beirlant et al. 2004: 147-154).

3.3 First and second order estimation of the EVI

As mentioned previously, the focus in this thesis will be on heavy-tailed distributions and, in particular, obtaining an estimate of the EVI (α) of a specific heavy-tailed distribution. The goal of estimating the EVI of a heavy-tailed distribution is to make inferences regarding a tail quantity far from the centre of the distribution.

For a sample of size n, this implies that only the k upper order statistics (excesses) can be used to estimate the EVI.

The following subsections serve as an introduction to obtaining the optimal number of excesses. These subsections specifically discuss certain definitions and concepts that motivate the use of the Pareto distribution.

3.3.1 First order regular variation

The definition of regularly varying tail function is given below (Peng and Qi, 2004: 306; Geluk, de Haan, Resnick and Starica, 1997: 139).

Definition 3.3.1: Regularly varying tail function

Suppose $X_1, X_2, \ldots X_n$ are independent, identically distributed random variables with common distribution function F on $[0, \infty)$ and tail function $\bar{F} = 1 - F$.

Then $\bar{F}$ is said to be regularly varying if

$$\lim_{t \to \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\alpha}$$

with $\alpha > 0$, for all $x > 0$.

The constant α is called the tail index or the first order regular variation parameter (De Haan and Ferreira, 2006).

The right-hand side of the equation can be recognised as the tail function of the Pareto distribution. The positive tail index $(\alpha > 0)$ implies a heavy-tailed distribution (EVI $= \gamma = \frac{1}{\alpha} > 0$).

Berning (2010: 18) summarises equivalent definitions of heavy-tailed distributions with distribution function F and tail function $\bar{F} = 1 - F$ as follows:

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

Berning (2010: 25) shows that, for all heavy-tailed distributions, the distribution of the multiplicative excesses over a threshold t tends to a Pareto distribution as t becomes large. Therefore, it follows that the multiplicative excesses are approximately Pareto distributed if the threshold is large.

In order to estimate the parameter of interest γ, maximum likelihood estimation is used to fit the known limiting distribution (Pareto distribution) to the multiplicative excesses. This leads to the well-known estimator of the EVI, namely the Hill estimator. The method was developed by Hill (1975) and is formally defined in the next subsection.

3.3.2 The Hill estimator

The definition of the Hill estimator as given by Beirlant et al. (2004: 101-103) is given below.

Definition 3.3.2: The Hill estimator

Let $X_{(1,n)} \leq X_{(2,n)} \leq \cdots \leq X_{(n,n)}$ denote the ascending order statistics of a series of independent, identically distributed random variables. Also, let k denote the number of excesses.

For positive EVI and k excesses, the Hill estimator of the EVI is defined as

$$H_{k,n} = \frac{1}{k}\sum_{i=1}^{k} logX_{(n-i+1,n)} - logX_{(n-k,n)}.$$

An important consideration is the choice of threshold. A number of methods can be used to determine the optimal threshold for the Hill estimator (See Beirlant et al, 2004: 123–129). Two of these methods will be considered in this study, namely the method of Guillou and Hall (2001) and the method of Drees and Kaufmann (1998). The procedures of these two methods are described in the following subsections.

3.3.2.1 Method of Guillou and Hall

Guillou and Hall (2001) propose a method that can be used to determine the optimal threshold for the Hill estimator. The steps to calculate the optimal threshold using the above mentioned method can be summarised as follows:

1. For an ordered set of observations $x_{(1,n)} \leq x_{(2,n)} \leq \cdots \leq x_{(n,n)}$ and a given value of k, calculate the Hill estimate $H_{k,n}$ (as defined in section 3.3.2).
2. $y_i = ilog\left(\frac{x_{(n-i+1,n)}}{x_{(n-i,n)}}\right)$ for $i = 1,2,\ldots,k$
3. $w_i = k - 2i + 1$ for $i = 1,2,\ldots,k$
4. $u_k = \sum_{i=1}^{k} w_i y_i$
5. $t_k = \sqrt{\frac{3}{k^3}}\left(\frac{u_k}{H_{k,n}}\right)$
6. $m = \lfloor k/2 \rfloor$, which is the integer part of $k/2$
7. $q_k = \sqrt{\frac{1}{2m+1}\sum_{j=-m}^{m} t_{k+j}^2}$

The optimal choice of the threshold is defined as k_{opt} and is the smallest integer k that meets the requirement $q_r > c_{crit}$ for all $r > k$. The values of the critical value c_{crit} that will be considered in this thesis are those recommended by the authors, namely 1.25 and 1.5.

In the case where $q_r \leq c_{crit}$ for all $k = 2,3,\ldots,k_{max}$, the optimal choice of k is equal to k_{max}, where k_{max} is defined as the maximum value of k for which $\sum_{j=-m}^{m} t_{k+j}^2$ can be calculated. In the case where $q_r > c_{crit}$ for all $k = 2,3,\ldots,k_{max}$, the procedure fails.

Berning (2010: 134) conducts an extensive simulation study, and the procedure never failed for $c_{crit} = 1.5$. In the case where the procedure failed for $c_{crit} = 1.25$, the procedure was repeated for that sample using $c_{crit} = 1.5$. The same principle will also be applied in this thesis.

3.3.2.2 Method of Drees and Kaufmann

Drees and Kaufmann (1998) develop a method that can be used to determine the optimal threshold for the Hill estimator.

The steps to calculate the optimal threshold can be summarised as follows:

1. For n observations, obtain an initial estimate of the EVI, calculated as $\hat{\gamma}_0 = H_{2\sqrt{n},n}$. Here $H_{2\sqrt{n},n}$ is the Hill estimate where $k = 2\sqrt{n}$.
2. For $r_n = 2.5\hat{\gamma}_0 n^{0.25}$, compute the minimum value of k for which there is an i such that $\left|\sqrt{i}(H_{i,n} - H_{k,n})\right| > r_n$, denoted by $\hat{k}_n(r_n)$. Here $k = 1,2, \dots, (n-1)$ and $i = 1,2, \dots, k$.
3. If $\hat{k}_n(r_n)$ does not exist, replace r_n by $0.9r_n$. Continue to do so until $\hat{k}_n(r_n)$ is well-defined.
4. In the same way, compute $\hat{k}_n(r_n^{0.7})$.
5. The optimal choice of k is given by $\hat{k}_{opt} = \frac{1}{3}\left(\frac{\hat{k}_n(r_n^{0.7})}{\left(\hat{k}_n(r_n^{0.7})\right)^{0.7}}\right)^{\frac{10}{3}} (2\hat{\gamma}_0)^{\frac{1}{3}}$.

The procedure will fail if $\hat{k}_n(r_n)$ or $\hat{k}_n(r_n^{0.7})$ does not exist. The procedure will also fail if $\hat{k}_{opt} < 2$ or if $\hat{k}_{opt} > n - 1$. In these cases, the initial estimate of the EVI, namely $\hat{\gamma}_0 = H_{2\sqrt{n},n}$, will be used.

According to Beirlant et al. (2004: 113), in most cases the Hill estimator overestimates the population value of γ. The main disadvantage of the Hill estimator is therefore that it is associated with a large bias term. Recall that the Hill estimator is the maximum likelihood estimator of the EVI. The Pareto distribution has the EVI as a parameter and is used as the limiting distribution of the multiplicative excesses when assuming first order regular variation.

In order to address the problem associated with the Hill estimator, the bias can be reduced by fitting a distribution to the multiplicative excesses which assumes second order regular variation of the tail function. Second order regular variation is formally defined in the next subsection.

3.3.3 Second order regular variation

The definition of second order regularly varying tail function is given below (De Haan and Stadtmüller, 1996: 383-384; Geluk et al. 1997: 139-140).

Definition 3.3.3: Second order regularly varying tail function

A regularly varying tail function $\bar{F}$ is second order regularly varying with first order index $-\frac{1}{\gamma}$, $\gamma > 0$, and second order index $\frac{\rho}{\gamma}$, $\rho < 0$, if a function $A(t)$ exists which tends to 0 and is eventually of constant sign as $t \to \infty$, such that

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

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

In the definition of second order regular variation, γ is the first order parameter (EVI) and ρ is called the second order parameter.

Table 3.3.1 is an expansion of table 2.3.1 and includes the first and second order parameters (Berning, 2010: 20):

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

Table 3.3.1: Pareto type distributions

All Pareto type distributions mentioned in this thesis are first and second order regularly varying. In fact, it is safe to assume that all Pareto type distributions encountered in practice will be first and second order regular varying.

Second order regular variation is a generalisation of first order regular variation. In order to determine the limiting distribution of the multiplicative excesses when assuming second order regular variation, consider the same setting as in the case when first order regular variation is assumed:

Let $X_1, X_2, \dots X_n$ be independent, identically distributed random variables with common heavy-tailed distribution function F on $[0, \infty)$. Since only multiplicative excesses are considered, denote Z as the multiplicative excess X/t given $X \geq t$, where t indicates a positive threshold. The tail function of Z is given by $\bar{F} = 1 - F$.

From the definition of second order regular variation, the following holds when $\rho < 0$:

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

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

Using this result and the fact that $\bar{F}(z) = \bar{F}(tz)/\bar{F}(t)$, it follows that

$$\bar{F}(z) \approx z^{-\frac{1}{\gamma}} + \frac{\gamma d}{\rho} z^{-\frac{1}{\gamma}} \left(z^{\frac{\rho}{\gamma}} - 1 \right) A(t)$$

for large t.

Therefore, as t becomes large, the survival function of the multiplicative excesses $\left(\bar{F}(z)\right)$ tends to

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

where $c = \gamma d A(t)/\rho$ (Berning, 2010: 29).

This well-known result is the survival function of the perturbed Pareto distribution, which is formally defined in the next subsection.

3.3.4 Perturbed Pareto distribution

The perturbed Pareto distribution (PPD) is defined below (Beirlant et al. 2004: 188).

Definition 3.3.4: Perturbed Pareto distribution

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

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

$$f(x) = \gamma^{-1}x^{-\frac{1}{\gamma}-1}\left(1 - c + c(1-\rho)x^{\frac{\rho}{\gamma}}\right)$$

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

The above definition indicates that the EVI is also a parameter of the perturbed Pareto distribution. If maximum likelihood estimation is used to estimate the EVI, the equation will result in a second order generalisation of the Hill estimate. This second order generalisation can be obtained by following the procedure proposed by Berning (2010: 31), namely to fit the perturbed Pareto distribution to the multiplicative excesses of a data set, instead of the Pareto distribution.

The advantage of using the perturbed Pareto distribution is that these bias reduced estimates significantly improve the estimate of the EVI.

Another advantage of using the perturbed Pareto distribution as limiting distribution, is that the estimate is less critically dependant on the choice of threshold than the Hill estimate. Berning (2010) illustrated that for a wide range of threshold values, the estimates of the EVI stay relatively close to the true value of the parameter when using the perturbed Pareto distribution.

The difficulty of estimating the EVI using the perturbed Pareto distribution is that an accurate estimate of the second order parameter ρ must be obtained. The value of the second order parameter can be determined by considering one of the following three options:

- The value of ρ can be fixed. Beirlant et al. (2004) suggests using $\rho = -1$;
- ρ can be estimated simultaneously with the other parameters γ and c;
- ρ can be estimated externally.

In this thesis the second order parameter will be estimated externally. The motivation and reasoning for this will be discussed in section 3.3.5.

The estimates of the two remaining parameters of the perturbed Pareto distribution, namely c and γ, will be obtained by means of both maximum likelihood estimation and Bayesian estimation. The procedure of obtaining these two parameter estimates is discussed in detail in section 4.7.

3.3.5 External estimation of the second order parameter

It is crucial to obtain an accurate estimate of ρ in order to calculate a second order reduced bias estimator of the EVI. In the past ρ was often set equal to -1. This was not only for the sake of simplicity, but also due to the fact that no satisfactory method of estimating ρ existed. Research developments overcame the difficulties in second order parameter estimation (Feuerverger and Hall, 1999; Gomes, Martins and Neves, 2000).

Gomes and Martins (2002) investigate the external estimation of the second order parameter. They consider four different estimates of ρ and recommend the following estimate for practical application. This estimate, denoted by $\hat{\rho}_1$, is a special case of a class of estimates proposed by Fraga Alves, Gomes and De Haan (2003) and is defined as follows:

$$\hat{\rho}_1 := -\left|\frac{3\left(T_n^{(0)}(k_1)-1\right)}{T_n^{(0)}(k_1)-3}\right|$$

where $k_1 = min(n-1, [2n/\log\log n])$,

$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\log\frac{X_{n-i+1,n}}{X_{n-k,n}}\right)^j$, $j \geq 1$ and

$$T_n^{(\tau)}(k) := \begin{cases} \dfrac{\left(M_n^{(1)}(k)\right)^{\tau}-\left(M_n^{(1)}(k)/2\right)^{\tau/2}}{\left(M_n^{(2)}(k)/2\right)^{\tau/2}-\left(M_n^{(3)}(k)/6\right)^{\tau/3}} & if\ \tau>0 \\[2em] \dfrac{\log\left(M_n^{(1)}(k)\right)-\frac{1}{2}\log\left(M_n^{(2)}(k)/2\right)}{\frac{1}{2}\log\left(M_n^{(2)}(k)/2\right)-\frac{1}{3}\log\left(M_n^{(3)}(k)/6\right)} & if\ \tau=0. \end{cases}$$

Another estimate of the second order parameter that will be considered in this thesis is one that was also proposed by Gomes and Martins (2001). The steps for obtaining this estimate can be summarised as follows:

1. For an ordered set of n positive observations $x_{(1,n)} \leq x_{(2,n)} \leq \cdots \leq x_{(n,n)}$, assumed to be from a second order regularly varying distribution, define the following:

   1.1. For given choices of α and k:

   Define $\hat{\gamma}_n^{(\alpha)}(k) = \frac{M_n^{(\alpha)}(k)}{\Gamma(\alpha+1)(H_{k,n})^{\alpha-1}}$, where $H_{k,n}$ denotes the Hill estimator, which is based on k excesses (as defined in section 3.3.2).

   1.2. For given choices of α, i_n and j_n:

   Let $\chi(\alpha, i_n, j_n)$ be the sample median of $\hat{\gamma}_n^{(\alpha)}(k)$, where $k = i_n, \dots, j_n$.

   1.3. For given choices of α, i_n and j_n:

   Let $\Sigma(\alpha, i_n, j_n) = \sum_{k=i_n}^{j_n} \left(\hat{\gamma}_n^{(\alpha)}(k) - \chi(\alpha, i_n, j_n)\right)^2$.

2. Obtain an estimate of α, defined as $\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \Sigma(\alpha, i_n, j_n)$.

3. Estimate ρ as the solution of $(1-\hat{\rho})^{\hat{\alpha}-1}\left(1+\hat{\rho}(\hat{\alpha}-2)\right) = 1$. Denote this estimate of ρ by $\hat{\rho}_2$.

The restriction $\hat{\alpha} \geq 2$ applies. Also, the value of $\hat{\alpha} = 2.09272$ corresponds to $\hat{\rho} = -10$ and the value of $\hat{\alpha} = 15.02746$ corresponds to $\hat{\rho} = -0.01$. Computationally the value of $\hat{\rho}$ is restricted to $-10 \leq \hat{\rho} \leq -0.01$. In other words, set $\hat{\rho} = -10$ if $2 \leq \hat{\alpha} \leq 2.09272$ and set $\hat{\rho} = -0.01$ if $\hat{\alpha} \geq 15.02746$.

Gomes and Martins (2001: 178) state that the method is quite robust regarding the choice of the values of i_n and j_n.

In the simulation study that will follow in chapter 5, the second order parameter will be externally estimated by using these two estimators.

3.3.6 Threshold selection when fitting the PPD

Berning (2013: 25) showed that a reasonable choice of a fixed (non-adaptive) threshold for the second order estimator of the EVI is $k = 2n^{2/3}$. Therefore in this thesis the value of k is fixed at $k_0 = 2n^{2/3}$ when fitting the perturbed Pareto distribution to the multiplicative excesses.

3.4 Other estimators of the EVI

The main goal of this thesis is to obtain an empirical Bayes estimate of the EVI. Standard empirical Bayes methodology will be discussed in chapter 6 and the empirical Bayes estimate of the EVI will be derived in detail in section 6.5.

In this section a brief description will be given of alternative approaches to the estimation of the EVI. These alternative approaches do not entail fitting a distribution to the multiplicative excesses. They are only mentioned for the sake of completeness and will not be elaborated on or be included in the simulation studies in this thesis.

The following alternative estimators of the EVI will be discussed:

- Zipf estimator;
- Moments estimator;
- Pickands estimator;
- Kernel-Type estimators;
- Estimator based on the exponential regression model.

3.4.1 Zipf estimator

Let $X_{(1,n)} \leq X_{(2,n)} \leq \cdots \leq X_{(n,n)}$ denote the order statistics of a finite sample $X_1, X_2, \ldots, X_n$. For a given value of k, Zipf (1949) defined an estimator of the EVI as follows:

$$\hat{\gamma}_{k,n}^{Z} = \frac{\frac{1}{k}\sum_{j=1}^{k}\left(log\frac{k+1}{j+1} - \frac{1}{k}\sum_{i=1}^{k} log\frac{k+1}{i+1}\right) logUH_{j,n}}{\frac{1}{k}\sum_{j=1}^{k} log^2\frac{k+1}{j+1} - \left(\frac{1}{k}\sum_{j=1}^{k} log\frac{k+1}{j+1}\right)^2}$$

where $UH_{j,n} = X_{(n-i,n)}H_{i,n}$.

In the same way as before, $H_{i,n}$ denotes the Hill estimator based on the i largest order statistics from the sample of n observations, and k denotes the number of largest order statistics which is used in the calculation of the estimate.

3.4.2 Moments estimator

Dekkers et al. (1989: 1833-1834) derived the moment estimator of the EVI. This estimator is a direct generalisation of the Hill estimator and is defined as follows:

Let $X_{(1,n)} \leq X_{(2,n)} \leq \cdots \leq X_{(n,n)}$ denote the order statistics. Let

$$M_n^{(1)} = \frac{1}{k}\sum_{i=0}^{k-1} logX_{(n-i,n)} - logX_{(n-k,n)} \qquad (k < n).$$

Note that $M_n^{(1)}$ is the Hill estimator if it is assumed that $\gamma > 0$.

Let the second moment be defined as

$$M_n^{(2)} = \frac{1}{k}\sum_{i=0}^{k-1}\left(logX_{(n-i,n)} - logX_{(n-k,n)}\right)^2.$$

Then the estimate of the EVI is given by

$$\hat{\gamma}_n = M_n^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\right)^{-1}.$$

The moment estimator is consistent for all $\gamma \in \mathbb{R}$.

3.4.3 Pickands estimator

For $X_{(1,n)} \leq X_{(2,n)} \leq \cdots \leq X_{(n,n)}$ and a given value of k, Pickands (1975) defined the following estimator of the EVI:

$$\hat{\gamma}_{k,n}^{Pi} = \frac{1}{\log(2)} log\left(\frac{X_{(n-k+1,n)} - X_{(n-2k+1,n)}}{X_{(n-2k+1,n)} - X_{(n-4k+1,n)}}\right)$$

where $i \leq n/4$.

The advantage of the Pickands estimator is that $\hat{\gamma}^{Pi}_{m_n,n}$ is consistent for all $\gamma \epsilon \mathbb{R}$ and any intermediate sequence $m_n \to \infty$ and $m_n/n \to 0$. Also the estimator is easy to compute and is not substantially influenced by scale transformations.

The disadvantage of the Pickands estimator is that the estimator is associated with large asymptotic variance. Drees (1995: 2060) addressed this problem by deriving refined Pickands estimators of the EVI. These refined estimators can be seen as a mixture of Pickands estimators and is defined as

$$\hat{\gamma}^{Pi}_{n,v} = \sum_{k=1}^{m_n} c_{nk}\hat{\gamma}^{Pi}_{k,n}.$$

In the above equation m_n is an intermediate sequence and c_{nk}, $1 \leq k \leq m_n$, refer to the scores obtained by a probability measure v, specifically $c_{nk} = v((k-1)/m_n, k/m_n]$.

The refined Pickands estimator also has the advantage of being more robust if an inappropriate number of upper order statistics are used in estimating the value of the EVI (Drees, 1995).

3.4.4 Kernel-type estimators

Csörgó, Deheuvels and Mason (1985) derived two nonparametric estimates of the first order regular variation parameter, based on a nonnegative non-increasing kernel.

Let $X_1, X_2, \ldots X_n$ be independent, identically distributed random variables with common distribution function F on $[0, \infty)$ and assume that the tail function $\bar{F} = 1 - F$ is regularly varying, implying:

$$\lim_{t\to\infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\alpha}$$

for any $x > 0$.

Considering the order statistics $X_{(1,n)} \leq X_{(2,n)} \leq \cdots \leq X_{(n,n)}$, the first estimate of the tail index α is defined as

$$\alpha_{n,\lambda} = \left(\int_0^{1/\lambda} \{logQ_n(1 - v\lambda)\} d\{vK(v)\}\right)^{-1} \left(\int_0^{1/\lambda} K(v)dv\right).$$

The above estimate is based on a chosen bandwidth parameter $\lambda = \lambda_n > 0$.

Also $\{K(u), u > 0\}$ is a nonnegative, non-increasing kernel that satisfies the condition

$$\int_0^{\infty} K(u)du = 1$$

and $\{Q_n(s), 0 < s \leq 1\}$ is the empirical quantile function, which is defined for $n \geq 1$ by $Q_n(s) = X_{(k,n)}$ if $(k-1)/n < s \leq k/n$ and $1 \leq k \leq n$.

If the scale term in above estimate is corrected for finite samples, the estimate becomes

$$\tilde{\alpha}_{n,\lambda} = \left(\sum_{j=1}^{n} \frac{j}{n\lambda} K\left(\frac{j}{n\lambda}\right)\{logX_{(n-j+1,n)} - logX_{(n-j,n)}\}\right)^{-1} \left(\sum_{j=1}^{n} \frac{j}{n\lambda} K\left(\frac{j}{n\lambda}\right)\right).$$

Note that the Hill estimator is a special case of the above estimator, corresponding to $K(u) = 1_{\{0<u<1\}}$ and $\lambda = k/n$.

Csörgó et al. (1985) state that the optimal kernel and bandwidth parameter may depend on the parameters of the underlying distribution, including the parameter of interest α. They suggest replacing the parameters of the distribution by preliminary estimates.

3.4.5 Estimator based on the exponential regression model

Matthys and Beirlant (2003) show that, for a given value of k and considering the order statistics $X_{(1,n)} \leq X_{(2,n)} \leq \cdots \leq X_{(n,n)}$, the log-ratio of spacings of the order statistics can be approximated by

$$ilog \frac{X_{(n-i+1,n)} - X_{(n-k,n)}}{X_{(n-i,n)} - X_{(n-k,n)}} \overset{D}{\approx} \frac{\gamma}{1 - \left(\frac{i}{k+1}\right)^{\gamma}} E_i$$

for $i = 1,2, \ldots, k-1$.

In the above approximation γ is the EVI and E_i, $i = 1,2, \ldots, (k-1)$ denote independent standard exponential random variables. Using the maximum likelihood method, the parameter γ can then be estimated.

The advantage of the above estimator is that the estimated EVI is consistent for all $\gamma \in \mathbb{R}$. Also the estimator is invariant with respect to a shift or a rescaling of the data.

3.5 Conclusion

The chapter focused on different methods that can be used to estimate the parameter of interest, namely the EVI. The first section discussed the three classes in which standard estimation methods can be categorised into.

In section 3.3 it was stated that the focus will be in obtaining an estimate of the EVI parameter of the generalised Pareto distribution. When assuming first order regular variation, the limiting distribution of the multiplicative excesses is the Pareto distribution. Since the EVI is a parameter of the Pareto distribution, maximum likelihood estimation can be used to estimate the EVI. The resulting estimator is known as the Hill estimator.

The importance of determining the optimal threshold was stressed and two methods where discussed that can be used to obtain the optimal threshold for the Hill estimator.

In section 3.3 it was shown that when assuming second order regular variation, the PPD is the resulting limiting distribution of the multiplicative excesses. Since the EVI is a parameter of the PPD, maximum likelihood estimation can also be used to estimate the EVI. The resulting estimator can be seen as a generalisation of the Hill estimator.

In the final section, five alternative estimators of the EVI were discussed.

CHAPTER 4

Overview of Bayesian theory

4.1 Introduction

Statistics uses two major paradigms, namely the frequentist (classical) paradigm and the Bayesian paradigm. The aim of this chapter is to introduce the latter and examine how methods in this paradigm can be used to estimate the EVI.

The chapter begins with the historical development of Bayesian theory. In section 4.3 the basic paradigm of Bayesian theory is discussed, which centres around Bayes' Theorem.

Section 4.4 discusses how inferences can be made by using Bayesian methodology. In section 4.5 the different types of prior distributions are mentioned and also different approaches that can be followed in order to obtain a prior distribution.

Section 4.6 discusses how Bayesian methodology can be used in an extreme value context and in section 4.7 these methods are used to obtain Bayesian estimates of the PPD parameters.

In section 4.8 Gibbs sampling is introduced, which is a Markov Chain Monte Carlo method that can be used to simulate values from the desired posterior distribution. The Gibbs sampler algorithm is also presented.

Different application areas of Bayesian methods are discussed in section 4.9 and in the last section some of the difficulties associated with the Bayesian approach are mentioned.

4.2 Historical development of Bayesian theory

Bayesian methods are primarily based on Bayes' Theorem, which was formulated by Thomas Bayes and first published in 1763. His ideas were further developed and implemented by well-known statisticians like Laplace and Gauss, but Bayesian statistics was largely ignored until the second half of the 20th century. In the 1950's statisticians like Savage and Kiefer implemented Bayesian methods to provide solutions to problems that could not be solved by using the frequentist approach (Carlin and Louis, 1996).

The recent development of computational software contributed to the ascendance of the Bayesian approach. Recent computing advances allowed the constraints on prior assumptions and models to be relaxed. This has led to the implementation of Bayesian statistics in numerous fields, including medicine, finance, auditing and monitoring environmental pollution (O'Hagan, 2003: 31).

4.3 Basic results of Bayesian theory

In simple terms the Bayesian methodology can be described as a probability distribution that reflects the current state of knowledge. When new data becomes available, the new data is considered and the probability distribution is updated. In this paradigm, the process of learning from the data is systematically implemented by making use of Bayes' Theorem. The available prior knowledge is therefore combined with the information provided by the data to produce the required posterior distribution (Bernardo, 2003).

Let $\boldsymbol{x} = (x_1, \dots, x_n)$ represent the data of a random variable X which follows a distribution with density function $f(\boldsymbol{x}|\boldsymbol{\theta})$. Also let $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$ represent a vector of unknown parameters in a statistical model. The assumption made when considering Bayesian inference is that there exists *a priori* knowledge regarding the distribution of the parameter vector $\boldsymbol{\theta}$. This is formalised by specifying a prior density function $\pi(\boldsymbol{\theta}|\boldsymbol{\eta})$. Given this framework, Bayes' Theorem can be used for statistical inference.

Bayes' Theorem is given in Carlin and Louis (1996) as follows:

Definition 4.3.1: Bayes' Theorem

Let $f(\boldsymbol{x}|\boldsymbol{\theta})$ *denote the likelihood function of* $\boldsymbol{\theta}$ *and* $\pi(\boldsymbol{\theta}|\boldsymbol{\eta})$ *denote a prior distribution, where* $\boldsymbol{\eta}$ *is a vector of hyperparameters. According to Bayes' Theorem, the posterior distribution of* $\boldsymbol{\theta}$ *is given by*

$$\pi(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{\eta}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\eta})}{\int f(\boldsymbol{x}|u)\pi(u|\boldsymbol{\eta})du}.$$

In some notation $\boldsymbol{\eta}$ is suppressed in the above expression. The posterior distribution can then be expressed as

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) \propto f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$$

which simply implies that the posterior distribution is proportional to the product of the likelihood and the prior. The proportionality follows from the fact that the numerator is a constant obtained by simply integrating out the parameter vector.

The marginal posterior distribution of parameter θ_i in the parameter vector $\boldsymbol{\theta}$ is obtained by integrating over all the remaining parameters. It can be difficult to obtain these integrals analytically and Markov Chain Monte Carlo methods are usually applied to estimate the marginal posterior distribution.

Geyer (1992: 473) defines Markov Chain Monte Carlo (MCMC) as a general method that can be used to simulate stochastic processes. The basic idea is that, if it is not possible to simulate independent realisations of a certain stochastic process, dependent realisations can be simulated. When such realisations are generated, the stationary distribution of the obtained Markov Chain is the distribution of interest. Since dependent realisations are generated, a larger sample is needed. Due to the advances in modern computer power it will always be possible to obtain such a large sample.

The Monte Carlo method that will be used in this study is Gibbs sampling, which will formally be introduced in section 4.8. This method will be used to obtain parameter estimates of the perturbed Pareto distribution (PPD).

4.4 Bayesian inference

Let $\boldsymbol{x}$ be the available data and $\boldsymbol{\theta}$ be the quantity of interest. A point estimator of $\boldsymbol{\theta}$ is some function of the data $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\boldsymbol{x})$ that draws inferences about the population, by using a single value or point (Keller & Warrack, 2000: 286).

Considering the estimator $\widehat{\boldsymbol{\theta}}$ and the loss function $l(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})$, the expected posterior loss is given as:

$$l(\widehat{\boldsymbol{\theta}}|\boldsymbol{x}) = \int l(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{x}) d\theta.$$

The loss function quantifies the errors made when estimating a parameter. For instance, $l(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ can be the squared error loss function, namely $l(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^2$. The corresponding Bayes estimator $\boldsymbol{\theta}^*$ is the function $\boldsymbol{\theta}^*(\boldsymbol{x})$ which minimises the expected posterior loss.

For any given model and data, the Bayes estimator depends on the chosen loss function.

The loss function is context specific and should be chosen by considering how the estimate will be used. A number of conventional loss functions have been suggested in the literature for situations where the estimator is not explicitly stated. These loss functions produce estimates which are simply measures of location for the posterior distribution. If square error loss is assumed, the Bayesian estimate of θ_i is the expected value of its marginal posterior distribution $f_i(\boldsymbol{\theta}|\boldsymbol{x})$, which is

$$\hat{\theta}_i = \int \boldsymbol{\theta} f_i(\boldsymbol{\theta}|\boldsymbol{x}) d\theta.$$

Similarly if the loss function is a zero-one function, the Bayesian estimate of θ_i is the posterior mode. Assuming absolute error loss, the Bayesian estimate of θ_i is the posterior median.

Sometimes it can be meaningless to obtain point estimates, except in a specific decision context. It is often more appropriate to describe the posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{x})$ in terms of regions of given probability (Bernardo, 2003: 18-19).

Definition 4.4.1: Posterior credible region

A $100(1-\alpha)\%$ *credible region for* $\boldsymbol{\theta}$ *is a subset* $R_{1-\alpha}$ *of* Θ *such that*

$$1-\alpha \leq P(R_{1-\alpha}|\boldsymbol{x}) = \int_{R_{1-\alpha}} f(\boldsymbol{\theta}|\boldsymbol{x})d\theta$$

(Bernardo, 2003: 21).

This implies that, given the data $\boldsymbol{x}$, the probability that the true value of $\boldsymbol{\theta}$ lies in the subset $R_{1-\alpha}$ is at least $(1-\alpha)$.

If the region also has the characteristic that all points in the region have larger probability density than all points outside, the region is called the highest probability density (HPD) region. Stated differently, for a specified probability $1-\alpha$, the HPD region is the shortest interval containing at least $100(1-\alpha)\%$ of the observations (Bernardo, 2003).

4.5 Choice of prior

The prior distribution is crucial in Bayesian statistics. Two main categories of priors exist. The first is the subjective prior. In this case the choice of prior used is based on the personal judgement of the analyst, before observing the data. If this type of subjective prior is used it is important to base the choice of prior on external evidence obtained from previous studies or on the opinions of experts in the field.

The second category consists of objective priors. Since subjectivity can vary from one expert to another, prior distributions are constructed which minimise the input of the analyst. Objective prior distributions are derived from the assumed probability function of the data.

Carlin and Louis (1996) mention three approaches that can be followed in order to obtain a prior. These approaches will now be discussed briefly.

4.5.1 Elicited priors

Elicited priors involve considering all possible values of $\boldsymbol{\theta}$ that are manageable and then to assign probability weights to each of these values. These weights should add up to one, where each weight reflects the prior beliefs of the researcher. The approach works well if the parameter values are discrete, but prove more difficult if $\boldsymbol{\theta}$ is continuous. Alternatively the assumption can be made that the prior of $\boldsymbol{\theta}$ is part of some parametric distributional family $\pi(\boldsymbol{\theta}|\boldsymbol{\eta})$. Then $\boldsymbol{\eta}$ can be chosen in such a way that it reflects the researcher's prior beliefs.

4.5.2 Conjugate priors

In a Bayesian context, if the prior distribution is of the same family as the posterior distribution, the prior is called a conjugate prior. Priors of this form are usually selected for mathematical convenience, but are usually flexible enough to give an indication of the prior knowledge.

4.5.3 Non-informative priors

If no prior information exists on the parameter vector $\boldsymbol{\theta}$ and a parametric prior is difficult to justify, non-informative priors can be used. Essentially this implies that all information contained in the posterior distribution is obtained from the data. This will result in an objective posterior distribution. According to Kass & Wasserman (1996), non-informative priors can usually be used as a starting point in any analysis.

4.5.4 Other objective priors

There are numerous other methods that can also be used to obtain an appropriate prior. Other proposals are for instance Jeffrey's prior, the reference prior or Zellner's maximal data information (MDI) prior.

The MDI prior provides maximal average data information on the vector of parameters $\boldsymbol{\theta}$ and is defined below (Zellner, 1971; Beirlant et al., 2004).

Definition 4.5.1: MDI prior

For a random variable X with density function $f(\boldsymbol{x}|\boldsymbol{\theta})$, the MDI prior of the parameter vector $\boldsymbol{\theta}$ is defined as

$$\pi(\boldsymbol{\theta}) \propto \exp\big(E(\log f(\boldsymbol{x}|\boldsymbol{\theta}))\big).$$

The MDI prior will be used in this thesis. The derivations of the prior for the Pareto distribution and the PPD will be discussed in section 4.7.

4.6 Bayesian methodology and extreme value theory

Ashour and El-Adl (1980) conducted a simulation study focussing on the distribution of the minima. They conclude that the Bayesian estimators are more efficient compared to maximum likelihood estimators, but that the Bayesian estimators are also more biased.

One of the difficulties that arise when applying Bayesian methodology to EVT is the formulation of the prior information. Since only extreme events are considered, it may be difficult to justify the choice of prior distribution (Kotz and Nadarajah, 2000).

Engelund and Rackwithz (1992) showed that the GEV distribution also does not allow conjugate priors, with the exception of GEV distributions that have only one parameter. Literature suggests specifying priors in terms of extreme quantiles for the underlying process (Coles and Tawn, 1996).

Another difficulty involves computing the posterior distribution. Complex models make it difficult to compute the integral analytically. However, simulation-based techniques like Markov Chain Monte Carlo (MCMC) methods can be used to obtain the posterior distribution (Coles, 2001: 171-173).

Despite these difficulties, there are numerous advantages of using Bayesian methods in an extreme value context.

Beirlant et al. (2004) discussed several advantages of using Bayesian methods when conducting an extreme value analysis. The first advantage is that additional knowledge can be taken into account, usually in the form of certain constraints. This can prove valuable since the available information is usually limited in an extreme value context.

The second advantage is that it is easy to make predictions using Bayesian methods. Let z be some future observation with probability density function $f(z|\boldsymbol{\theta})$. Also let $f(\boldsymbol{\theta}|\boldsymbol{x})$ denote the posterior distribution based on the observed data $\boldsymbol{x}$. Then the predictive density of z given $\boldsymbol{x}$ is

$$f(z|x) = \int_{\Theta} f(z|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{x}) d\theta \,.$$

The $f(\boldsymbol{\theta}|\boldsymbol{x})$ term in the predictive density indicates the uncertainty in the model and the $f(z|\boldsymbol{\theta})$ term indicates the uncertainty caused by variability in the future observations.

Since the goal of an extreme value analysis is usually to predict the occurrence of some extreme event, Bayesian methods can easily be implemented here.

Another advantage of Bayesian methods is that they are not dependent on regularity assumptions, as opposed to other methods like maximum likelihood. See for instance Beirlant et. al (2004: 429-430).

The following section will discuss how to obtain Bayesian estimates of the PPD parameters.

4.7 Bayesian estimation of the PPD parameters

Consider again the definition of the PPD given in section 3.3.4:

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

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

$$f(x) = \gamma^{-1} x^{-\frac{1}{\gamma}-1}\left(1 - c + c(1-\rho)x^{\frac{\rho}{\gamma}}\right)$$

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

The objective prior that will be used in obtaining the Bayesian estimates of the PPD parameters is Zellner's MDI prior. Beirlant et al. (2004: 432) defines the MDI prior as:

$$\pi(\gamma) \propto \exp\big(E(\log f(\boldsymbol{x}|\gamma))\big).$$

The MDI prior of the PPD cannot be obtained in closed form, but since the PPD is a second order generalisation of the Pareto distribution, the MDI of the Pareto distribution can be used as a substitute for the MDI of the PPD.

Consider the definition of the Pareto distribution:

$$\bar{F}(x) = x^{-\frac{1}{\gamma}} \text{ and } f(x) = \gamma^{-1}x^{-\frac{1}{\gamma}-1}.$$

By using the definition of the MDI prior, Beirlant et al. (2004: 438) gives the MDI prior of the Pareto distribution as

$$\pi(\gamma) \propto \frac{1}{\gamma}e^{-\gamma}.$$

The above prior of the Pareto distribution will be used as an approximation of the MDI prior of the PPD, denoted as

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

Assuming that the observations $z_1, z_2, \dots, z_k$ are from a PPD, the posterior distribution of the PPD is given by:

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

In order to obtain estimates of the parameters of the posterior distribution, Gibbs sampling is used.

4.8 Gibbs sampling

Casella and George (1992) define Gibbs sampling as a technique that generates random variables indirectly from a marginal distribution. This implies that it is not necessary to calculate or approximate the probability density function.

The PPD consists of three unknown parameters. In order to sample from a posterior distribution $f(a, b, c|\boldsymbol{x})$, Kroese, Taimre and Botev (2011: 675) gives the following algorithm, based on the Gibbs sampler:

1. Obtain initial values of a, b and c. Now iterate the following steps:
2. Draw a from $f(a|b, c, \boldsymbol{x})$.
3. Draw b from $f(b|a, c, \boldsymbol{x})$.
4. Draw c from $f(c|a, b, \boldsymbol{x})$.

In this way a dependent sample $\{(a_t, b_t, c_t)\}$ can be obtained from the posterior distribution $f(a, b, c|\boldsymbol{x})$.

The algorithm is adjusted in order to obtain Bayesian estimates of the PPD parameters. The adjusted algorithm will be discussed in the next subsection.

4.8.1 Gibbs sampling procedure

Given the multiplicative excesses $z_1, z_2, \dots, z_k$, the Bayesian estimates of the PPD parameters γ, ρ and c are determined in order to fit a PPD to the multiplicative excesses.

As discussed in section 3.3.5, the PPD parameter ρ will be externally estimated. Let $\hat{\rho}$ denote the external estimate of the parameter ρ.

In order to obtain the remaining two PPD parameters, the Gibbs sampling procedure described by Berning (2010: 36-42) will be used. The procedure is as follows:

1. Obtain an initial estimate of the unknown parameter c, denoted by $c^{(0)}$.

2. For $i = 1,2, \dots N$:

   2.1. Generate $\gamma^{(i)}$, given $\hat{\rho}$ and $c^{(i-1)}$ from

$$\pi\left(\gamma\middle|\hat{\rho}\,, c^{(i-1)}, \boldsymbol{z}\right) \propto \frac{1}{\gamma} e^{-\gamma} \prod_{i=1}^{k} \gamma^{-1} z_i^{-\frac{1}{\gamma}-1} \left(1 - c^{(i-1)} + c^{(i-1)}(1 - \hat{\rho}\,) z_i^{\frac{\hat{\rho}}{\gamma}}\right).$$

   This formulation implies that $\gamma^{(i)}$ is a single simulated value of γ, generated from the conditional posterior of γ, given $\hat{\rho}$ and $c^{(i-1)}$.

2.2. Generate $c^{(i)}$, given $\gamma^{(i)}$ and $\hat{\rho}$ from

$$\pi\left(c\middle|\gamma^{(i)},\hat{\rho}\,,\boldsymbol{z}\right) \propto \frac{1}{\gamma^{(i)}}e^{-\gamma^{(i)}}\prod_{i=1}^{k}\left(\gamma^{(i)}\right)^{-1}z_i^{-\frac{1}{\gamma^{(i)}}-1}\left(1-c+c(1-\hat{\rho}\,)z_i^{\frac{\hat{\rho}}{\gamma^{(i)}}}\right).$$

This implies that $c^{(i)}$ is a single simulated value of c, generated from the conditional posterior of c, given $\gamma^{(i)}$ and $\hat{\rho}$.

3. The resulting vectors of simulated values for each parameter are denoted by $\boldsymbol{\gamma} = \left(\gamma^{(1)},\gamma^{(2)},\ldots,\gamma^{(N)}\right)$ and $\boldsymbol{c} = \left(c^{(1)},c^{(2)},\ldots,c^{(N)}\right)$.

   Assuming square error loss, the estimates of the expected value of the marginal posteriors are given by the arithmetic mean of these vectors. Therefore the Bayes estimates of the parameters are given by:

$$\hat{\gamma} = \frac{1}{N}\sum_{i=1}^{N}\gamma^{(i)}$$

$$\hat{c} = \frac{1}{N}\sum_{i=1}^{N}c^{(i)}.$$

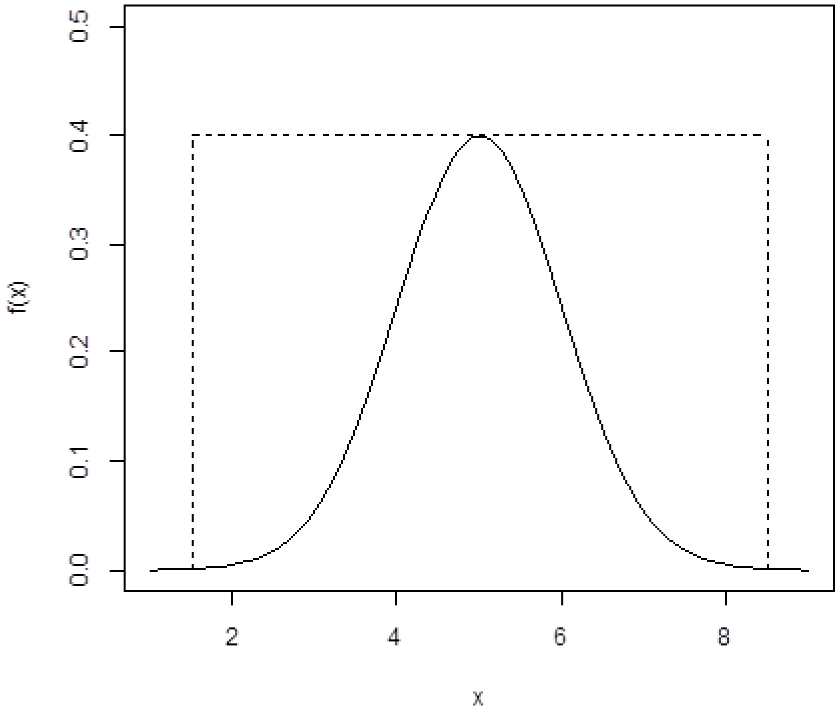
   Assuming absolute error loss, the Bayesian estimates of the parameters are given by the median of the marginal posterior distribution, obtained by calculating the median of the simulated vectors.

   Assuming zero-one loss, the Bayesian estimates of the parameters are given by the mode of the marginal posterior distribution. Generally squared error loss is assumed, unless another loss function is superior to a specific setting.

Considering steps 2.1 and 2.2, the rejection method will be used in this thesis to simulate one value from the given conditional posterior density. The rejection method will be explained in the next subsection and is based on the explanation given by Berning (2010: 37 – 41).

4.8.2 The rejection method

Suppose the goal is to simulate values from a given density function $f(x)$. This density function is graphically illustrated in figure 4.8.2.



Source: Berning (2010)

Figure 4.8.1: Display of density from which values are simulated

With the aim of applying the rejection method, it is necessary to construct the rectangle indicated in the graph above. In order to do this, the maximum of the density function and the interval in which the density is "significant" should be determined. Considering the graph, the "significant" interval is defined as all values of x for which $f(x) > 0.01\max(f(x))$. In the above example, $\max(f(x)) = 0.4$ and the interval of x values is $[x_{min} = 1.5,\ x_{max} = 8.5]$ (indicated by the dashed lines).

In order to simulate a single value from the density $f(x)$, the following steps are carried out:

i) Simulate a value x^* from a Uniform (x_{min}, x_{max}) distribution.

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

iii) If $y^* \leq f(x^*)$, then x^* is taken as the simulated value. Otherwise, reject x^* and repeat steps i) – iii).

Considering Step 2.1 of the Gibbs sampling procedure, the rejection method can now be used to simulate a value $\gamma^{(i)}$ from the conditional posterior

$$\pi\left(\gamma\middle|\hat{\rho}, c^{(i-1)}, \boldsymbol{z}\right) \propto \frac{1}{\gamma} e^{-\gamma} \prod_{i=1}^{k} \gamma^{-1} z_i^{-\frac{1}{\gamma}-1}\left(1 - c^{(i-1)} + c^{(i-1)}(1-\hat{\rho}) z_i^{\frac{\hat{\rho}}{\gamma}}\right).$$

For the sake of simplicity, let

$$f(\gamma) = \frac{1}{\gamma} e^{-\gamma} \prod_{i=1}^{k} \gamma^{-1} z_i^{-\frac{1}{\gamma}-1}\left(1 - c + c(1-\rho) z_i^{\frac{\rho}{\gamma}}\right).$$

Since the normalising constant of $f(\gamma)$ is unknown, Berning (2010: 40) suggests that $\log\left(f(\gamma)\right)$ should be computed first. After these values have been computed, another function, say $g(\gamma)$, should be obtained by scaling $\log\left(f(\gamma)\right)$ appropriately and then $\exp\left(g(\gamma)\right)$ should be calculated. In theory $f(\gamma)$ and $\exp\left(g(\gamma)\right)$ will give the same results, but working with $f(\gamma)$ can lead to some numerical problems.

The log of the conditional posterior is given by:

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

The procedure of obtaining a simulated value from the conditional posterior is as follows:

i) Choose a range of γ values, for example $\boldsymbol{\gamma} = (0.01, 0.02, \dots, 2)$.

ii) Calculate $\log\left(f(\boldsymbol{\gamma})\right) = \left(\log\left(f(0.01)\right), \log\left(f(0.02)\right), \dots, \log\left(f(2)\right)\right)$.

iii) Calculate $g(\boldsymbol{\gamma}) = \log\left(f(\boldsymbol{\gamma})\right) - \max\left(\log\left(f(\boldsymbol{\gamma})\right)\right)$.

iv) Calculate $\exp\left(g(\boldsymbol{\gamma})\right)$. A vector of “density values” at points in $\boldsymbol{\gamma}$ is obtained, rescaled so the maximum value is 1.

v) Keep only those values of $\exp\left(g(\boldsymbol{\gamma})\right)$ which are greater than 0.01 and their corresponding $\boldsymbol{\gamma}$ values.

vi) In order to simulate a single observation from $\pi(\boldsymbol{\gamma}|\rho, c, \boldsymbol{z})$, use the result obtained in v) and apply the rejection method.

The Gibbs sampling procedure described above will be applied in chapter 5, where the performance of the estimators for the PPD parameters will be evaluated in a simulation study.

4.9 Application of Bayesian methods

According to Spiegelhalter and Rice (2009), Bayesian statistical methods are primarily used in three situations.

The first situation is when prior knowledge and information is included in the analysis due to the lack of sufficient data. For example, if a policy decision must be made and information is available from multiple sources, Bayesian methods can be used.

The second situation is when there exist multiple sources of evidence that relates to a specific problem. In this situation hierarchical models can be constructed on the assumption that the prior distributions are from the same family. The parameters can then be estimated from the data. Application areas include meta-analysis, disease mapping and multi-centre studies.

The third situation is when a huge joint probability model is created. In this situation, thousands of observations and parameters are possible, and the only practical way of conducting inferences is by using Bayesian methods. Examples of such areas include image processing, spam filtering and signal analysis.

Bayesian statistics can also be applied in machine learning. In this area Bayesian methods are used to include each additional bit of information. Cui, Wong and Lui (2006) discuss using Bayesian networks to model consumer responses in order to improve the performance of marketing operations.

Bayesian methods are in general robust when little information is available and when uncertain or poorly determined parameters are used in making inferences about a population. Eddy (2004: 1177-1178) mentions that in a phylogenetic analysis, the probability of an evolutionary tree given some observed DNA sequence depends, amongst other things, on an evolutionary model and specific branch lengths of the tree. Both parameters however are subjected to substantial uncertainty. If Bayesian methods are used, these methods will result in probability models that are more realistic.

Lee and Wagenmakers (2005) discuss the use of Bayesian methods in the field of psychology. They argue that certain problems in this field can only be solved by using Bayesian inference.

It is clear that Bayesian methods can be applied in numerous fields and that, in certain situations, these methods can deliver superior results when compared to standard statistical methods.

4.10 Problems associated with Bayesian statistics

According to Eddy (2004: 1177-1178), there are mainly three difficulties associated with the use of the Bayesian approach. Firstly Bayesian calculations involve integrating over uncertain parameters. These integrations usually do not have an analytical solution and the calculations required to solve these integrations are computationally intensive.

Secondly, it may occur that Bayes' Theorem is applied to situations in which there is no underlying random process. This is known as the "method of inverse probability" (Edwards, 1972). According to this method, "the deductive argument leading from hypothesis to probability of results is inverted to form an inductive argument from results to probability of the hypothesis. Especially when combined with the use of subjectively determined probabilities as inputs, this method is commonly called the 'Bayesian approach' " (Ferson, 20-21).

Lastly, it is necessary in Bayesian statistics to specify a certain prior probability distribution. As mentioned this prior distribution is usually not known in advance and the choice of prior can prove to be a difficult task.

Recent articles in the literature suggest using empirical Bayes methods (Carlin and Louis, 1996). Empirical Bayes statistics involves methods and procedures in which the prior distribution is estimated directly from the data. Empirical Bayes statistics will be discussed in chapter 6.

4.11 Conclusion

In this chapter the basic results of Bayesian methodology were discussed and Bayes' Theorem was introduced. Bayes' Theorem is used to obtain the posterior distribution of the parameter of interest. This posterior distribution is obtained by using a specific prior distribution and the likelihood function of the parameter of interest.

In section 4.4 Bayesian inferences were presented which included a brief description of point estimation and interval estimation in the Bayesian context.

Section 4.5 discussed the two main categories of prior distributions, namely subjective priors and objective priors. The section also discussed some approaches that can be applied in order to obtain the optimal prior distribution.

Section 4.6 introduced Bayesian methodology in the extreme value context. The section mentioned some difficulties and advantages of using Bayesian methods in an EVT context.

In section 4.7 Gibbs sampling was introduced as the method that will be used in order to obtain Bayesian estimates of the PPD parameters. Zellner's MDI prior and the posterior distribution of the perturbed Pareto distribution were presented. Section 4.8 discussed Gibbs sampling in detail and the Gibbs algorithm was also described.

Section 4.9 mentioned some of the areas of application of Bayesian methods. The last section discussed some of the difficulties associated with Bayesian methods in general.

CHAPTER 5

Simulation Study

5.1 Introduction

The goal of the simulation study is to measure the performance of the different EVI estimators, as discussed in chapters 3 and 4.

Three first order estimators of the EVI will be considered in the simulation study. Subjective and adaptive threshold selection methods will be used to determine the optimal number of excesses. The two adaptive methods that will be considered are the method of Guillou and Hall (2001) and the method of Drees and Kaufmann (1998).

The second order parameter (ρ) of the perturbed Pareto distribution will be estimated externally. Two estimators of ρ will be considered in the simulation study. Maximum likelihood estimation and Bayesian estimation will then be used to estimate the remaining two parameters of the PPD.

Section 5.2 provides an outline of the methodology and the design of the simulation study. The results of the simulation study are discussed in section 5.3.

5.2 Methodology and design

The objective of this study is to measure the performance of different estimators of the EVI using some measure of error. The simulation study design followed in this chapter is based on the design proposed by Berning (2013, 11-12).

5.2.1 Simulation procedure

The following procedure is carried out for each estimator of the EVI in order to measure the performance of the estimator:

1. Choose a distribution for which the true underlying EVI is known. The distributions that will be considered are discussed in more detail in section 5.2.2.
2. Choose a sample size n. Sample sizes of $n = 200$, $n = 500$, $n = 1000$, $n = 2000$ and $n = 5000$ will be used in the simulation study.
3. Choose a threshold or, stated differently, the number of order statistics k. In this simulation study three methods will be used to determine the threshold. The first method is to use a fixed number of order statistics, namely to let $k = 2\sqrt{n}$. The rest of the methods are adaptive methods of threshold selection, namely the method of Guillou and Hall (2001) and the method of Drees and Kaufmann (1998).
4. Generate a sample of size n from the distribution chosen in step 2.
5. Given the threshold chosen in step 3, calculate the EVI estimate $\hat{\gamma}$.
6. Calculate the square error of estimation $(\hat{\gamma} - \gamma)^2$.
7. Repeat steps 4 – 6 a large number of times. In this simulation study these steps will be repeated 1000 times.
8. Calculate the mean of the square errors in order to obtain the mean square error (MSE). The MSE will be used to measure the performance of the EVI estimator under consideration.

5.2.2 Distributions

As mentioned in the previous section, the measure of error that will be used in this study is the MSE. In order to obtain the MSE of any estimator of the EVI for a specific distribution, it is necessary to know the theoretical expression for the EVI of that distribution. It is also necessary to know the theoretical expression of the second order parameter of that distribution, in order to choose values of the second order parameter which can realistically occur in practice.

Another important point to note is that observations simulated from one of the limiting distributions will result in an estimate of the EVI that will almost always lead to superior results. Therefore, the following distributions will not be included in the simulation study:

- Pareto distribution;
- Generalised Pareto distribution;
- Perturbed Pareto distribution;
- Generalised extreme value distribution.

Based on the above mentioned criteria, the following families of distributions will be considered:

Burr distribution

Recall that the distribution function of the Burr distribution is given by

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

defined for $x > 0$, with $\beta > 0,\ \tau > 0$ and $\lambda > 0$.

The theoretical expressions for the EVI and the second parameter ρ is given by

$$\gamma = \frac{1}{\lambda\tau}$$

$$\rho = -\frac{1}{\lambda}.$$

Rewriting these terms, it follows that $\tau = -\frac{\rho}{\gamma}$ and $\lambda = -\frac{1}{\rho}$. It will therefore be possible to rewrite the distribution function in terms of β, γ and ρ:

$$F(x) = 1 - \left(\frac{\beta}{\beta + x^{-\frac{\rho}{\gamma}}}\right)^{-\frac{1}{\rho}}.$$

The following six Burr distributions will be considered in the simulation study:

- $Burr(\beta = 1, \gamma = 0.25, \rho = -2)$
- $Burr(\beta = 1, \gamma = 0.5, \rho = -2)$
- $Burr(\beta = 1, \gamma = 1, \rho = -2)$
- $Burr(\beta = 1, \gamma = 0.25, \rho = -0.5)$
- $Burr(\beta = 1, \gamma = 0.5, \rho = -0.5)$
- $Burr(\beta = 1, \gamma = 1, \rho = -0.5)$

Frechét distribution

The distribution function of the Frechét distribution is given by

$$F(x) = \exp(-x^{-\alpha}),$$

defined for $x > 0$, with $\alpha > 0$.

The theoretical expressions for the EVI and the second parameter ρ is given by

$$\gamma = \frac{1}{\alpha}$$

$$\rho = -1.$$

The following three Frechét distributions will be considered in the simulation study:

- $Frech\acute{e}t(\alpha = 1)$, for which $\gamma = 1$.
- $Frech\acute{e}t(\alpha = 2)$, for which $\gamma = 0.5$.
- $Frech\acute{e}t(\alpha = 4)$, for which $\gamma = 0.25$.

Loggamma distribution

The distribution function of the loggamma distribution is given by

$$F(x) = 1 - \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \int_{1}^{x} w^{-\lambda-1} (\log w)^{\alpha-1} dw,$$

defined for $x > 1$, with $\lambda > 0$ and $\alpha > 0$.

The theoretical expressions for the EVI and the second parameter ρ is given by

$$\gamma = \frac{1}{\lambda}$$

$$\rho = 0.$$

Since $\alpha = 1$ results in the Pareto distribution, $\alpha = 2$ will be used.

The following three loggamma distributions will be considered in the simulation study:

- $log\Gamma(\lambda = 1, \alpha = 2)$, for which $\gamma = 1$.
- $log\Gamma(\lambda = 2, \alpha = 2)$, for which $\gamma = 0.5$.
- $log\Gamma(\lambda = 4, \alpha = 2)$, for which $\gamma = 0.25$.

Student t distribution

The distribution function of the Student t distribution is given by

$$F(x) = 1 - \frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \int_x^\infty \left(1 + \frac{w^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} dw,$$

defined for $x > 0$, with $n > 0$.

Only observations on the positive half-line $(0, \infty)$ of the t distribution will be used, since only positive values will be used in the simulation study. The distribution function becomes

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

The theoretical expressions for the EVI and the second parameter ρ is given by

$$\gamma = \frac{1}{n}$$

$$\rho = -\frac{2}{n}.$$

The following three t distributions will be considered in the simulation study:

- $|t_1|$ distribution
- $|t_2|$ distribution
- $|t_4|$ distribution

The simulated data for each of the 15 distributions are obtained by means of the program *R*. All calculations are also performed using *R*.

5.2.3 Estimators of the EVI

The simulation study will include the following six estimators of the EVI:

1. $\hat{\gamma}_0$: Hill estimate where $k = 2\sqrt{n}$.
2. $\hat{\gamma}_{GH.1.5}$: Hill estimate where k is determined by using the method of Guillou and Hall (2001), with critical value 1.5.
3. $\hat{\gamma}_{GH.1.25}$: Hill estimate where k is determined by using the method of Guillou and Hall (2001), with critical value 1.25.
4. $\hat{\gamma}_{DK}$: Hill estimate where k is determined by using the method of Drees and Kauffman (1998).
5. $\hat{\gamma}_0^{MLE}$: EVI estimate obtained by fitting the PPD to the relative excesses using MLE, after externally estimating the second order parameter.
6. $\hat{\gamma}_0^{Bayes}$: EVI estimate obtained by fitting the PPD to the relative excesses using Bayesian methodology, after externally estimating the second order parameter.

In this case of the last two estimators ($\hat{\gamma}_0^{MLE}$ and $\hat{\gamma}_0^{Bayes}$), the number of order statistics used when fitting the PPD is fixed at $k = 2n^{2/3}$ (Berning, 2013: 22).

5.2.4 Computational issues

When estimating the EVI by means of fitting the PPD to the excesses using MLE, it is necessary to first estimate the second order parameter ρ. The approach that will be followed in this study is to externally estimate the second order parameter ρ. This involves obtaining an estimate of ρ and then using the estimated value $\hat{\rho}$ as substitute for ρ in Definition 3.3.4. MLE will then be used to estimate γ and c, the remaining two parameters.

As mentioned in section 3.3.5, two estimators of ρ will be investigated. Recall that the first estimator of ρ is defined as

$$\hat{\rho}_1 := -\left|\frac{3\left(T_n^{(0)}(k_1)-1\right)}{T_n^{(0)}(k_1)-3}\right|$$

where $k_1 = min(n-1, [2n/\log\log n])$,

$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\log\frac{X_{n-i+1,n}}{X_{n-k,n}}\right)^j$, $j \geq 1$ and

$$T_n^{(\tau)}(k) := \begin{cases} \dfrac{\left(M_n^{(1)}(k)\right)^{\tau} - \left(M_n^{(1)}(k)/2\right)^{\tau/2}}{\left(M_n^{(2)}(k)/2\right)^{\tau/2} - \left(M_n^{(3)}(k)/6\right)^{\tau/3}} & if\ \tau > 0 \\ \\ \dfrac{\log\left(M_n^{(1)}(k)\right) - \frac{1}{2}\log\left(M_n^{(2)}(k)/2\right)}{\frac{1}{2}\log\left(M_n^{(2)}(k)/2\right) - \frac{1}{3}\log\left(M_n^{(3)}(k)/6\right)} & if\ \tau = 0. \end{cases}$$

The second estimator of ρ is defined as

$\hat{\rho}_2$ as the solution of $(1-\hat{\rho})^{\hat{\alpha}-1}\big(1+\hat{\rho}(\hat{\alpha}-2)\big) = 1$

where $i_n = 50\%$ of n,

$j_n = 90\%$ of n,

$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}}\Sigma(\alpha, i_n, j_n)$,

$\Sigma(\alpha, i_n, j_n) = \sum_{k=i_n}^{j_n}\left(\hat{\gamma}_n^{(\alpha)}(k) - \chi(\alpha, i_n, j_n)\right)^2$,

$\chi(\alpha, i_n, j_n) =$ sample median of $\hat{\gamma}_n^{(\alpha)}(k)$, where $k = i_n, \ldots, j_n$, and

$\hat{\gamma}_n^{(\alpha)}(k) = \frac{M_n^{(\alpha)}(k)}{\Gamma(\alpha+1)(H_{k,n})^{\alpha-1}}$, where $H_{k,n}$ denotes the Hill estimator, based on k excesses.

The performance of these two estimators will be discussed in section 5.3.1.

The second computational issue involves restricting the range of the parameter c. By carrying out extensive simulation studies, Berning (2013: 7) found that applying the restriction $c \leq 0.5$ improves the accuracy of the EVI estimates significantly.

5.3 Discussion of simulation results

5.3.1 Results for estimators of the second order parameter

The following tables show the calculated MSE values of the two estimators of $\hat{\rho}$ with respect to the true underlying value ρ.

In the event of the calculated value of $\hat{\rho}$ being smaller than -4, the calculated value is set equal to -4. The reason for this is that if all extremely low estimates of $\hat{\rho}$ are included in the MSE calculation, these low estimates can influence the results significantly. It might even indicate that the estimator should not even be considered in the simulation study.

Studies have shown (Fraga Alves (2002); Gomes, Caeiro and Figueiredo (2004); and Gomes, Martins and Neves (2007)) that -2 is a realistic lower bound for ρ. The value -4 is used since this is twice the lower bound of ρ.

Note that the above mentioned restriction is not applied in section 5.3.2, where the MSE values for the EVI estimates are calculated.

The standard error of a specific MSE value is shown below the corresponding MSE value in brackets. The 15 distributions are grouped together according to the theoretical value of the second order parameter.

The mean value of each group is obtained by calculating the average of the MSEs of the distributions in that group. The overall mean value is obtained by calculating the average of the MSEs of the 15 distributions.

The standard error reported below the mean value is calculated as follows. Let s_j denote the standard error of distribution j (as given in brackets) and m the number of distributions in the group. The standard error is given by

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

The overall standard error is given by

$$\sqrt{\frac{\sum_{j=1}^{15} s_j^2}{15^2}}.$$

			$n = 200$		$n = 500$		$n = 1000$	
Distr.	γ	ρ	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_1$	$\hat{\rho}_2$
$Burr$	0.25	-2	1.0207	1.4431	1.1092	0.8733	1.1747	0.6629
			(0.0095)	(0.0392)	(0.0058)	(0.0265)	(0.0045)	(0.0205)
$Burr$	0.5	-2	1.0189	1.4611	1.1072	0.9157	1.1792	0.6653
			(0.0098)	(0.0409)	(0.0056)	(0.0295)	(0.0043)	(0.0203)
$Burr$	1	-2	1.0240	1.4669	1.0972	0.8976	1.1738	0.6328
			(0.0090)	(0.0399)	(0.0063)	(0.0293)	(0.0042)	(0.0201)
$\lvert t_1 \rvert$	1	-2	1.6222	3.2577	1.4573	1.1935	1.6173	3.1472
			(0.0011)	(0.0127)	(0.0022)	(0.0201)	(0.0003)	(0.0045)
Mean		**-2**	**1.1714**	**1.9072**	**1.1927**	**0.9700**	**1.2862**	**1.2771**
			(0.0073)	**(0.0332)**	**(0.0050)**	**(0.0117)**	**(0.0033)**	**(0.0163)**
$Frechét(4)$	0.25	-1	0.3125	0.6332	0.1066	0.4201	0.0670	0.2556
			(0.0369)	(0.0297)	(0.0103)	(0.0245)	(0.0092)	(0.0197)
$Frechét(2)$	0.5	-1	0.3112	0.6109	0.0910	0.4143	0.0577	0.2540
			(0.0397)	(0.0305)	(0.0094)	(0.0251)	(0.0035)	(0.0167)
$Frechét(1)$	1	-1	0.2009	0.6378	0.1084	0.4294	0.0568	0.2231
			(0.0242)	(0.0330)	(0.0121)	(0.0268)	(0.0041)	(0.0146)
$\lvert t_2 \rvert$	0.5	-1	0.0662	0.4032	0.0670	0.3539	0.0685	0.3509
			(0.0003)	(0.0063)	(0.0002)	(0.0039)	(0.0001)	(0.0030)
Mean		**-1**	**0.2227**	**0.5713**	**0.0933**	**0.4044**	**0.0625**	**0.2709**
			(0.0253)	**(0.0249)**	**(0.0080)**	**(0.0201)**	**(0.0042)**	**(0.0135)**
$Burr$	0.25	-0.5	0.0650	0.0643	0.0635	0.0336	0.0618	0.0282
			(0.0003)	(0.0023)	(0.0002)	(0.0010)	(0.0001)	(0.0006)
$Burr$	0.5	-0.5	0.0644	0.0620	0.0634	0.0342	0.0619	0.0280
			(0.0003)	(0.0022)	(0.0002)	(0.0010)	(0.0001)	(0.0006)
$Burr$	1	-0.5	0.0650	0.0602	0.0634	0.0354	0.0618	0.0278
			(0.0003)	(0.0023)	(0.0002)	(0.0011)	(0.0001)	(0.0006)
$\lvert t_4 \rvert$	0.25	-0.5	0.1010	0.4581	0.0520	0.0799	0.0787	0.2767
			(0.0015)	(0.0324)	(0.0001)	(0.0012)	(0.0004)	(0.0100)
Mean		**-0.5**	**0.0738**	**0.1611**	**0.0606**	**0.0458**	**0.0661**	**0.0901**
			(0.0006)	**(0.0098)**	**(0.0001)**	**(0.0011)**	**(0.0002)**	**(0.0029)**
$\log\Gamma(4,2)$	0.25	0	2.6486	1.4453	2.6875	1.4165	2.3638	1.3109
			(0.1187)	(0.0765)	(0.1046)	(0.0651)	(0.0742)	(0.0522)
$\log\Gamma(2,2)$	0.5	0	2.7056	1.4752	2.4818	1.3540	2.3979	1.3252
			(0.1245)	(0.0792)	(0.0893)	(0.0605)	(0.0698)	(0.0530)
$\log\Gamma(1,2)$	1	0	2.9600	1.4340	2.8488	1.3777	2.3897	1.3612
			(0.1318)	(0.0790)	(0.1124)	(0.0629)	(0.0730)	(0.0542)
Mean		**0**	**2.7714**	**1.4515**	**2.6727**	**1.3827**	**2.3838**	**1.3324**
			(0.1250)	**(0.0782)**	**(0.1021)**	**(0.0628)**	**(0.0723)**	**(0.0531)**
Overall mean			**1.0598**	**1.0228**	**1.0048**	**0.7007**	**0.9496**	**0.7426**
			(0.0396)	**(0.0365)**	**(0.0288)**	**(0.0239)**	**(0.0200)**	**(0.0215)**

Table 5.3.1 Overall MSE for estimates of ρ for samples sizes $n = 200,\ 500$ and 1000

			$n = 2000$		$n = 5000$	
Distr.	γ	ρ	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_1$	$\hat{\rho}_2$
$Burr$	0.25	-2	0.7297	0.4617	0.3780	0.3856
			(0.0049)	(0.0123)	(0.0057)	(0.0080)
$Burr$	0.5	-2	0.7408	0.4750	0.3893	0.3990
			(0.0046)	(0.0112)	(0.0057)	(0.0079)
$Burr$	1	-2	0.7371	0.4901	0.3913	0.4018
			(0.0048)	(0.0126)	(0.0053)	(0.0080)
$\lvert t_1 \rvert$	1	-2	1.2839	1.1057	1.7163	3.1350
			(0.0020)	(0.0113)	(0.0007)	(0.0018)
Mean		**-2**	**0.8729**	**0.6331**	**0.7187**	**1.0803**
			(0.0041)	**(0.0118)**	**(0.0043)**	**(0.0064)**
$Frechét(4)$	0.25	-1	0.1000	0.1218	0.1078	0.0545
			(0.0056)	(0.0095)	(0.0048)	(0.0031)
$Frechét(2)$	0.5	-1	0.1136	0.1411	0.0907	0.0454
			(0.0064)	(0.0099)	(0.0041)	(0.0026)
$Frechét(1)$	1	-1	0.1024	0.1257	0.1079	0.0564
			(0.0045)	(0.0089)	(0.0054)	(0.0032)
$\lvert t_2 \rvert$	0.5	-1	0.0632	0.3386	0.0610	0.3404
			(0.0001)	(0.0021)	(0.0002)	(0.0013)
Mean		**-1**	**0.0948**	**0.1818**	**0.0918**	**0.1242**
			(0.0042)	**(0.0076)**	**(0.0036)**	**(0.0025)**
$Burr$	0.25	-0.5	0.0657	0.0268	0.0568	0.0233
			(0.0002)	(0.0005)	(0.0002)	(0.0003)
$Burr$	0.5	-0.5	0.0660	0.0255	0.0561	0.0241
			(0.0002)	(0.0004)	(0.0002)	(0.0003)
$Burr$	1	-0.5	0.0660	0.0252	0.0562	0.0242
			(0.0002)	(0.0005)	(0.0002)	(0.0003)
$\lvert t_4 \rvert$	0.25	-0.5	0.0480	0.0746	0.2724	0.2324
			(0.0001)	(0.0005)	(0.0019)	(0.0036)
Mean		**-0.5**	**0.0614**	**0.0380**	**0.1104**	**0.0760**
			(0.0001)	**(0.0005)**	**(0.0006)**	**(0.0011)**
$\log\Gamma(4,2)$	0.25	0	1.8183	1.0957	1.3490	0.9833
			(0.0435)	(0.0359)	(0.0248)	(0.0196)
$\log\Gamma(2,2)$	0.5	0	1.8985	1.1461	1.3486	0.9626
			(0.0531)	(0.0349)	(0.0297)	(0.0186)
$\log\Gamma(1,2)$	1	0	1.9307	1.1831	1.3579	0.9466
			(0.0547)	(0.0395)	(0.0268)	(0.0182)
Mean		**0**	**1.8825**	**1.1416**	**1.3518**	**0.9642**
			(0.0504)	**(0.0368)**	**(0.0271)**	**(0.0188)**
Overall mean			**0.7279**	**0.4986**	**0.5682**	**0.5612**
			(0.0147)	**(0.0142)**	**(0.0089)**	**(0.0072)**

Table 5.3.2 Overall MSE for estimates of ρ for sample sizes $n = 2000$ and 5000

The results indicate that $\hat{\rho}_2$ performs overall better than $\hat{\rho}_1$. Based on this result, $\hat{\rho}_2$ will be used to externally estimate the second order parameter ρ when calculating $\hat{\gamma}_0^{MLE}$ and $\hat{\gamma}_0^{Bayes}$.

5.3.2 Results for estimators of the EVI

The results provided in the tables below indicate the calculated MSE values of the six estimators of the EVI. The standard error of a specific MSE value is again shown below the corresponding MSE value in brackets. Note that the MSE values and their corresponding standard errors are multiplied with a factor of 1000 to ease comparison.

The 15 distributions are grouped together according to the theoretical value of the EVI. Therefore three groups can be formed, each group consisting of five distributions.

As in section 5.3.1, the mean value of each group is obtained by calculating the average of the MSEs of the distributions in that group. The overall mean value is obtained by calculating the average of the MSEs of the 15 distributions.

Again let s_j denote the standard error of distribution j (as given in brackets). The standard error is given by

$$\sqrt{\frac{\sum_{j=1}^{5} s_j^2}{5^2}}.$$

The overall standard error is given by

$$\sqrt{\frac{\sum_{j=1}^{15} s_j^2}{15^2}}.$$

For the sake of convenience, the lowest MSE value in each row is highlighted.

Sample size $n = 200$

Dist	γ	ρ	$\hat{\gamma}_0$	$\hat{\gamma}_{GH1.5}$	$\hat{\gamma}_{GH1.25}$	$\hat{\gamma}_{DK}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_0^{Bayes}$
$Burr$	0.25	-2	2.357	1.753	2.055	1.462	2.241	1.809
			(0.112)	(0.099)	(0.127)	(0.064)	(0.094)	(0.072)
$Burr$	0.25	-0.5	11.277	18.855	15.484	12.516	5.635	8.253
			(0.409)	(0.569)	(0.529)	(0.442)	(0.307)	(0.326)
$Frechét(4)$	0.25	-1	2.335	2.937	2.969	2.125	2.539	2.240
			(0.114)	(0.173)	(0.210)	(0.110)	(0.102)	(0.097)
$\lvert t_4 \rvert$	0.25	-0.5	21.240	30.458	24.404	20.735	10.573	15.925
			(0.604)	(0.846)	(0.742)	(0.660)	(0.359)	(0.426)
$\log\Gamma(4,2)$	0.25	0	6.755	9.805	8.956	6.740	4.567	4.649
			(0.316)	(0.227)	(0.262)	(0.316)	(0.192)	(0.187)
Mean			**8.793**	**12.762**	**10.774**	**8.716**	**5.111**	**6.575**
			(0.162)	**(0.213)**	**(0.196)**	**(0.173)**	**(0.106)**	**(0.116)**
$Burr$	0.5	-2	9.849	6.483	9.015	5.957	9.189	7.292
			(0.440)	(0.296)	(0.602)	(0.286)	(0.378)	(0.305)
$Burr$	0.5	-0.5	43.470	75.524	60.577	48.714	22.711	31.282
			(1.455)	(2.155)	(1.929)	(1.670)	(1.098)	(1.158)
$Frechét(2)$	0.5	-1	9.658	12.054	14.577	9.382	9.737	8.500
			(0.475)	(0.445)	(1.999)	(0.447)	(0.369)	(0.342)
$\lvert t_2 \rvert$	0.5	-1	13.191	26.233	24.481	15.574	10.637	9.473
			(0.626)	(0.950)	(1.028)	(0.669)	(0.786)	(0.491)
$\log\Gamma(2,2)$	0.5	0	24.362	36.810	33.680	24.324	17.359	16.966
			(1.004)	(0.968)	(1.145)	(1.010)	(0.624)	(0.648)
Mean			**20.106**	**31.421**	**28.466**	**20.790**	**13.926**	**14.703**
			(0397)	**(0.520)**	**(0.647)**	**(0.426)**	**(0.316)**	**(0.298)**
$Burr$	1	-2	34.196	25.645	32.256	21.078	34.199	26.512
			(1.546)	(1.371)	(2.166)	(1.013)	(1.448)	(1.155)
$Burr$	1	-0.5	183.147	310.797	255.443	200.319	89.666	107.944
			(6.154)	(8.943)	(7.949)	(6.916)	(4.094)	(3.939)
$Frechét(1)$	1	-1	38.194	46.530	51.373	38.934	40.593	32.091
			(1.799)	(1.953)	(4.586)	(1.830)	(1.576)	(1.272)
$\lvert t_1 \rvert$	1	-2	38.824	37.378	50.374	27.916	45.795	32.264
			(1.818)	(1.814)	(6.828)	(1.298)	(2.144)	(1.372)
$\log\Gamma(1,2)$	1	0	101.281	150.155	134.504	101.804	69.515	56.537
			(4.356)	(3.531)	(3.632)	(4.301)	(2.812)	(2.424)
Mean			**79.128**	**114.101**	**104.790**	**78.010**	**55.954**	**51.07**
			(1.622)	**(2.014)**	**(2.439)**	**(1.702)**	**(1.164)**	**(1.024)**
Overall mean			**36.009**	**52.761**	**48.010**	**35.839**	**24.997**	**24.116**
			(0.559)	**(0.697)**	**(0.844)**	**(0.588)**	**(0.403)**	**(0.358)**

Table 5.3.3 MSEs x 1000 for estimates of the EVI for sample size $n = 200$

Sample size $n = 500$

Dist	γ	ρ	$\hat{\gamma}_0$	$\hat{\gamma}_{GH1.5}$	$\hat{\gamma}_{GH1.25}$	$\hat{\gamma}_{DK}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_0^{Bayes}$
$Burr$	0.25	-2	1.343	0.797	1.158	0.630	1.236	1.183
			(0.060)	(0.035)	(0.143)	(0.028)	(0.053)	(0.052)
$Burr$	0.25	-0.5	6.377	11.064	9.187	8.111	3.244	4.713
			(0.221)	(0.316)	(0.295)	(0.241)	(0.164)	(0.171)
$Frechét(4)$	0.25	-1	1.501	1.507	1.635	0.903	1.671	1.592
			(0.069)	(0.067)	(0.103)	(0.043)	(0.080)	(0.070)
$\|t_4\|$	0.25	-0.5	11.206	16.711	13.406	12.061	4.969	7.272
			(0.308)	(0.450)	(0.398)	(0.344)	(0.160)	(0.204)
$\log\Gamma(4,2)$	0.25	0	4.519	6.971	5.978	4.554	3.501	4.157
			(0.179)	(0.145)	(0.143)	(0.181)	(0.136)	(0.150)
Mean			**4.989**	**7.410**	**6.273**	**5.252**	**2.924**	**3.784**
			(0.086)	**(0.115)**	**(0.109)**	**(0.092)**	**(0.057)**	**(0.063)**
$Burr$	0.5	-2	5.534	3.264	4.103	2.478	5.239	4.733
			(0.267)	(0.264)	(0.468)	(0.116)	(0.232)	(0.196)
$Burr$	0.5	-0.5	25.337	43.158	35.571	30.527	12.158	16.463
			(0.897)	(1.227)	(1.096)	(0.950)	(0.735)	(0.637)
$Frechét(2)$	0.5	-1	5.512	5.965	6.521	3.797	5.962	5.766
			(0.232)	(0.239)	(0.363)	(0.167)	(0.236)	(0.247)
$\|t_2\|$	0.5	-1	7.965	14.084	14.217	8.043	7.993	5.820
			(0.408)	(0.486)	(0.630)	(0.319)	(0.548)	(0.286)
$\log\Gamma(2,2)$	0.5	0	16.002	28.459	24.124	16.111	13.188	15.315
			(0.620)	(0.596)	(0.582)	(0.614)	(0.510)	(0.559)
Mean			**12.070**	**18.986**	**16.907**	**12.191**	**8.908**	**9.619**
			(0.243)	**(0.298)**	**(0.303)**	**(0.239)**	**(0.220)**	**(0.190)**
$Burr$	1	-2	20.908	19.214	34.538	9.104	19.368	16.622
			(0.915)	(5.201)	(15.059)	(0.424)	(0.755)	(0.730)
$Burr$	1	-0.5	97.816	170.134	143.243	117.552	45.962	57.819
			(3.525)	(4.932)	(5.252)	(3.942)	(2.394)	(2.324)
$Frechét(1)$	1	-1	23.160	23.642	26.897	17.497	23.806	20.890
			(1.036)	(0.908)	(2.135)	(0.788)	(0.996)	(0.852)
$\|t_1\|$	1	-2	23.216	17.483	28.702	12.136	31.347	21.270
			(1.079)	(1.992)	(5.350)	(0.562)	(1.840)	(0.996)
$\log\Gamma(1,2)$	1	0	73.412	112.912	97.828	74.126	61.512	57.159
			(2.901)	(2.477)	(2.528)	(2.923)	(2.416)	(2.079)
Mean			**47.702**	**68.677**	**66.241**	**46.083**	**36.399**	**34.752**
			(0.978)	**(1.579)**	**(3.429)**	**(1.004)**	**(0.813)**	**(0.692)**
Overall mean			**21.587**	**31.691**	**29.807**	**21.175**	**16.077**	**16.052**
			(0.337)	**(0.537)**	**(1.148)**	**(0.345)**	**(0.281)**	**(0.240)**

Table 5.3.4 MSEs x 1000 for estimates of the EVI for sample size $n = 500$

Sample size $n = 1000$

Dist	γ	ρ	$\hat{\gamma}_0$	$\hat{\gamma}_{GH1.5}$	$\hat{\gamma}_{GH1.25}$	$\hat{\gamma}_{DK}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_0^{Bayes}$
$Burr$	0.25	-2	1.042	0.528	0.856	0.336	0.936	0.923
			(0.049)	(0.065)	(0.172)	(0.015)	(0.046)	(0.041)
$Burr$	0.25	-0.5	3.940	7.035	6.130	5.568	1.812	2.750
			(0.137)	(0.201)	(0.267)	(0.159)	(0.087)	(0.110)
$Frech\acute{e}t(4)$	0.25	-1	1.005	0.954	1.106	0.516	1.066	1.084
			(0.052)	(0.049)	(0.077)	(0.021)	(0.061)	(0.049)
$\|t_4\|$	0.25	-0.5	6.839	10.235	9.093	8.005	2.737	3.943
			(0.2)	(0.38)	(0.421)	(0.214)	(0.107)	(0.129)
$\log\Gamma\,(4{,}2)$	0.25	0	3.402	5.585	4.817	3.783	2.943	3.827
			(0.125)	(0.11)	(0.106)	(0.121)	(0.110)	(0.123)
Mean			**3.246**	**4.867**	**4.400**	**3.642**	**1.899**	**2.505**
			(0.056)	**(0.090)**	**(0.109)**	**(0.059)**	**(0.038)**	**(0.044)**
$Burr$	0.5	-2	4.144	2.209	3.794	1.399	3.410	3.363
			(0.211)	(0.141)	(0.751)	(0.071)	(0.152)	(0.148)
$Burr$	0.5	-0.5	15.947	28.656	26.268	20.493	7.375	10.425
			(0.537)	(0.806)	(2.306)	(0.602)	(0.393)	(0.403)
$Frech\acute{e}t(2)$	0.5	-1	4.074	3.818	3.788	2.373	4.020	4.276
			(0.198)	(0.159)	(0.234)	(0.093)	(0.171)	(0.176)
$\|t_2\|$	0.5	-1	4.898	8.940	8.819	5.014	5.322	3.939
			(0.249)	(0.357)	(0.487)	(0.187)	(0.323)	(0.191)
$\log\Gamma\,(2{,}2)$	0.5	0	13.235	21.973	19.754	14.454	11.540	14.501
			(0.514)	(0.446)	(0.595)	(0.498)	(0.436)	(0.472)
Mean			**8.460**	**13.119**	**12.485**	**8.747**	**6.333**	**7.301**
			(0.167)	**(0.202)**	**(0.511)**	**(0.162)**	**(0.142)**	**(0.138)**
$Burr$	1	-2	16.561	7.49	9.912	5.177	13.769	13.044
			(0.816)	(0.587)	(1.045)	(0.233)	(0.675)	(0.548)
$Burr$	1	-0.5	63.081	114.326	99.236	74.932	29.251	38.005
			(2.107)	(2.978)	(3.155)	(2.399)	(1.512)	(1.495)
$Frech\acute{e}t(1)$	1	-1	17.062	15.826	16.775	10.716	15.787	15.546
			(0.835)	(0.742)	(1.962)	(0.428)	(0.737)	(0.725)
$\|t_1\|$	1	-2	15.114	8.654	19.091	6.359	19.468	14.281
			(0.71)	(0.471)	(7.307)	(0.306)	(0.963)	(0.625)
$\log\Gamma\,(1{,}2)$	1	0	52.916	88.606	77.122	58.481	45.375	73.320
			(1.934)	(1.732)	(1.735)	(1.864)	(1.660)	(3.896)
Mean			**32.947**	**46.980**	**44.427**	**31.133**	**24.730**	**30.839**
			(0.634)	**(0.721)**	**(1.689)**	**(0.618)**	**(0.528)**	**(0.863)**
Overall mean			**14.884**	**21.656**	**20.437**	**14.507**	**10.987**	**13.549**
			(0.219)	**(0.251)**	**(0.589)**	**(0.214)**	**(0.183)**	**(0.292)**

Table 5.3.5 MSEs x 1000 for estimates of the EVI for sample size $n = 1000$

Sample size $n = 2000$

Dist	γ	ρ	$\hat{\gamma}_0$	$\hat{\gamma}_{GH1.5}$	$\hat{\gamma}_{GH1.25}$	$\hat{\gamma}_{DK}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_0^{Bayes}$
$Burr$	0.25	-2	0.746	0.307	0.372	0.192	0.584	0.649
			(0.036)	(0.030)	(0.038)	(0.009)	(0.025)	(0.026)
$Burr$	0.25	-0.5	2.553	4.751	4.072	3.878	1.084	1.620
			(0.092)	(0.135)	(0.152)	(0.103)	(0.062)	(0.069)
$Frech\acute{e}t(4)$	0.25	-1	0.707	0.567	0.619	0.364	0.698	0.748
			(0.031)	(0.022)	(0.038)	(0.014)	(0.033)	(0.033)
$\|t_4\|$	0.25	-0.5	4.403	6.671	5.612	5.625	1.448	2.156
			(0.131)	(0.180)	(0.181)	(0.148)	(0.057)	(0.079)
$\log\Gamma(4,2)$	0.25	0	2.601	4.345	3.820	4.221	2.423	3.221
			(0.098)	(0.080)	(0.082)	(0.086)	(0.092)	(0.103)
Mean			**2.202**	**3.328**	**2.899**	**2.856**	**1.247**	**1.679**
			(0.039)	**(0.048)**	**(0.051)**	**(0.040)**	**(0.026)**	**(0.031)**
$Burr$	0.5	-2	2.767	1.035	1.795	0.651	2.180	2.361
			(0.120)	(0.060)	(0.379)	(0.029)	(0.097)	(0.107)
$Burr$	0.5	-0.5	10.260	18.696	15.963	14.334	4.498	6.564
			(0.342)	(0.508)	(0.488)	(0.397)	(0.254)	(0.272)
$Frech\acute{e}t(2)$	0.5	-1	3.312	2.529	2.584	1.581	2.893	3.189
			(0.149)	(0.113)	(0.136)	(0.058)	(0.142)	(0.134)
$\|t_2\|$	0.5	-1	3.425	5.339	5.281	3.237	4.334	3.056
			(0.161)	(0.235)	(0.304)	(0.136)	(0.253)	(0.149)
$\log\Gamma(2,2)$	0.5	0	10.703	17.771	15.387	16.829	9.580	12.361
			(0.363)	(0.325)	(0.317)	(0.311)	(0.342)	(0.407)
Mean			**6.093**	**9.074**	**8.202**	**7.326**	**4.697**	**5.506**
			(0.112)	**(0.132)**	**(0.154)**	**(0.105)**	**(0.105)**	**(0.108)**
$Burr$	1	-2	11.623	4.868	6.184	2.798	8.620	9.068
			(0.516)	(0.382)	(0.617)	(0.132)	(0.366)	(0.371)
$Burr$	1	-0.5	43.878	77.720	68.089	52.704	17.676	24.551
			(1.452)	(2.103)	(2.269)	(1.648)	(0.955)	(1.007)
$Frech\acute{e}t(1)$	1	-1	11.163	9.973	10.170	6.465	10.807	11.333
			(0.514)	(0.827)	(1.074)	(0.242)	(0.493)	(0.480)
$\|t_1\|$	1	-2	10.804	5.170	7.039	3.571	14.447	11.461
			(0.513)	(0.289)	(0.611)	(0.155)	(0.909)	(0.563)
$\log\Gamma(1,2)$	1	0	44.115	71.579	63.383	68.047	39.258	46.915
			(1.574)	(1.375)	(1.585)	(1.415)	(1.377)	(1.498)
Mean			**24.317**	**33.862**	**30.973**	**26.717**	**18.162**	**20.665**
			(0.464)	**(0.538)**	**(0.619)**	**(0.439)**	**(0.401)**	**(0.397)**
Overall mean			**10.871**	**15.421**	**14.025**	**12.300**	**8.035**	**9.283**
			(0.160)	**(0.185)**	**(0.213)**	**(0.151)**	**(0.461)**	**(0.138)**

Table 5.3.6 MSEs x 1000 for estimates of the EVI for sample size $n = 2000$

Sample size $n = 5000$

Dist	γ	ρ	$\hat{\gamma}_0$	$\hat{\gamma}_{GH1.5}$	$\hat{\gamma}_{GH1.25}$	$\hat{\gamma}_{DK}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_0^{Bayes}$
$Burr$	0.25	-2	0.451	0.137	0.171	0.086	0.341	0.393
			(0.022)	(0.009)	(0.016)	(0.004)	(0.015)	(0.018)
$Burr$	0.25	-0.5	1.645	2.959	3.049	2.726	0.691	1.026
			(0.055)	(0.082)	(0.406)	(0.063)	(0.034)	(0.045)
$Frechét(4)$	0.25	-1	0.467	0.350	0.338	0.213	0.405	0.487
			(0.022)	(0.022)	(0.030)	(0.008)	(0.018)	(0.022)
$\|t_4\|$	0.25	-0.5	2.554	4.149	3.488	3.540	0.704	1.020
			(0.075)	(0.117)	(0.112)	(0.082)	(0.031)	(0.042)
$\log\Gamma(4,2)$	0.25	0	2.167	3.429	2.963	3.926	2.067	2.651
			(0.067)	(0.058)	(0.058)	(0.045)	(0.067)	(0.076)
Mean			**1.457**	**2.205**	**2.002**	**2.098**	**0.842**	**1.115**
			(0.024)	**(0.031)**	**(0.085)**	**(0.023)**	**(0.017)**	**(0.020)**
$Burr$	0.5	-2	1.850	0.524	0.674	0.374	1.319	1.447
			(0.083)	(0.026)	(0.080)	(0.017)	(0.062)	(0.070)
$Burr$	0.5	-0.5	6.110	10.681	9.337	8.745	2.367	3.325
			(0.211)	(0.282)	(0.367)	(0.224)	(0.152)	(0.145)
$Frechét(2)$	0.5	-1	1.695	1.268	1.231	0.762	1.492	1.739
			(0.080)	(0.116)	(0.118)	(0.028)	(0.073)	(0.083)
$\|t_2\|$	0.5	-1	1.894	2.624	3.622	1.696	2.643	1.945
			(0.084)	(0.126)	(0.900)	(0.067)	(0.134)	(0.079)
$\log\Gamma(2,2)$	0.5	0	8.705	13.401	11.786	15.622	7.939	10.112
			(0.279)	(0.219)	(0.234)	(0.186)	(0.245)	(0.281)
Mean			**4.051**	**5.700**	**5.330**	**5.440**	**3.152**	**3.713**
			(0.076)	**(0.079)**	**(0.202)**	**(0.060)**	**(0.066)**	**(0.069)**
$Burr$	1	-2	7.152	1.881	3.503	1.368	4.745	5.200
			(0.331)	(0.096)	(0.747)	(0.061)	(0.221)	(0.241)
$Burr$	1	-0.5	25.662	46.188	39.791	31.368	10.078	13.814
			(0.915)	(1.232)	(1.229)	(0.917)	(0.569)	(0.624)
$Frechét(1)$	1	-1	7.036	5.182	6.051	3.401	5.978	1.739
			(0.342)	(0.235)	(0.516)	(0.125)	(0.296)	(0.083)
$\|t_1\|$	1	-2	6.736	2.787	3.463	1.739	7.487	7.092
			(0.316)	(0.260)	(0.347)	(0.082)	(0.483)	(0.306)
$\log\Gamma(1,2)$	1	0	33.733	54.176	46.691	58.892	31.818	38.417
			(1.039)	(0.912)	(0.924)	(0.832)	(1.036)	(1.146)
Mean			**16.064**	**22.043**	**19.900**	**19.354**	**12.021**	**14.249**
			(0.299)	**(0.315)**	**(0.364)**	**(0.250)**	**(0.266)**	**(0.280)**
Overall mean			**7.191**	**9.982**	**9.077**	**8.964**	**5.338**	**6.359**
			(0.103)	**(0.109)**	**(0.142)**	**(0.086)**	**(0.092)**	**(0.096)**

Table 5.3.7 MSEs x 1000 for estimates of the EVI for sample size $n = 5000$

The methods based on the fitting of the PPD perform significantly better than the methods based on the Hill estimator.

The following tables provide a summary of which estimation method performed best overall. The tables indicate the number of distributions for which that specific method resulted in the lowest MSE value, expressed as a percentage of the total number of distributions.

Table 5.3.8 shows the results of the four methods based on the Hill estimator and table 5.3.9 shows the results of the two methods based on the fitting of the PPD. Table 5.3.10 contains the results of all six estimation methods.

Sample size	$\hat{\gamma}_0$	$\hat{\gamma}_{GH1.5}$	$\hat{\gamma}_{GH1.25}$	$\hat{\gamma}_{DK}$
200	40%	0%	0%	60%
500	53%	0%	0%	47%
1000	53%	0%	0%	47%
2000	47%	0%	0%	53%
5000	47%	0%	0%	53%

Table 5.3.8 Percentage distributions yielding the lowest MSE: first order estimation

Considering the four methods based on the Hill estimator, the threshold selection method by Drees and Kaufmann outperforms the other methods.

Sample size	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_0^{Bayes}$
200	33%	67%
500	40%	60%
1000	60%	40%
2000	87%	13%
5000	87%	13%

Table 5.3.9 Percentage distributions yielding the lowest MSE: second order estimation

Considering the two methods based on the fitting of the PPD, the method of maximum likelihood estimation leads to superior results when compared to the method based on Bayesian methodology.

The only two exceptions are the distributions with an true EVI of 1, for sample sizes $n = 200$ and $n = 500$. The overall mean value is also the lowest for sample sizes $n = 200$ and $n = 500$ when Bayesian methodology is used. The results indicating that the estimation method based on Bayesian methodology leads to improved results when the sample size is small, is to be expected.

Sample size	$\hat{\gamma}_0$	$\hat{\gamma}_{GH1.5}$	$\hat{\gamma}_{GH1.25}$	$\hat{\gamma}_{DK}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_0^{Bayes}$
200	0%	0%	0%	33%	33%	33%
500	0%	0%	0%	47%	40%	13%
1000	0%	0%	0%	47%	47%	7%
2000	0%	0%	0%	47%	47%	7%
5000	0%	0%	0%	53%	47%	0%

Table 5.3.10 Percentage distributions yielding the lowest MSE

From these tables it is evident that the method of Drees and Kauffmann (1998) performs better for certain distributions. For example the Burr distribution with $\gamma = -2$ and $\gamma = -0.25$, irrespective of the sample size. For other distributions, like the Burr distribution with $\gamma = -0.5$ and $\gamma = -0.25$, the method of maximum likelihood is superior. This is particularly evident when there is more data available. In the situation where data is limited, Bayesian methods lead to improved estimation of the EVI.

5.4 Conclusion

In this chapter the design of the simulation study and the results obtained were presented. In section 5.2 the simulation procedure was given, together with a description of all the distributions that were considered in the simulation study. The six estimators of the EVI that are of interest were discussed, together with some computational issues, the most important being how the second order parameter is estimated.

Section 5.3 provided the results of the simulation study. In section 5.3.1 the results of the two estimators of the second order parameter were given. It was found that $\hat{\rho}_2$ performed superior to $\hat{\rho}_1$. This estimator of the second order parameter was therefore used in the simulation study carried out in section 5.3.2.

In section 5.3.2 the performance of the six estimators of the EVI were compared. It was found that the second order estimation methods performed better overall than the first order estimation methods.

Considering the second order estimation methods, the method based on Bayesian methodology performed better when the sample size was small. For large sample sizes, the method of maximum likelihood was superior.

CHAPTER 6

Empirical Bayes estimation of the EVI

6.1 Introduction

In this chapter the aim is to introduce empirical Bayes (EB) methodology. The chapter begins with a discussion of the development of the field and an overview of the EB approach.

The two categories of EB methods are discussed in section 6.3. In section 6.4 the EB methodology is described in terms of an ANOVA setting. The general ANOVA setting is explained by considering a practical situation. A first order EB estimator and a second order EB estimator are also derived.

6.2 Overview of empirical Bayes theory

Bayesian statistics takes past experiences, guesses or assumptions into account in the form of the prior distribution. In a pure Bayesian approach the distribution of some random variable θ is calculated by considering the prior distribution together with the likelihood function.

The prior distribution can also depend on some unknown parameters. These unknown parameters also have a prior density, referred to as a second-stage prior. The parameters of the prior distribution are referred to as hyperparameters. These hyperparameters are assumed to be independent from the parameters of the distribution. Different methods exist which can be used to estimate the hyperparameters.

A fully Bayesian analysis requires a prior density for the hyperparameters (Novick, Jackson, Thayer and Cole, 1972). The posterior density of the parameter vector is found by integrating out the hyperparameters. These integrals are seldom expressible in a closed form and numerical integration procedures are required.

The EB approach, in contrast, uses the data to obtain estimates of the hyperparameters. The estimated hyperparameters are then used to calculate the prior and posterior densities (Carlin and Louis, 1996).

According to Casella (1992: 108) the EB approach can be seen as a set of procedures and methods that combine the strengths of the frequentist (classical) approach with the strengths of the Bayesian approach. In other words, the best of both worlds.

The EB approach can be described by considering the following two-stage model:

Let X be a random variable with likelihood function $f(\boldsymbol{x}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of unknown parameters and $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ is a sample of n observations. Also assume that $\boldsymbol{\theta}$ has a family of prior distributions $\pi(\boldsymbol{\theta}|\boldsymbol{\eta})$, where $\boldsymbol{\eta}$ is a vector of hyperparameters. The objective is to make an inference about $\boldsymbol{\theta}$ based on the sample $\boldsymbol{x}$.

Using Bayes' Theorem, the posterior distribution is defined as:

$$\pi(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{\eta}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\eta})}{\int f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\eta})d\theta}.$$

In EB methodology, the vector of hyperparameters $\boldsymbol{\eta}$ is assumed to be unknown and is estimated by $\hat{\boldsymbol{\eta}}$. In order to obtain an estimate of $\boldsymbol{\eta}$, marginal maximum likelihood estimation can be used, where the marginal distribution of the data is denoted by:

$$m(\boldsymbol{x}|\boldsymbol{\eta}) = \int f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\eta})d\theta.$$

Inferences about $\boldsymbol{\theta}$ are then based on the estimated posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{x},\hat{\boldsymbol{\eta}})$ (Carlin and Louis, 1996).

6.3 Categories of empirical Bayes methods

EB methods can be classified into two categories, namely parametric methods and non-parametric methods. Parametric EB methods make the assumption that the prior for $\boldsymbol{\theta}$ belongs to a parametric class $\pi(\boldsymbol{\theta}|\boldsymbol{\eta})$, where only the hyperparameter $\boldsymbol{\eta}$ is unknown (Morris, 1983).

In the case of non-parametric methods, it is assumed that the unknown parameters θ_i are independent and identically distributed without assuming any specific family of distributions. The data are used directly to estimate the prior distribution (Robbins, 1983).

In this study parametric EB methods will be applied, as outlined in the next section.

6.4 Empirical Bayes in the ANOVA setting

Casella (1992: 114-116) argues that EB methods can be applied in simpler statistical testing procedures. Examples of such procedures include ANOVA and the t test. In this thesis EB methodology will specifically be applied in the context of an ANOVA setting, therefore implying two or more different populations or "treatments".

In order to derive an EB estimator of the EVI in an ANOVA setting, consider the scenario where the data of five insurance portfolios are available. Four of these five insurance portfolios consist of a large number of claims, whereas the fifth portfolio consists of a much smaller number.

The situation can be represented as follows:

	Portfolio 1	**Portfolio 2**	**Portfolio 3**	**Portfolio 4**	**Portfolio 5**
	Claim 1	Claim 1	Claim 1	Claim 1	Claim 1
	Claim 2	Claim 2	Claim 2	Claim 2	Claim 2
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	Claim n_1	Claim n_2	Claim n_3	Claim n_4	Claim n_5
EVI	$\boldsymbol{\gamma_1}$	$\boldsymbol{\gamma_2}$	$\boldsymbol{\gamma_3}$	$\boldsymbol{\gamma_4}$	$\boldsymbol{\gamma_5}$

The five portfolios are referred to as "treatments". The parameters of interest are the EVIs of the treatments. This differs from the classical ANOVA, where the parameters of interest are the means of the treatments.

The assumption will be made that some relationship exists between the treatments. However, assumptions like normality and homoscedasticity of errors, which are associated with the classic ANOVA, will not be made.

When considering the five portfolios as individual samples, the estimate of the EVI of portfolio 5 may be inaccurate, since the sample size of this portfolio is considerably smaller than the sample sizes of the other four portfolios.

The objective is to determine whether the accuracy of the estimated value of γ_5 can be improved upon by also taking the data from the other four portfolios into account. This will be done by applying EB methodology in this ANOVA-type setting. The procedure and derivations given below are those derived by De Wet and Berning (2013).

6.4.1 General setting

In an attempt to generalise the above-mentioned ANOVA setting, assume that the observations are from r heavy-tailed distributions, referred to as the treatments.

In order to obtain the EB estimator of the EVI for a specific treatment, consider the following, where x_{ij} denote the j^{th} observation of the i^{th} treatment:

Treatment	**Observations**	**EVI**
1	$x_{11}, x_{12}, \dots, x_{1n_1}$	$\gamma_1 > 0$
2	$x_{21}, x_{22}, \dots, x_{2n_2}$	$\gamma_2 > 0$
$\vdots$	$\vdots$	$\vdots$
r	$x_{r1}, x_{r2}, \dots, x_{rn_r}$	$\gamma_r > 0$

The k_i multiplicative excesses for treatment i are calculated, for all $i = 1, 2, \dots, n$.

The notation z_{ij} is used to indicate the j^{th} observed excess of the i^{th} treatment:

Treatment	Observations	EVI
1	$z_{11}, z_{12}, \dots, z_{1k_1}$	$\gamma_1 > 0$
2	$z_{21}, z_{22}, \dots, z_{2k_2}$	$\gamma_2 > 0$
$\vdots$	$\vdots$	$\vdots$
r	$z_{r1}, z_{r2}, \dots, z_{rk_r}$	$\gamma_r > 0$

The assumption is made that the thresholds t_i $(i = 1,2, \dots, r)$ that correspond to the multiplicative excesses k_i $(i = 1,2, \dots, r)$ are large enough so that the underlying distribution of the excesses can be modelled by the limiting distribution. This implies that the approximate underlying distributions of the observed excesses are from the same family of distributions. The family is the strict Pareto distribution if first order regular variation is assumed. If second order regular variation is assumed, the family of distributions is the PPD.

Let $f^*(\cdot\,|\boldsymbol{\theta}^*)$ denote the density function of the family of distributions and let the prior distribution of the parameter vector $\boldsymbol{\theta}^*$ be denoted by $g^*(\boldsymbol{\theta}^*|\boldsymbol{\eta}^*)$, where $\boldsymbol{\eta}^*$ is the vector of hyperparameters.

It follows that:

Treatment	Excesses	EVI	Density	Prior
1	$z_{11}, z_{12}, \dots, z_{1k_1}$	$\gamma_1 > 0$	$f^*(\cdot\,|\boldsymbol{\theta}_1^*)$	$g^*(\cdot\,|\boldsymbol{\eta}^*)$
2	$z_{21}, z_{22}, \dots, z_{2k_2}$	$\gamma_2 > 0$	$f^*(\cdot\,|\boldsymbol{\theta}_2^*)$	$g^*(\cdot\,|\boldsymbol{\eta}^*)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
r	$z_{r1}, z_{r2}, \dots, z_{rk_r}$	$\gamma_r > 0$	$f^*(\cdot\,|\boldsymbol{\theta}_r^*)$	$g^*(\cdot\,|\boldsymbol{\eta}^*)$

Since it is more convenient to use $\alpha_i = 1/\gamma_i$, this adjustment is made and the parameter of interest now becomes α_i (instead of γ_i). It is also easier to work with the logs of the multiplicative excesses, denoted by $y_{ij} = \log(z_{ij})$.

The table is therefore adjusted as follows:

Treatment	Log of excesses	EVI	Density	Prior
1	$y_{11}, y_{12}, \dots, y_{1k_1}$	$\gamma_1 = \frac{1}{\alpha_1} > 0$	$f(\cdot\,\|\boldsymbol{\theta}_1)$	$g(\cdot\,\|\boldsymbol{\eta})$
2	$y_{21}, y_{22}, \dots, y_{2k_2}$	$\gamma_2 = \frac{1}{\alpha_2} > 0$	$f(\cdot\,\|\boldsymbol{\theta}_2)$	$g(\cdot\,\|\boldsymbol{\eta})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
r	$y_{r1}, y_{r2}, \dots, y_{rk_r}$	$\gamma_r = \frac{1}{\alpha_r} > 0$	$f(\cdot\,\|\boldsymbol{\theta}_r)$	$g(\cdot\,\|\boldsymbol{\eta})$

where $f(\cdot\,|\boldsymbol{\theta})$ denote the density function of the log of the multiplicative excesses and $g(\boldsymbol{\theta}|\boldsymbol{\eta})$ the prior distribution of the parameter vector $\boldsymbol{\theta}$.

6.4.2 First order EB estimation of the EVI

A first order EB estimator of the EVI can be derived by assuming first order regular variation. When first order regular variation is assumed, the distribution of the excesses tends to a strict Pareto distribution as the threshold becomes large. Considering the setting described in the previous section, it follows that for treatment i

$$\bar{F}^*_{t_i}(z|\alpha_i) \approx z^{-\alpha_i}$$

for sufficiently large t_i, where t_i denotes the threshold associated with treatment i.

From this point onwards, it will be assumed that the conditions for a reasonable approximation hold. The above expression will therefore be written as

$$\bar{F}^*_{t_i}(z|\alpha_i) = z^{-\alpha_i}.$$

By applying the log transformation to the excesses, that is $Y = \log(Z)$, it can be shown that the distribution of Y is approximately exponential. This is done as follows:

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(\log Z \le y) \\ &= P(Z \le e^y) \\ &= F_Z(e^y). \end{aligned}$$

Since $\bar{F}_Z(z) = z^{-\alpha}$, it follows that $\bar{F}_Y(y) = e^{-\alpha y}$ and therefore:

$$\begin{aligned} F_{t_i}(y|\alpha_i) &= 1 - \bar{F}_{t_i}(y|\alpha_i) \\ &= 1 - e^{-\alpha_i y}. \end{aligned}$$

In summary, the assumed setting is:

Treatment	**Log of excesses**	**EVI**	**Density**	**Prior**
1	$y_{11}, y_{12}, \dots, y_{1k_1}$	$\gamma_1 = \frac{1}{\alpha_1} > 0$	$f(\cdot\,|\alpha_1)$	$g(\cdot\,|\boldsymbol{\eta})$
2	$y_{21}, y_{22}, \dots, y_{2k_2}$	$\gamma_2 = \frac{1}{\alpha_2} > 0$	$f(\cdot\,|\alpha_2)$	$g(\cdot\,|\boldsymbol{\eta})$
⋮	⋮	⋮	⋮	⋮
r	$y_{r1}, y_{r2}, \dots, y_{rk_r}$	$\gamma_r = \frac{1}{\alpha_r} > 0$	$f(\cdot\,|\alpha_r)$	$g(\cdot\,|\boldsymbol{\eta})$

where $f(\cdot\,|\alpha_i)$ is the exponential density with parameter $\alpha_i > 0$.

In order to obtain the first order EB estimator, the following steps are carried out (De Wet & Berning, 2013):

Step 1: Choice of prior

Objective priors, such as Zellner's MDI prior, cannot be applied in an EB framework. This is because these priors do not have hyperparameters. Since EB methods fundamentally involve estimating the hyperparameters, objective priors cannot be used.

An alternative choice will be to consider a conjugate prior, chosen for mathematical convenience. For the exponential distribution the conjugate prior is the gamma distribution. The chosen prior $\boldsymbol{g}(\cdot\,|\boldsymbol{\eta})$, and therefore also the resulting posterior $\boldsymbol{g}(\cdot\,|\boldsymbol{\eta}, \boldsymbol{y})$, will belong to the gamma family. This is confirmed by the following theorem:

Theorem 6.4.1: Posterior distribution of exponentially distributed variables

Let $X_1, X_2, \dots, X_n$ be independent and identically distributed random variables of a sample from an exponential distribution with unknown parameter $\lambda > 0$. Also assume that the parameter λ have a gamma prior distribution with parameters $\alpha > 0$ and $\beta > 0$. It follows that, given $X_i = x_i$ for $i = 1,2, \dots, n$, the posterior distribution of λ is gamma with parameters $\alpha + n$ and $\beta + \sum_{i=1}^{n} x_i$.

Proof

Let $f(x) = \lambda \exp(-\lambda x)$ be the underlying density function of the random variables $X_1, X_2, \dots, X_n$. Since these random variables are independent, the likelihood function of λ is given by

$$L(\lambda|x_1, \dots, x_n) = \prod_{i=1}^{n} \lambda \exp(-\lambda x_i) = \lambda^n \left(-\lambda \sum_{i=1}^{n} x_i\right).$$

Let the gamma prior density of the parameter λ be given by

$$g(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta\lambda).$$

According to Bayes' Theorem, the posterior density is proportional to the product of the prior and the likelihood. Therefore the posterior density of λ is given by

$$\begin{aligned} g(\lambda|x_1, \dots, x_n) &\propto g(\lambda) L(\lambda|x_1, \dots, x_n) \\ &\propto \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta\lambda)\, \lambda^n \left(-\lambda \sum_{i=1}^{n} x_i\right) \\ &\propto \lambda^{\alpha+n-1} \exp\left(-\lambda\left(\beta + \sum_{i=1}^{n} x_i\right)\right). \end{aligned}$$

It follows that the posterior density of λ is proportional to a gamma density with parameters $\alpha + n$ and $\beta + \sum_{i=1}^{n} x_i$. ■

In order to simplify the method of obtaining the first order EB estimator, the two parameters of the gamma distribution are reduced to only one parameter. This is done by considering the chi-square distribution, which is a special case of the gamma distribution. A chi-square distribution with $v > 0$ degrees of freedom is the same as a gamma distribution with parameters $v/2$ and $1/2$ (Wackerly, Mendenhall and Scheaffer, 2008: 187).

The prior that will be used in the calculation of the EB estimate is a chi-square prior, given by

$$g(\alpha|v) = \frac{1}{2^{v/2}\Gamma(v/2)} \alpha^{v/2-1} e^{-\alpha/2}$$

where $v > 0$ denotes the hyperparameter.

Step 2: Obtaining the posterior distribution

Applying Theorem 6.4.1, it follows that for treatment i

$$\alpha_i | y_{i1}, y_{i2}, \dots, y_{ik_i} \sim gamma\left(\frac{v}{2} + k_i,\ \frac{1}{2} + \sum_{j=1}^{k_i} y_{ij}\right).$$

Step 3: Obtaining the Bayes estimate of the parameter for treatment i

Assuming squared error loss, the Bayes estimate of a parameter is given by the expected value of the parameter, with respect to its posterior distribution. Considering the gamma distribution, the expected value is given by the ratio of its two parameters.

The Bayes estimate of α_i is given by:

$$\hat{\alpha}_i = \frac{\frac{v}{2} + k_i}{\frac{1}{2} + \sum_{j=1}^{k_i} y_{ij}} = \frac{v + 2k_i}{1 + 2\sum_{j=1}^{k_i} y_{ij}}.$$

The next steps involve estimating the unknown hyperparameter v.

Step 4: Determining the unconditional distribution of the log multiplicative excesses

In order to estimate the unknown parameter v, it is necessary to determine the unconditional distribution of the log multiplicative excesses Y.

Previously it was shown that an exponential distribution can be assumed:

$$f(y|\alpha_i) = \alpha_i \exp(-\alpha_i y)$$

for treatment i.

In order to determine the density $f(y)$, it is necessary to integrate out the parameter α_i:

$$
\begin{aligned}
f(y) &= \int_0^\infty g(u) f(y|u) du \\
&= \int_0^\infty \frac{1}{2^{v/2}\Gamma(v/2)} u^{v/2-1} e^{-u/2}\, u e^{-uy} du \\
&= \frac{1}{2^{v/2}\Gamma(v/2)} \int_0^\infty u^{v/2+1-1} e^{-u(x+1/2)} du.
\end{aligned}
$$

The above integral is related to a gamma density. It follows directly that:

$$
\begin{aligned}
f(y) &= \frac{1}{2^{v/2}\Gamma(v/2)} \left(\frac{\Gamma(v/2+1)}{(y+1/2)^{v/2+1}} \right) \\
&= \frac{v/2}{2^{v/2}(y+1/2)^{v/2+1}} \\
&= \frac{v/2}{2^{v/2}(1+2y)^{v/2+1}} 2^{v/2+1} \\
&= v(1+2y)^{-(v/2+1)}.
\end{aligned}
$$

The unconditional distribution of Y is therefore the generalised Pareto distribution, expressed as:

$$
Y \sim GP\left(\frac{2}{v}, \frac{1}{v}\right).
$$

Step 5: Estimating the hyperparameter

The density $f(y) = v(1+2y)^{-(v/2+1)}$ is not conditional on the treatment parameter. This implies that all the data across all the treatments can be used to estimate the hyperparameter v. The method of maximum likelihood is used to estimate v and the process of obtaining the estimate is explained below.

Let the likelihood function of the hyperparameter be given by:

$$
L(v) = \prod_{i=1}^{r} \prod_{j=1}^{k_i} v(1+2y_{ij})^{-(v/2+1)}.
$$

The log likelihood function is:

$$l(v) = k_{Tot}\log(v) - (v/2+1)\sum_{i=1}^{r}\sum_{j=1}^{k_i}\log(1+2y_{ij})$$

where $k_{Tot} = \sum_{i=1}^{r} k_i$, the total number of excesses across all treatments.

By differentiating the log likelihood with respect to v and setting it equal to zero, the value of v where $l(v)$ is a maximum, is obtained:

$$\frac{\partial l(v)}{\partial v} = k_{Tot}\left(\frac{1}{v}\right) - \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{k_i}\log\left(1+2y_{ij}\right).$$

Setting $\frac{\partial l(v)}{\partial v} = 0$:

$$\frac{k_{Tot}}{\hat{v}} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{k_i}\log\left(1+2y_{ij}\right)}{2}$$

$$\hat{v} = \frac{2k_{Tot}}{\sum_{i=1}^{r}\sum_{j=1}^{k_i}\log\left(1+2y_{ij}\right)}.$$

Therefore the above expression states the MLE of the parameter v.

Step 6: EB estimation of the treatment parameter

In order to obtain the EB estimator of the treatment parameter, the estimated value $\hat{v}$ obtained in Step 5 is substituted into the estimator obtained in Step 3:

$$\hat{\alpha}_i^{EB} = \frac{\hat{v} + 2k_i}{1 + 2\sum_{j=1}^{k_i} y_{ij}} \qquad (1)$$

for $i = 1,2,\ldots,r$.

Step 7: EB estimates of the EVIs

The EB estimate of the EVI for treatment i is then given by

$$\hat{\gamma}_i^{EB} = \frac{1}{\hat{\alpha}_i^{EB}} \tag{2}$$

for $i = 1,2, \ldots, r$.

In the situation where there is little or no data available, that is $k_i \rightarrow 0$ for $i = 1,2, \ldots, r$, equation (1) indicates that $\hat{\alpha}_i^{EB} \rightarrow \hat{v}$. Since the prior of α_i is gamma with parameters $v/2$ and $1/2$, the prior mean is v. This implies that if less data is available, the estimate shrinks towards the estimated prior mean v.

Equation (1) can be rewritten as:

$$\hat{\alpha}_i^{EB} = \frac{\frac{\hat{v}}{2k_i} + 1}{\frac{1}{2k_i} + \bar{x}_i} \tag{3}$$

where $\bar{x}_i = \frac{1}{k_i}\sum_{j=1}^{k_i} y_{ij}$, the mean of the log multiplicative excesses for treatment i. It is also true that $\bar{x}_i$ is the Hill estimate of the EVI for treatment i.

In the situation where there is a large amount of data available, that is $k_i \rightarrow \infty$ for $i = 1,2, \ldots, r$, equation (3) indicates that $\hat{\alpha}_i^{EB} \approx 1/\bar{x}_i$, which is the MLE of the parameter α_i of the exponential distribution. Using equation (2), it also follows that $\hat{\gamma}_i^{EB} = 1/\hat{\alpha}_i^{EB} \approx \bar{x}_i$ for large k_i.

This result shows that in the case where little or no data is available, the estimator of the EVI is dominated by the prior Bayesian assumptions made. If more data becomes available, the estimated tends to a frequentist estimator, namely the MLE of the parameter, based on only the data for that treatment.

To summarise, a first order EB estimate of the EVI for treatment i is given by:

$$\hat{\gamma}_i^{EB} = \frac{1}{\hat{\alpha}_i^{EB}} = \frac{1 + 2\sum_{j=1}^{k_i} y_{ij}}{\hat{v} + 2k_i},$$

where $y_{i1}, y_{i2}, \ldots, y_{ik_i}$ are the logs of the k_i multiplicative excesses of treatment i. Also the hyperparameter v is estimated by:

$$\hat{v} = \frac{2\sum_{i=1}^{r} k_i}{\sum_{i=1}^{r}\sum_{j=1}^{k_i} \log(1 + 2y_{ij})}$$

where $y_{i1}, y_{i2}, \ldots, y_{ik_i}$ are the logs of the multiplicative excesses of treatment i and r is the total number of treatments.

6.4.3 Second order EB estimation of the EVI

Consider again the setting described in section 6.4.1:

Treatment	Excesses	EVI	Density	Prior
1	$z_{11}, z_{12}, \ldots, z_{1k_1}$	$\gamma_1 > 0$	$f^*(\cdot\,\|\boldsymbol{\theta}_1^*)$	$g^*(\cdot\,\|\boldsymbol{\eta}^*)$
2	$z_{21}, z_{22}, \ldots, z_{2k_2}$	$\gamma_2 > 0$	$f^*(\cdot\,\|\boldsymbol{\theta}_2^*)$	$g^*(\cdot\,\|\boldsymbol{\eta}^*)$
⋮	⋮	⋮	⋮	⋮
r	$z_{r1}, z_{r2}, \ldots, z_{rk_r}$	$\gamma_r > 0$	$f^*(\cdot\,\|\boldsymbol{\theta}_r^*)$	$g^*(\cdot\,\|\boldsymbol{\eta}^*)$

A second order EB estimator of the EVI can be derived by assuming second order regular variation. When assuming second order regular variation, the distribution of the excesses tends to a PPD as the threshold becomes large. Therefore, for treatment i:

$$\bar{F}_{t_i}^*(z|\alpha_i, c_i, \rho_i) \approx (1 - c_i)z^{-\alpha_i} + c_i z^{-(1-\rho_i)\alpha_i}$$

for sufficiently large t_i, where t_i denotes the threshold associated with treatment i.

As mentioned in chapter 3, the second order parameter ρ will estimated externally. Therefore ρ is replaced by the estimated value $\hat{\rho}$ in the above equation:

$$\bar{F}_{t_i}^*(z|\alpha_i, c_i) \approx (1 - c_i)z^{-\alpha_i} + c_i z^{-(1-\hat{\rho})\alpha_i}.$$

The log transformation is applied to the variable. It follows that, for $Y = \log(Z)$:

$$\bar{F}_{t_i}(y|\alpha_i, c_i) \approx (1 - c_i)e^{-\alpha_i y} + c_i e^{-(1-\hat{\rho})\alpha_i y}.$$

By differentiating the above equation with respect to y, the density function of Y is obtained:

$$f_{t_i}(y|\alpha_i, c_i) \approx \alpha_i(1 - c_i)e^{-\alpha_i y} + \alpha_i c_i(1 - \hat{\rho})e^{-\alpha_i(1-\hat{\rho})y}.$$

The above result can be considered as a "perturbed exponential" distribution and consists of two exponential parts, one with parameter $\alpha_i > 0$ and the other with parameter $\alpha_i(1 - \hat{\rho}) > 0$.

This implies:

Treatment	**Log of excesses**	**EVI**	**Density**	**Prior**
1	$y_{11}, y_{12}, \dots, y_{1k_1}$	$\gamma_1 = \frac{1}{\alpha_1} > 0$	$f(\cdot\,|\alpha_1, c_1)$	$g(\cdot\,|\boldsymbol{\eta})$
2	$y_{21}, y_{22}, \dots, y_{2k_2}$	$\gamma_2 = \frac{1}{\alpha_2} > 0$	$f(\cdot\,|\alpha_2, c_2)$	$g(\cdot\,|\boldsymbol{\eta})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
r	$y_{r1}, y_{r2}, \dots, y_{rk_r}$	$\gamma_r = \frac{1}{\alpha_r} > 0$	$f(\cdot\,|\alpha_r, c_r)$	$g(\cdot\,|\boldsymbol{\eta})$

In order to obtain a second order EB estimator, the same steps followed in section 6.4.2 are carried out (De Wet & Berning, 2013):

Step 1: Choice of prior

In this step the results obtained in section 6.4.2 are used. The following assumptions are made in order to simplify the mathematical calculations and for the sake of convenience.

It is assumed that the prior on the parameter α is independent of the prior on the parameter c. The prior used for α in section 6.4.2, namely the chi-square distribution with v degrees of freedom, will again be considered:

$$g_\alpha(\alpha|v) = \frac{1}{2^{v/2}\Gamma(v/2)}\alpha^{v/2-1}e^{-\alpha/2}$$

where $v > 0$ denotes the hyperparameter.

The same restriction as earlier is placed on the parameter c:

$$\frac{1}{\rho} \leq c \leq 0.5.$$

The simplest prior that can be chosen is to choose c uniform on $\left(\frac{1}{\rho}, 0.5\right)$. The prior of c is therefore chosen as:

$$\begin{aligned} g_c(c) &= \frac{1}{0.5 - \frac{1}{\hat{\rho}}} \\ &= \frac{-\hat{\rho}}{1 - 0.5\hat{\rho}} \end{aligned}$$

where $\hat{\rho}$ is the estimate of ρ.

By combining the two expressions given above, the prior that will be used in the calculation of the second order EB estimate is obtained:

$$g(\alpha, c|v) = g_\alpha(\alpha|v) g_c(c) = \frac{-\hat{\rho}\alpha^{v/2-1}e^{-\alpha/2}}{(1 - 0.5\hat{\rho})2^{v/2}\Gamma(v/2)}.$$

Step 2: Obtaining the posterior distribution of the parameter vector

Considering once again Bayes' Theorem, the posterior density of the parameter vector is given by:

$$\begin{aligned} h(\alpha_i, c_i | v, y_{i1}, y_{i2}, \dots, y_{ik_i}) &\propto g(\alpha_i, c_i|v) \prod_{j=1}^{k_i} f_{t_i}(y_{ij}|\alpha_i, c_i) \\ &= \frac{-\hat{\rho}\alpha_i{}^{v/2-1}e^{-\alpha_i/2}}{(0.5 - \hat{\rho})2^{v/2}\Gamma(v/2)} \prod_{j=1}^{k_i} \left\{\alpha_i(1 - c_i)e^{-\alpha_i y_{ij}} + \alpha_i c_i (1 - \hat{\rho}) e^{-\alpha_i(1-\hat{\rho})y_{ij}}\right\} \\ &\propto \alpha_i{}^{k_i + v/2 - 1} e^{-\alpha_i/2} \prod_{j=1}^{k_i} \left\{(1 - c_i)e^{-\alpha_i y_{ij}} + c_i(1 - \hat{\rho})e^{-\alpha_i(1-\hat{\rho})y_{ij}}\right\}. \end{aligned}$$

It is not possible to recognise the distribution as one of the well-known families of distributions.

Step 3: Estimating the hyperparameter

Since the unconditional distribution of Y cannot be determined in closed form, the estimate of the hyperparameter derived in section 6.4.2 will be used:

$$\hat{v} = \frac{2\sum_{i=1}^{r} k_i}{\sum_{i=1}^{r}\sum_{j=1}^{k_i}\log(1+2y_{ij})}.$$

Step 4: EB estimation of the treatment parameters

In order to obtain the EB estimator of the treatment parameters, the estimate $\hat{v}$ is substituted into the expression for the posterior density obtained in Step 2. The Gibbs sampler is applied to the resulting posterior, as described in section 4.8, to obtain the EB estimates $\hat{\alpha}_i^{EB}$ and $\hat{c}_i^{EB}$.

Step 5: EB estimates of the EVIs

The EB estimate of the EVI for treatment i is then given by

$$\hat{\gamma}_i^{EB} = \frac{1}{\hat{\alpha}_i^{EB}}$$

for $i = 1,2,\dots,r$.

In summary, a second order EB estimate of the EVI for treatment i is given by:

$$\hat{\gamma}_i^{EB} = \frac{1}{\hat{\alpha}_i^{EB}}.$$

The hyperparameter v is estimated by:

$$\hat{v} = \frac{2\sum_{i=1}^{r} k_i}{\sum_{i=1}^{r}\sum_{j=1}^{k_i}\log(1+2y_{ij})}$$

where $y_{i1}, y_{i2}, \dots, y_{ik_i}$ are the logs of the multiplicative excesses of treatment i and r is the total number of treatments.

With the aim of obtaining the EB estimate $\hat{\alpha}_i^{EB}$, the Gibbs sampler is applied to the following posterior:

$$h\left(\alpha_i, c_i \middle| \hat{v}, \hat{\rho}, y_{i1}, y_{i2}, \dots, y_{ik_i}\right) \propto \alpha_i{}^{k_i+\hat{v}/2-1} e^{-\alpha_i/2} \prod_{j=1}^{k_i} \left\{(1-c_i)e^{-\alpha_i y_{ij}} + c_i(1-\hat{\rho})e^{-\alpha_i(1-\hat{\rho})y_{ij}}\right\}$$

where $y_{i1}, y_{i2}, \dots, y_{ik_i}$ are the logs of the k_i multiplicative excesses of treatment i.

For computational purposes, it is easier to parameterise in terms of γ_i than α_i. This leads to

$$h\left(\gamma_i, c_i \middle| \hat{v}, \hat{\rho}, y_{i1}, y_{i2}, \dots, y_{ik_i}\right) \propto \gamma_i{}^{\left(1-k_i-\frac{\hat{v}}{2}\right)} e^{-1/2\gamma_i} \prod_{j=1}^{k_i} \left\{(1-c_i)e^{-\frac{y_{ij}}{\gamma_i}} + c_i(1-\hat{\rho})e^{-\frac{(1-\hat{\rho})y_{ij}}{\gamma_i}}\right\}.$$

The log of the above expression is

$$\left(1-k_i-\frac{\hat{v}}{2}\right)\log(\gamma_i) - \frac{1}{2\gamma_i} - \frac{1}{\gamma_i}\sum_{j=1}^{k_i} y_{ij} + \sum_{j=1}^{k_i} \log\left(1-c_i+c_i(1-\hat{\rho})e^{\frac{\hat{\rho} y_{ij}}{\gamma_i}}\right).$$

6.5 Conclusion

The focus of this chapter was to derive an EB estimate of the EVI. In section 6.2 an overview of EB theory was given. The concept of a hyperparameter was also introduced in this section. In section 6.3 the two categories of EB methods were discussed, namely parametric EB methods and non-parametric EB methods.

In section 6.4 the EB methodology in an ANOVA setting was examined. The section began with a description of the application of EB methods in the general ANOVA-type setting, using an example of five insurance portfolios. In the next subsection the steps were given in order to obtain a first order EB estimate of the EVI. In the last subsection a second order EB estimate of the EVI was derived. This estimate can be obtained by using Gibbs sampling.

CHAPTER 7

Case study

7.1 Introduction

In this chapter the data of a specific case study will be analysed, using the techniques discussed in the previous chapters. The objective is to determine how the different estimation methods compare. More specifically, the goal is to determine how the empirical Bayes estimation method performs in practice and how its performance compares to traditional estimation methods.

The case study data consist of the claim sizes associated with five insurance portfolios (Berning, 2010). A general description of the data is given in section 7.2. Section 7.3 discusses general aspects of the estimation procedure that will be followed in order to obtain the different estimates of the EVI. Section 7.4 and section 7.5 illustrate how inference is done in practice. In section 7.4 the focus is on estimating the EVI of one insurance portfolio, using first order estimation methods. This section also includes a discussion of the results obtained.

Section 7.5 gives the procedure and results of estimating the EVI of one insurance portfolio, using second order estimation methods.

7.2 Description of the data

The insurance claims data was obtained from a South African short term insurer. The name of the insurer is not disclosed for confidentiality reasons. The data set contains the sizes of claims associated with five insurance portfolios. The risks of the portfolios (for example automobile, household items, etc.) are unknown and may differ from one portfolio to the next.

The claim dates are from 1 July 2004 to 21 July 2006 for portfolios 1, 2 and 3. For portfolios 4 and 5 the claim dates are from 31 December 2001 to 24 February 2004. The dates specified are the dates on which the claims occurred, and not the dates on which the claims were registered.

The claim amounts reported were the total claim amounts and do not include any excesses paid by the client. The amounts were also adjusted for inflation, by using July 2006 as base month.

The data set contains no zero or negative claim amounts.

Table 7.2.1 gives the sample sizes of the five insurance portfolios and able 7.2.2 gives a summary of the ten largest claims in each portfolio.

Portfolio	Sample size (Number of claims)
1	16 197
2	16 104
3	15 990
4	18 198
5	14 917

Table 7.2.1: Sample sizes of five insurance portfolios

		Portfolio				
		1	2	3	4	5
Largest claims (Rand)	1	602 768	835 568	1 435 283	12 691 360	9 770 438
	2	321 815	819 669	909 032	6 938 160	6 192 213
	3	299 165	732 276	411 798	5 488 720	5 149 854
	4	297 013	411 387	394 108	4 346 204	4 144 118
	5	287 167	378 205	385 209	4 058 660	3 443 016
	6	269 005	373 408	313 713	3 988 139	3 130 848
	7	261 362	359 964	305 713	3 833 309	3 046 545
	8	257 904	353 241	304 735	3 707 895	2 735 170
	9	257 147	335 234	299 035	3 419 720	2 728 505
	10	251 614	328 630	293 162	3 088 995	2 489 547

Table 7.2.2: Ten largest claims per portfolio

Considering the ten largest claims per portfolio, it is evident that the claims differ greatly in size. The occurrence of extremely large observations (claims) clearly illustrates the heavy-tail effect.

7.3 General aspects of estimating the EVI

The four estimates of the EVI that will be considered in the following sections are:

1. Hill estimate,
2. First order EB estimate,
3. Bayesian (PPD) estimate,
4. Second order EB estimate.

The performance of these four estimates will be assessed by comparing their respective MSE values. In order to calculate the MSE values, it is necessary to obtain an accurate estimate of the EVI for every portfolio. This estimate will then be used as the "true" value of the EVI when calculating the MSE.

All the portfolios have between $n = 10\,000$ and $n = 20\,000$ observations. It is safe to assume that the estimate of the EVI obtained for a specific portfolio will be fairly accurate, when using all the data of that portfolio. The estimated EVI of every portfolio will be referred to as the benchmark estimator.

In order to get a feel of the data, the Hill plot and the Pareto plot are drawn for each portfolio.

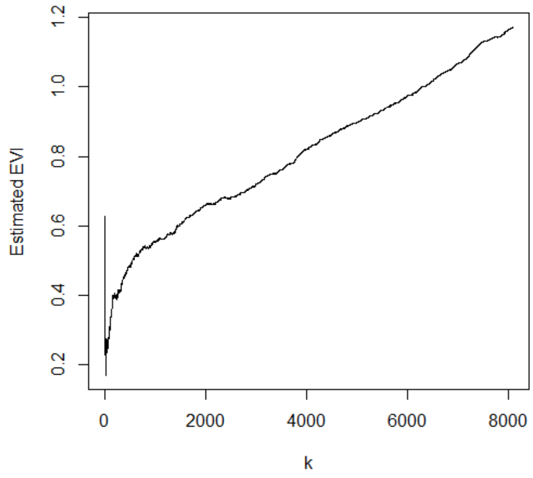
The Hill plot is obtained by plotting the Hill estimates $(H_{k,n})$ against the number of excesses $(1 \leq k \leq n/2)$.

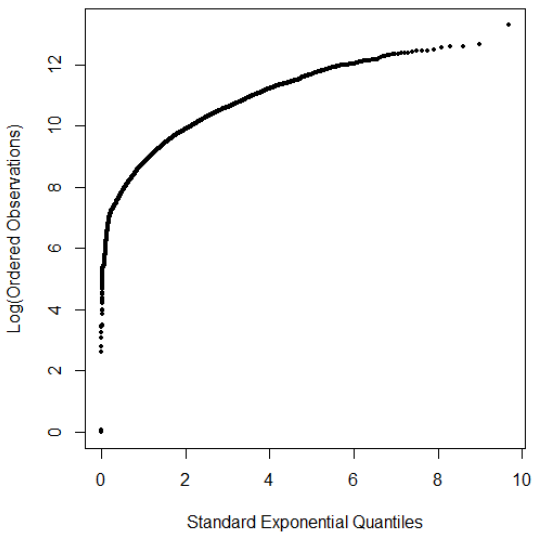
The Pareto plot is a plot of the log of the ordered observations against the standard exponential quantiles, as defined below.

Definition 7.1: Pareto plot

Given observations $x_1, x_2, \ldots, x_n$, a plot of $\left(-\log\left(1 - \frac{i}{n+1}\right); \log(x_{i,n})\right), i = 1,2,\ldots,n$, 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.

The Hill plot and the Pareto plot of the five portfolios are shown below.

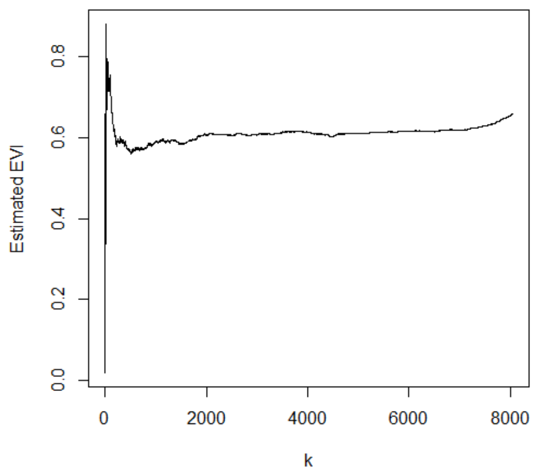


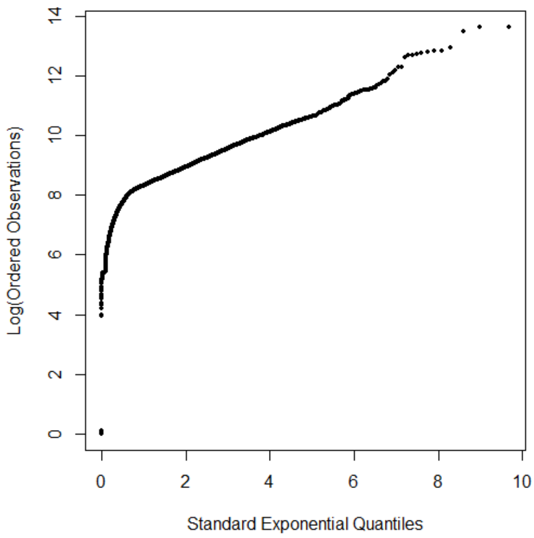


Source: Berning (2010)

Figure 7.3.1: Hill plot of portfolio 1 ***Figure 7.3.2: Pareto plot of portfolio 1***

For portfolio 1, the Hill plot indicates that the Hill estimates increase as a function of k. Considering the Pareto plot, there is no point on the graph where the Pareto plot becomes linear. Therefore, the choice of threshold is fundamental.

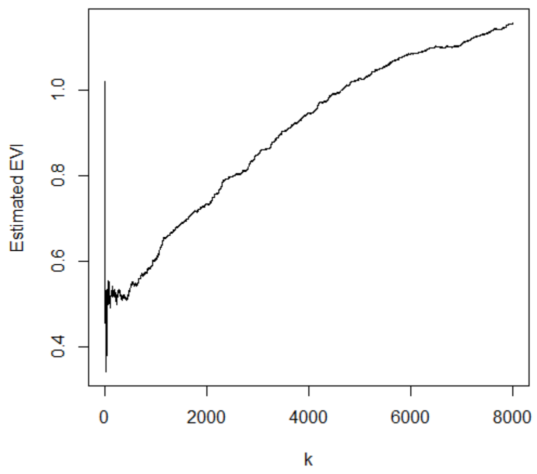


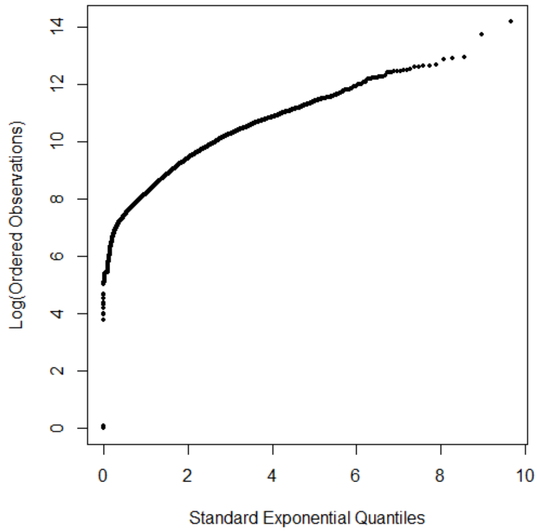


Source: Berning (2010)

Figure 7.3.3.: Hill plot of portfolio 2 ***Figure 7.3.4: Pareto plot of portfolio 2***

The Hill plot for portfolio 2 indicates that the estimated EVI values are stable at around 0.6. The Pareto plot is also linear over a large region of the graph. This is regarded as "good" behaviour. Therefore, it will be possible to obtain reliable estimates of the EVI.

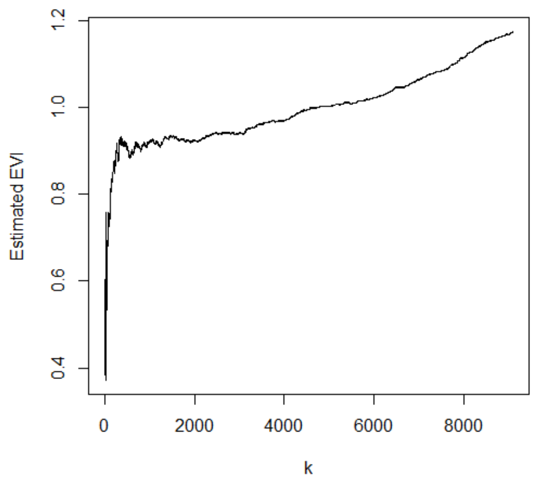


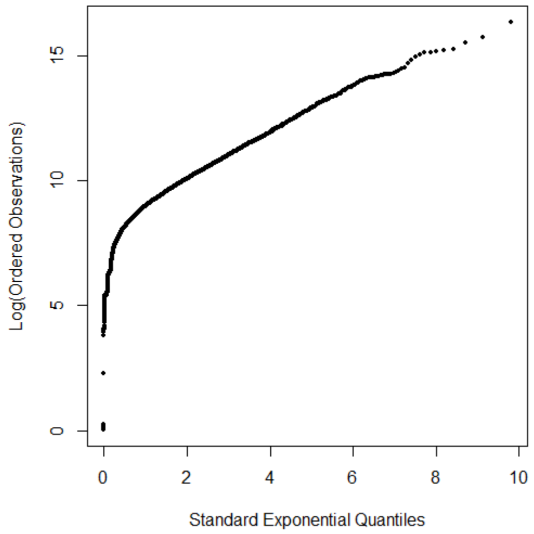


Source: Berning (2010)

Figure 7.3.5: Hill plot of portfolio 3 ***Figure 7.3.6: Pareto plot of portfolio 3***

Considering only the left-hand side of the Hill plot, it seems as if the estimates are relatively constant around 0.52 for portfolio 3. This might prove to be a reliable estimate of the EVI. However, the Pareto plot does not become linear at any point.

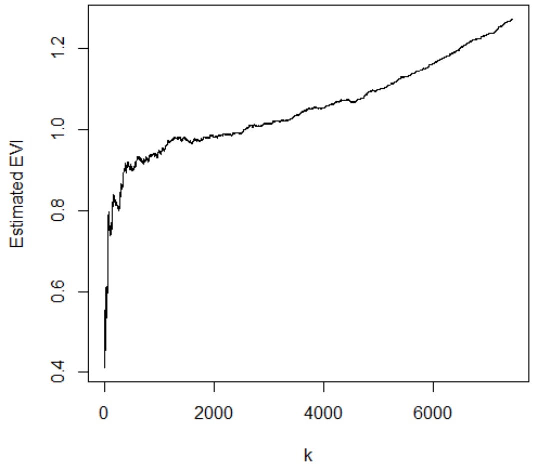


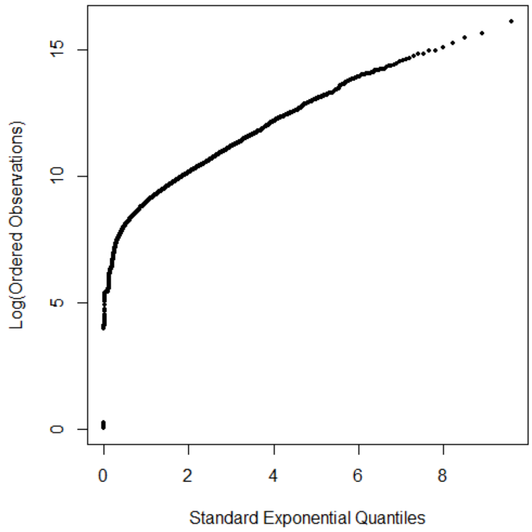


Source: Berning (2010)

Figure 7.3.7: Hill plot of portfolio 4 ***Figure 7.3.8: Pareto plot of portfolio 4***

For portfolio 4, the Hill plot indicates that 0.91 will be a relative good estimate of the EVI. The Pareto plot is linear over a wide range. This implies that a reliable estimate of the EVI can be obtained from the data of portfolio 4.





Source: Berning (2010)

Figure 7.3.9: Hill plot of portfolio 5 ***Figure 7.3.10: Pareto plot of portfolio 5***

The Hill plot for portfolio 5 indicates that the Hill estimates increase as a function of k. Since there is no stable region, this is an indication that the EVI will strongly depend on the choice of threshold. Considering the top right-hand side of the Pareto plot, the plot indicates a straight line.

In order to obtain a benchmark estimate that is more accurate, a second order estimate of the EVI is obtained. This is done by fitting the PPD to the relative excesses using MLE. The second order parameter is externally estimated by $\hat{\rho}_2$ and the number of excesses used is fixed at $k = 2n^{2/3}$. This choice of threshold was justified in section 3.3.6.

The resulting benchmark estimators are given below.

Portfolio	Benchmark estimator
1	0.4236
2	0.5858
3	0.4959
4	0.8703
5	0.8346

Table 7.3.1: Benchmark estimators of five portfolios

These benchmark estimators will be used as “true” values of the EVI when calculating the MSE in the following sections.

7.4 First order estimation of the EVI

Section 7.4.1 presents the procedure that will be followed in order to make inferences of a specific insurance portfolio. The results are shown in section 7.4.2.

7.4.1 Procedure of estimation

The following procedure is carried out in order to compare the first order estimates of the EVI:

1. Choose a target portfolio for which inferences will be made.
2. Choose a sample size n. Sample sizes of $n = 100$, $n = 200$, $n = 500$ and $n = 1000$ will be used in this thesis.
3. Take a bootstrap sample (without replacement) of size n from the target portfolio.
4. Calculate the Hill estimate of the bootstrap sample, denoted by $\hat{\gamma}_0$. The number of excesses is fixed at $k = 2\sqrt{n}$. This choice of threshold corresponds to the threshold proposed by Drees and Kauffmann (1998), which was used previously to obtain an initial estimate of the EVI (see section 3.3.2.2).
5. Calculate the first order EB estimate of the bootstrap sample, denoted by $\hat{\gamma}_{EB1}$. This is done by including all the data of the other four portfolios in the calculation. Considering the target portfolio, the number of excesses is fixed at $k = 2\sqrt{n}$. Considering the other four portfolios, the number of excesses is fixed at $k = 2\sqrt{n_i}$, where n_i is the total sample size of portfolio $i, i = 1, 2, 3, 4$.
6. Calculate the square error of estimation $(\hat{\gamma} - \gamma)^2$, where γ is the calculated benchmark estimator.
7. Repeat steps 3 – 6 a large number of times. In this thesis these steps will be repeated 10 000 times.
8. Calculate the MSE.

7.4.2 Results: first order estimation of the EVI

The following tables show the calculated MSE values of the two estimators of the EVI. The standard error of a specific MSE value is shown below the corresponding MSE value in brackets. The MSE values and their corresponding standard errors are multiplied with a factor of 1000 to ease comparison.

$n = 100$			$n = 200$		
Portfolio	$\hat{\gamma}_0$	$\hat{\gamma}_{EB1}$	**Portfolio**	$\hat{\gamma}_0$	$\hat{\gamma}_{EB1}$
1	133.0122	115.0658	1	74.0219	66.3955
	(0.3510)	(0.2998)		(0.1838)	(0.1631)
2	18.9848	16.1025	2	13.0786	11.5539
	(0.0935)	(0.0800)		(0.0662)	(0.0555)
3	175.9731	147.9303	3	93.1430	81.3288
	(0.4699)	(0.3991)		(0.2437)	(0.2200)
4	51.9354	39.2768	4	33.7445	27.7258
	(0.2626)	(0.1947)		(0.1695)	(0.1371)
5	80.5218	59.5647	5	52.9261	41.3696
	(0.3674)	(0.2762)		(0.2391)	(0.1907)
Mean	**92.0855**	**75.5880**	**Mean**	**53.3828**	**45.6747**
	(0.1492)	**(0.1216)**		**(0.0857)**	**(0.0730)**

Table 7.4.1: MSEs x 1000 for EVI estimates for $n = 100$ and $n = 200$: first order estimation

$n = 500$			$n = 1000$		
Portfolio	$\hat{\gamma}_0$	$\hat{\gamma}_{EB1}$	**Portfolio**	$\hat{\gamma}_0$	$\hat{\gamma}_{EB1}$
1	38.2162	35.5478	1	20.7772	19.6446
	(0.0982)	(0.0893)		(0.0486)	(0.0513)
2	7.6533	7.2122	2	5.3573	5.1507
	(0.0363)	(0.0323)		(0.0238)	(0.0241)
3	43.2284	39.2546	3	19.6160	18.1074
	(0.1168)	(0.1073)		(0.0626)	(0.0596)
4	20.2147	17.3915	4	14.0965	12.3888
	(0.0946)	(0.0844)		(0.0656)	(0.0575)
5	36.3431	30.2138	5	22.7169	19.5421
	(0.1457)	(0.1300)		(0.0931)	(0.0830)
Mean	**29.1311**	**25.9240**	**Mean**	**16.5128**	**14.9667**
	(0.0468)	**(0.0422)**		**(0.0282)**	**(0.0260)**

Table 7.4.2: MSEs x 1000 for EVI estimates for $n = 500$ and $n = 1000$: first order estimation

The minimum MSE for each portfolio-sample size combination, across the different estimation techniques, are highlighted. This is the minimum across the estimators considered in table 7.4.1, 7.4.2, 7.5.1, 7.5.2, 7.5.5 and 7.5.6.

Tables 7.4.1 and 7.4.2 indicate that $\hat{\gamma}_{EB1}$ outperforms $\hat{\gamma}_0$ for all sample sizes considered.

In order to assess the bias resulting from estimating the EVI, the mean value of the 10 000 Hill estimates and the mean value of the 10 000 first order EB estimates are calculated. These mean values are given in the following table, together with the benchmark estimator for each portfolio.

$n = 100$				$n = 200$			
Portfolio	γ	$\bar{\gamma}_0$	$\bar{\gamma}_{EB1}$	**Portfolio**	γ	$\bar{\gamma}_0$	$\bar{\gamma}_{EB1}$
1	0.4236	0.7569	0.7331	1	0.4236	0.6738	0.6602
2	0.5858	0.6120	0.5953	2	0.5858	0.6073	0.5954
3	0.4959	0.8793	0.8459	3	0.4959	0.7734	0.7541
4	0.8703	0.9706	0.9252	4	0.8703	0.9362	0.9050
5	0.8346	1.0172	0.9707	5	0.8346	0.9859	0.9537

Table 7.4.3: Mean estimates of EVI, together with benchmark estimate of five portfolios

$n = 500$				$n = 1000$			
Portfolio	γ	$\bar{\gamma}_0$	$\bar{\gamma}_{EB1}$	**Portfolio**	γ	$\bar{\gamma}_0$	$\bar{\gamma}_{EB1}$
1	0.4236	0.6025	0.5959	1	0.4236	0.5557	0.5518
2	0.5858	0.5891	0.5821	2	0.5858	0.5866	0.5816
3	0.4959	0.6838	0.6742	3	0.4959	0.6161	0.6106
4	0.8703	0.9257	0.9064	4	0.8703	0.9211	0.9074
5	0.8346	0.9732	0.9533	5	0.8346	0.9411	0.9276

Table 7.4.4: Mean estimates of EVI, together with benchmark estimate of five portfolios

Considering all five portfolios and all sample sizes, the above tables indicate that $\bar{\gamma}_{EB1}$ is closer to the benchmark estimator than $\bar{\gamma}_0$. Considering only the results corresponding to sample size $n = 1000$, both mean estimates differ greatly from the benchmark estimator for portfolio 1, 3 and 5. This result corresponds with the results obtained from the Hill plots and the Pareto plots.

7.5 Second order estimation of the EVI

Section 7.5.1 presents the procedure that will be followed in order to make inferences of a specific insurance portfolio. The results are shown in section 7.5.2.

7.5.1 Procedure of estimation

The following procedure is carried out in order to compare the second order estimates of the EVI:

1. Choose a target portfolio for which inferences will be made.
2. Choose a sample size n. Sample sizes of $n = 100$, $n = 200$, $n = 500$ and $n = 1000$ will be used in this study.
3. Take a bootstrap sample (without replacement) of size n from the target portfolio.
4. Calculate the second order EVI estimate of the bootstrap sample, denoted by $\hat{\gamma}_0^{MLE}$. This estimate is obtained by fitting the PPD to the relative excesses using MLE, after externally estimating the second order parameter by $\hat{\rho}_2$. The number of excesses is fixed at $k = 2n^{2/3}$.
5. Calculate the second order EB estimate of the bootstrap sample, denoted by $\hat{\gamma}_{EB2}$. This is done by including all the data of the other four portfolios in the calculation. Considering the target portfolio, the number of excesses is fixed at $k = 2n^{2/3}$. Considering the other four portfolios, the number of excesses is fixed at $k = 2n_i^{2/3}$, where n_i is the sample size of portfolio $i, i = 1, 2, 3, 4$.
6. Calculate the square error of estimation $(\hat{\gamma} - \gamma)^2$, where γ is the calculated benchmark estimator.
7. Repeat steps 3 – 6 a large number of times. In this study these steps will be repeated 10 000 times.
8. Calculate the MSE.

7.5.2 Results: second order estimation of the EVI

The following tables show the calculated MSE values of the two estimators of the EVI.

$n = 100$			$n = 200$		
Portfolio	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	Portfolio	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$
1	72.5734	152.0570	1	37.2665	76.9701
	(0.2099)	(0.3741)		(0.1113)	(0.1853)
2	26.5548	18.7343	2	20.6003	13.0903
	(0.1250)	(0.1069)		(0.1093)	(0.0707)
3	122.5152	239.9252	3	83.9488	156.8020
	(0.3573)	(0.5527)		(0.2235)	(0.3421)
4	48.0822	68.6304	4	34.7670	44.6966
	(0.2946)	(0.3458)		(0.1824)	(0.2128)
5	58.2598	112.0705	5	39.8267	68.0097
	(0.3442)	(0.4379)		(0.2514)	(0.2710)
Mean	**65.5971**	**118.2835**	**Mean**	**43.2819**	**71.9137**
	(0.1253)	**(0.1753)**		**(0.0826)**	**(0.1049)**

Table 7.5.1: MSEs x 1000 for EVI estimates for second order estimation, with $k = 2n^{2/3}$

$n = 500$			$n = 1000$		
Portfolio	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	Portfolio	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$
1	13.3530	26.4718	1	7.3924	15.0683
	(0.0465)	(0.0787)		(0.0263)	(0.0456)
2	14.1738	9.1931	2	10.6592	7.4032
	(0.0632)	(0.0419)		(0.0507)	(0.0330)
3	42.0632	69.5441	3	22.3163	34.0372
	(0.1066)	(0.1632)		(0.0520)	(0.0758)
4	21.7314	26.2908	4	16.1541	19.0746
	(0.1250)	(0.1172)		(0.0877)	(0.0794)
5	25.7061	46.1027	5	18.2515	32.9614
	(0.1327)	(0.1704)		(0.0956)	(0.1120)
Mean	**23.4055**	**35.5205**	**Mean**	**14.9547**	**21.7089**
	(0.0451)	**(0.0556)**		**(0.0302)**	**(0.0333)**

Table 7.5.2: MSEs x 1000 for EVI estimates for second order estimation, with $k = 2n^{2/3}$

Considering all the sample sizes, the second order EB estimator shows superior performance only for portfolio 2.

The mean values of the two estimates for each portfolio are given in the following tables.

$n = 100$				$n = 200$			
Portfolio	γ	$\bar{\gamma}_0^{MLE}$	$\bar{\gamma}_{EB2}$	**Portfolio**	γ	$\bar{\gamma}_0^{MLE}$	$\bar{\gamma}_{EB2}$
1	0.4236	0.6670	0.7887	1	0.4236	0.5968	0.6814
2	0.5858	0.5551	0.6165	2	0.5858	0.5820	0.6228
3	0.4959	0.8079	0.9520	3	0.4959	0.7603	0.8667
4	0.8703	0.8695	1.0128	4	0.8703	0.8906	0.9970
5	0.8346	0.9248	1.0807	5	0.8346	0.9130	1.0280

Table 7.5.3: Mean estimates of EVI, together with benchmark estimate of five portfolios

$n = 500$				$n = 1000$			
Portfolio	γ	$\bar{\gamma}_0^{MLE}$	$\bar{\gamma}_{EB2}$	**Portfolio**	γ	$\bar{\gamma}_0^{MLE}$	$\bar{\gamma}_{EB2}$
1	0.4236	0.5204	0.5675	1	0.4236	0.4941	0.5315
2	0.5858	0.5952	0.6273	2	0.5858	0.5942	0.6257
3	0.4959	0.6855	0.7425	3	0.4959	0.6342	0.6686
4	0.8703	0.8738	0.9608	4	0.8703	0.8782	0.9480
5	0.8346	0.9088	1.0013	5	0.8346	0.9041	0.9800

Table 7.5.4: Mean estimates of EVI, together with benchmark estimate of five portfolios

Considering all five portfolios and all sample sizes, the above tables indicate that $\bar{\gamma}_0^{MLE}$ is closer to the benchmark estimator than $\bar{\gamma}_{EB2}$. Looking only at the results corresponding to sample size $n = 1000$, both mean estimates differ greatly from the benchmark estimator for portfolio 1, 3 and 5. For portfolio 2 and 4, the EVI estimate obtained by MLE is close to the benchmark estimator.

To further investigate whether the number of excesses used in the EVI estimation process has a significant influence on the final estimate of the EVI, alternative methods of estimating the optimal threshold are considered.

The first method is to fixed the optimal threshold of the bootstrap sample at $k = 2\sqrt{n}$. This is the optimal threshold used in the first order estimation process.

The following tables show the calculated MSE values of the two estimators of the EVI, where $k = 2\sqrt{n}$.

$n = 100$			$n = 200$		
Portfolio	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	**Portfolio**	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$
1	27.3079	81.5880	1	15.0798	38.1855
	(0.1519)	(0.3142)		(0.0818)	(0.1540)
2	45.7144	39.5469	2	35.7197	26.7904
	(0.2401)	(0.2805)		(0.1662)	(0.1577)
3	57.5763	162.0947	3	29.1141	77.9672
	(0.3094)	(0.5781)		(0.1670)	(0.3063)
4	87.5943	100.2430	4	60.6360	61.9380
	(0.4182)	(0.5492)		(0.2404)	(0.2883)
5	77.2404	117.5588	5	57.2630	81.4058
	(0.4108)	(0.6056)		(0.2751)	(0.4149)
Mean	**59.0867**	**100.2063**	**Mean**	**39.5625**	**57.2574**
	(0.1442)	**(0.2173)**		**(0.0885)**	**(0.1261)**

Table 7.5.5: MSEs x 1000 for EVI estimates for second order estimation, with $k = 2\sqrt{n}$

$n = 500$			$n = 1000$		
Portfolio	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	**Portfolio**	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$
1	7.0799	13.6249	1	4.7616	6.2544
	(0.0336)	(0.0579)		(0.0167)	(0.0263)
2	26.4059	16.6411	2	23.9573	12.7848
	(0.1129)	(0.0919)		(0.0908)	(0.0618)
3	14.2265	30.9496	3	9.7051	13.4968
	(0.0941)	(0.1255)		(0.0471)	(0.0596)
4	39.1565	40.7771	4	28.0232	26.6730
	(0.1778)	(0.1984)		(0.1087)	(0.1230)
5	35.6038	51.6918	5	22.6466	29.9788
	(0.1601)	(0.2281)		(0.0987)	(0.1260)
Mean	**24.4945**	**30.7369**	**Mean**	**17.8188**	**17.8376**
	(0.0566)	**(0.0690)**		**(0.0359)**	**(0.0395)**

Table 7.5.6: MSEs x 1000 for EVI estimates for second order estimation, with $k = 2\sqrt{n}$

In the situation where $k = 2\sqrt{n}$ and considering all sample sizes, the second order EB estimate performs again only better than the benchmark estimator for portfolio 2. For sample size $n = 1000$, the second order EB estimate performs also better than the benchmark estimator for portfolio 4.

It should be noted that for sample size $n = 1000$, the difference between the mean value obtained by $\hat{\gamma}_0^{MLE}$ and the mean value obtained by $\hat{\gamma}_{EB2}$ is only 0.0188. In the situation where $k = 2n^{2/3}$, the difference between the two mean values was 6.7542. The choice of threshold therefore has a significant influence on the overall results.

The following tables compare the mean values of the two estimates to the benchmark estimates.

$n = 100$				$n = 200$			
Portfolio	γ	$\bar{\gamma}_0^{MLE}$	$\bar{\gamma}_{EB2}$	**Portfolio**	γ	$\bar{\gamma}_0^{MLE}$	$\bar{\gamma}_{EB2}$
1	0.4236	0.5096	0.6637	1	0.4236	0.4663	0.5811
2	0.5858	0.5364	0.6440	2	0.5858	0.5504	0.6346
3	0.4959	0.6429	0.8345	3	0.4959	0.5784	0.7244
4	0.8703	0.8380	1.0027	4	0.8703	0.8585	0.9820
5	0.8346	0.8563	1.0346	5	0.8346	0.8743	1.0144

Table 7.5.7: Mean estimates of EVI, together with benchmark estimate of five portfolios

$n = 500$				$n = 1000$			
Portfolio	γ	$\bar{\gamma}_0^{MLE}$	$\bar{\gamma}_{EB2}$	**Portfolio**	γ	$\bar{\gamma}_0^{MLE}$	$\bar{\gamma}_{EB2}$
1	0.4236	0.4194	0.5063	1	0.4236	0.3985	0.4691
2	0.5858	0.5724	0.6271	2	0.5858	0.5770	0.6159
3	0.4959	0.5258	0.6316	3	0.4959	0.4868	0.5698
4	0.8703	0.8703	0.9745	4	0.8703	0.8623	0.9519
5	0.8346	0.8577	0.9834	5	0.8346	0.8394	0.9459

Table 7.5.8: Mean estimates of EVI, together with benchmark estimate of five portfolios

As in the case where $k = 2n^{2/3}$, $\bar{\gamma}_0^{MLE}$ is overall closer to the benchmark estimator than $\bar{\gamma}_{EB2}$.

Comparing the two threshold selection methods, improved results are obtained for portfolio 2, 4 and 5 in the situation where $k = 2n^{2/3}$, since these portfolios exhibited “good” behaviour. However, when the number of excesses is fixed at $k = 2\sqrt{n}$, improved results are obtained for portfolio 1 and 3, since these portfolios did not show “good” behaviour.

The second method that will be used to determine whether the choice of threshold has a significant influence on the estimate of the EVI, is to express the number of excesses k as a proportion of the sample size n. The possible values of k which will be considered are $k = 5\%$, $k = 10\%$, $k = 20\%$, $k = 30\%$, $k = 40\%$ and $k = 50\%$ of n, rounded to the nearest integer.

The results obtained are given in the following tables.

Threshold: $k = 5\%$ of n

	$n = 100$		$n = 200$		$n = 500$		$n = 1000$	
Portfolio	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$
1	30.3966	202.7945	19.0168	52.4557	10.1705	12.4978	6.0241	5.7361
	(0.1119)	(0.7988)	(0.0726)	(0.2377)	0.0368	(0.0641)	(0.0216)	(0.0261)
2	95.0769	260.0452	71.2206	104.9423	40.6179	31.2471	26.9579	15.6128
	(0.4048)	(1.3503)	(0.3051)	(0.6458)	(0.1703)	(0.1681)	(0.0915)	(0.0733)
3	49.2196	278.2495	39.4840	101.5303	21.9792	29.0283	11.6508	13.0599
	(0.2201)	(1.0630)	(0.2290)	(0.5147)	(0.1033)	(0.1428)	(0.0548)	(0.0575)
4	160.4147	402.0463	94.9457	195.1276	57.0107	68.5713	29.1262	30.6872
	(0.6363)	(1.6920)	(0.3748)	(0.9885)	(0.2146)	(0.3398)	(0.1213)	(0.1320)
5	146.0731	416.0406	83.9569	187.4035	41.7517	62.7902	24.3405	32.6037
	(0.5805)	(1.7437)	(0.3319)	(0.8983)	(0.1677)	(0.3100)	(0.0993)	(0.1471)
Mean	**96.2362**	**311.8352**	**61.7248**	**128.2919**	**34.3060**	**40.8269**	**19.6199**	**19.5400**
	(0.1966)	**(0.6163)**	**(0.1267)**	**(0.3177)**	**(0.0679)**	**(0.1028)**	**(0.0382)**	**(0.0440)**

Table 7.5.9: MSEs x 1000 for EVI estimates where $k = 5\%$ of n, for all sample sizes

Threshold: $k = 10\%$ of n

	$n = 100$		$n = 200$		$n = 500$		$n = 1000$	
Portfolio	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$
1	23.7889	85.8649	13.7999	34.9675	6.5249	14.1120	3.7082	8.1993
	(0.1365)	(0.3771)	(0.0690)	(0.1550)	(0.0320)	(0.0551)	(0.0175)	(0.0347)
2	67.1320	90.4940	39.7440	34.3223	26.1203	16.3423	18.8793	10.3819
	(0.3496)	(0.6296)	(0.1655)	(0.1885)	(0.1100)	(0.0865)	(0.0714)	(0.0476)
3	44.0690	161.7611	26.9966	71.6256	13.3586	33.3613	6.7210	18.9687
	(0.2899)	(0.7842)	(0.1605)	(0.3015)	(0.0782)	(0.1259)	(0.0426)	(0.0695)
4	117.4630	195.0522	73.2987	81.8854	40.8294	39.7230	23.8526	22.7755
	(0.5011)	(0.9485)	(0.3016)	(0.4239)	(0.1659)	(0.1740)	(0.1033)	(0.0968)
5	101.4586	224.8421	63.8039	106.6975	34.0028	48.3622	18.9758	30.9113
	(0.4154)	(1.0013)	(0.2948)	(0.5238)	(0.1474)	(0.1989)	(0.0847)	(0.1129)
Mean	**70.7823**	**151.6029**	**43.5286**	**65.8997**	**24.1672**	**30.3802**	**14.4274**	**18.2473**
	(0.1611)	**(0.3496)**	**(0.0971)**	**(0.1555)**	**(0.0523)**	**(0.0620)**	**(0.0317)**	**(0.0349)**

Table 7.5.10: MSEs x 1000 for EVI estimates where $k = 10\%$ of n, for all sample sizes

Threshold: $k = 20\%$ of n

	$n = 100$		$n = 200$		$n = 500$		$n = 1000$	
Portfolio	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$
1	25.9334	75.7684	17.2029	46.5002	10.0167	25.5367	7.5762	15.1945
	(0.1641)	(0.2920)	(0.0904)	(0.1640)	(0.0453)	(0.0833)	(0.0277)	(0.0492)
2	39.1809	35.6801	28.8360	20.2773	17.2519	10.6867	9.6534	6.4830
	(0.1894)	(0.2310)	(0.1400)	(0.1240)	(0.0820)	(0.0544)	(0.0425)	(0.0280)
3	52.7984	155.6087	36.6855	94.0872	26.1269	54.1240	21.1910	32.5444
	(0.2741)	(0.5537)	(0.1541)	(0.2875)	(0.0857)	(0.1423)	(0.0463)	(0.0692)
4	85.3238	94.4286	54.5477	56.0696	26.6965	29.2056	14.1188	18.9870
	(0.3752)	(0.4581)	(0.2581)	(0.2699)	(0.1223)	(0.1251)	(0.0633)	(0.0746)
5	79.2243	122.8984	52.7520	68.8991	29.5027	45.1751	18.0256	33.1772
	(0.3924)	(0.6254)	(0.2845)	(0.3378)	(0.1411)	(0.1643)	(0.0918)	(0.1068)
Mean	**56.4922**	**96.8768**	**38.0048**	**57.1667**	**21.9189**	**32.9456**	**14.1130**	**21.2772**
	(0.1316)	**(0.2046)**	**(0.0892)**	**(0.1117)**	**(0.0452)**	**(0.0540)**	**(0.0262)**	**(0.0316)**

Table 7.5.11: MSEs x 1000 for EVI estimates where $k = 20\%$ of n, for all sample sizes

Threshold: $k = 30\%$ of n

	$n = 100$		$n = 200$		$n = 500$		$n = 1000$	
Portfolio	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$
1	38.7398	105.5904	29.6337	67.5696	23.4124	37.5483	20.8483	25.7673
	(0.1703)	(0.3204)	(0.1123)	(0.1995)	(0.0610)	(0.0916)	(0.0420)	(0.0511)
2	32.5415	25.3202	21.7180	14.6513	12.9392	8.9114	7.6466	6.0975
	(0.1620)	(0.1483)	(0.0958)	(0.0788)	(0.0633)	(0.0405)	(0.0408)	(0.0276)
3	80.3808	198.0291	67.9162	138.9451	59.9455	85.9082	55.9485	65.7325
	(0.3273)	(0.5711)	(0.2016)	(0.3340)	(0.1154)	(0.1679)	(0.0763)	(0.0941)
4	66.6109	73.1400	40.4438	48.4533	19.1990	28.7951	8.6937	17.1290
	(0.3030)	(0.3633)	(0.2179)	(0.2345)	(0.1068)	(0.1267)	(0.0447)	(0.0702)
5	70.0606	99.8350	46.2028	71.5921	23.2862	46.2457	14.1885	30.8027
	(0.3675)	(0.4866)	(0.2355)	(0.2751)	(0.1238)	(0.1598)	(0.0674)	(0.1115)
Mean	**57.6667**	**100.3829**	**41.1829**	**68.2423**	**27.7565**	**41.4817**	**21.4651**	**29.1058**
	(0.1248)	**(0.1811)**	**(0.0813)**	**(0.1074)**	**(0.0437)**	**(0.0565)**	**(0.0251)**	**(0.0344)**

Table 7.5.12: MSEs x 1000 for EVI estimates where $k = 30\%$ of n, for all sample sizes

Threshold: $k = 40\%$ of n

	$n = 100$		$n = 200$		$n = 500$		$n = 1000$	
Portfolio	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$
1	59.5336	134.3013	49.8375	88.8144	45.3284	56.7229	41.6256	45.9968
	(0.1922)	(0.3248)	(0.1193)	(0.1970)	(0.0791)	(0.1026)	(0.0554)	(0.0594)
2	30.4550	20.1490	20.4011	13.2514	10.7488	7.9015	5.4809	5.1203
	(0.1624)	(0.1129)	(0.1147)	(0.0766)	(0.0543)	(0.0348)	(0.0341)	(0.0209)
3	119.4123	240.8345	109.5424	186.7718	107.9232	137.8901	102.1100	113.4957
	(0.3842)	(0.6052)	(0.2387)	(0.3687)	(0.1479)	(0.2042)	(0.0966)	(0.1080)
4	56.1537	78.7366	27.8708	44.9270	11.0392	21.7977	4.8400	8.6728
	(0.3004)	(0.3802)	(0.1592)	(0.2202)	(0.0628)	(0.1014)	(0.0253)	(0.0456)
5	58.9671	107.6084	31.0300	66.1412	15.7206	35.6756	9.7191	16.9918
	(0.3195)	(0.4978)	(0.1831)	(0.2563)	(0.0761)	(0.1372)	(0.0378)	(0.0612)
Mean	**64.9043**	**116.3260**	**47.7363**	**79.9812**	**38.1520**	**51.9976**	**32.7551**	**38.0555**
	(0.1270)	**(0.1873)**	**(0.0757)**	**(0.1086)**	**(0.0404)**	**(0.0575)**	**(0.0250)**	**(0.0293)**

Table 7.5.13: MSEs x 1000 for EVI estimates where $k = 40\%$ of n, for all sample sizes

Threshold: $k = 50\%$ of n

	$n = 100$		$n = 200$		$n = 500$		$n = 1000$	
Portfolio	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$	$\hat{\gamma}_0^{MLE}$	$\hat{\gamma}_{EB2}$
1	95.9392	175.6685	84.7572	118.6885	81.3692	91.1819	81.0276	85.4189
	(0.2261)	(0.3675)	(0.1420)	(0.1939)	(0.0954)	(0.1069)	(0.0723)	(0.0736)
2	23.9923	18.3926	15.9463	11.9675	8.3090	6.9862	5.9301	5.0349
	(0.1132)	(0.0949)	(0.0719)	(0.0589)	(0.0335)	(0.0280)	(0.0179)	(0.0163)
3	138.4549	259.0205	136.6324	212.4979	135.6145	162.3644	137.2342	148.5659
	(0.3585)	(0.5508)	(0.2482)	(0.3631)	(0.1553)	(0.1847)	(0.1073)	(0.1214)
4	32.9528	67.2438	18.7475	40.8044	6.7826	11.7861	3.7444	5.1289
	(0.1697)	(0.3150)	(0.1025)	(0.1868)	(0.0322)	(0.0616)	(0.0176)	(0.0237)
5	40.2554	103.0826	28.4196	64.4279	18.0450	29.2190	14.3224	18.0484
	(0.2163)	(0.3964)	(0.1615)	(0.2525)	(0.0671)	(0.1099)	(0.0427)	(0.0525)
Mean	**66.3189**	**124.6816**	**56.9006**	**89.6772**	**50.0240**	**60.3075**	**48.4518**	**52.4394**
	(0.1035)	**(0.1678)**	**(0.0703)**	**(0.1042)**	**(0.0399)**	**(0.0499)**	**(0.0277)**	**(0.0308)**

Table 7.5.14: MSEs x 1000 for EVI estimates where $k = 50\%$ of n, for all sample sizes

The mean values of the two estimates (for each k value) are shown in the following graphs. The dashed line indicates the benchmark estimator of the relative portfolio. The first vertical line indicates where $k = 2\sqrt{n}$ and the second vertical line indicates where $k = 2n^{2/3}$.

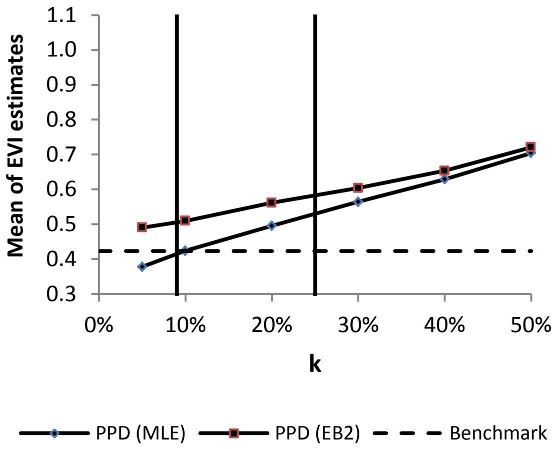


Figure 7.5.1: Mean of EVI estimates: P1, $n = 500$

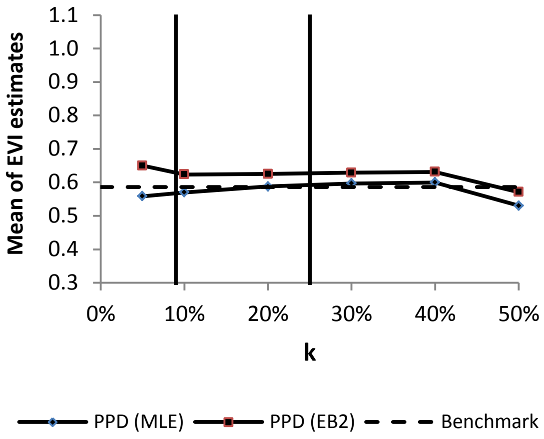


Figure 7.5.2: Mean of EVI estimates: P2, $n = 500$

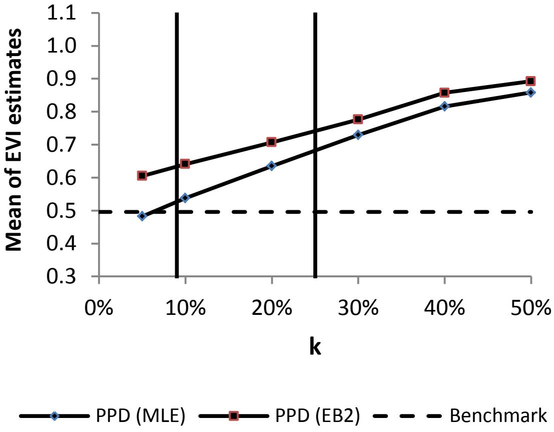


Figure 7.5.3: Mean of EVI estimates: P3, $n = 500$

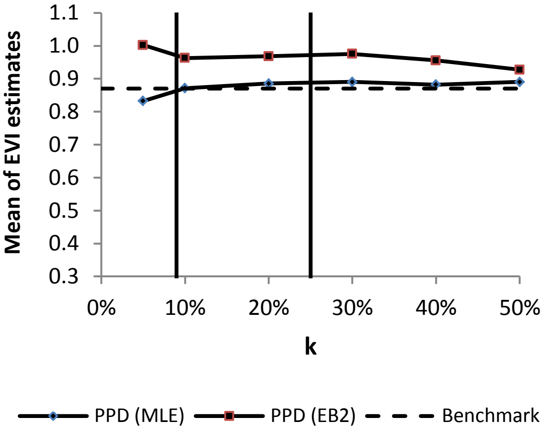


Figure 7.5.4: Mean of EVI estimates: P4, $n = 500$

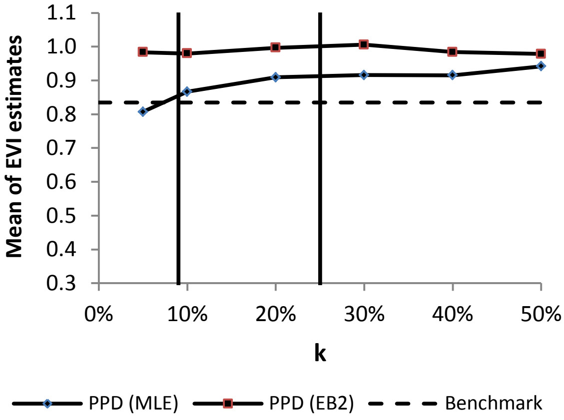


Figure 7.5.5: Mean of EVI estimates: P5, $n = 500$

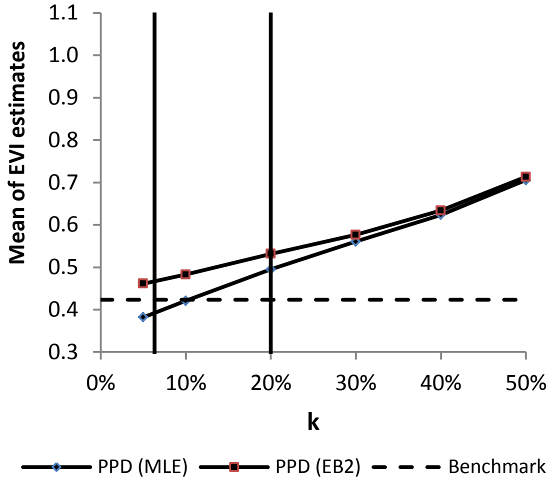


Figure 7.5.6: Mean of EVI estimates: P1, $n = 1000$

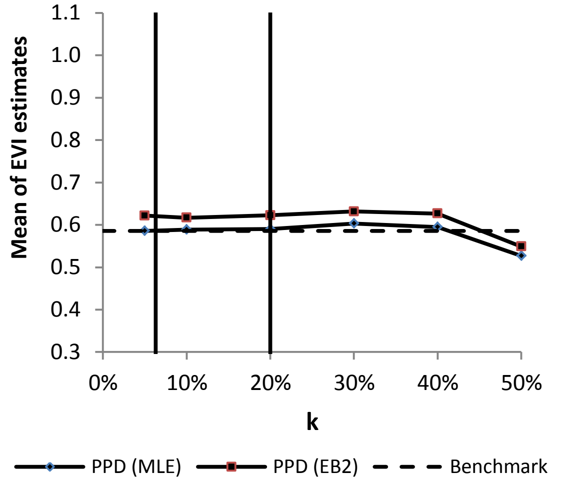


Figure 7.5.7: Mean of EVI estimates: P2, $n = 1000$

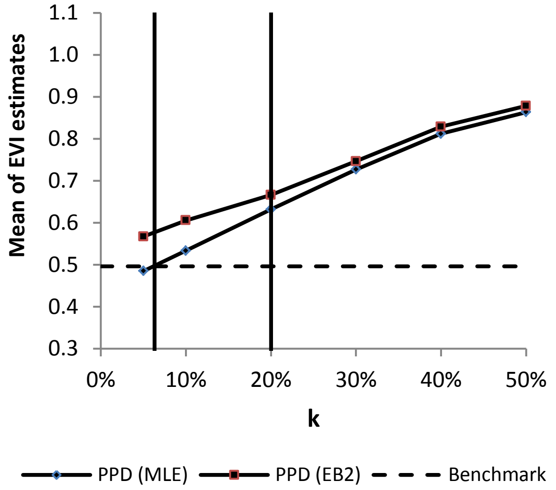


Figure 7.5.8: Mean of EVI estimates: P3, $n = 1000$

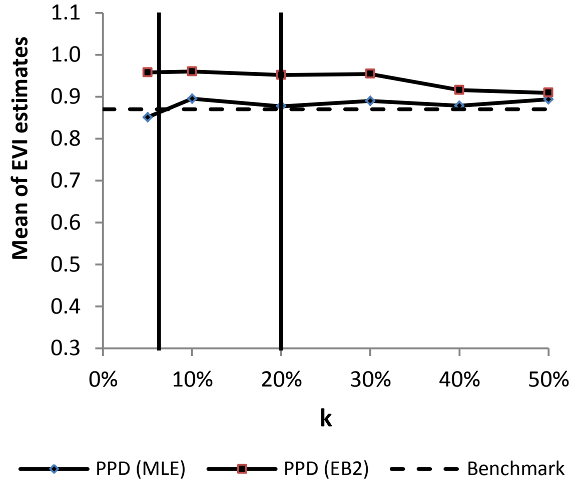


Figure 7.5.9: Mean of EVI estimates: P4, $n = 1000$

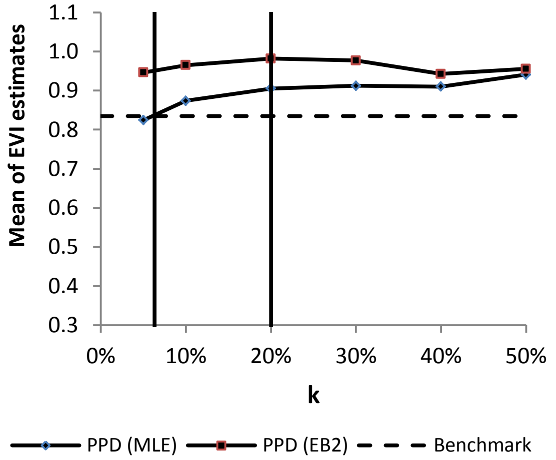


Figure 7.5.10: Mean of EVI estimates: P5, $n = 1000$

The graphs indicate that the choice of threshold does not have a significant influence on the results of portfolio 2, 4 and 5, since the two mean values stay relatively stable across all k values. For portfolio 1 and 3, the choice of threshold influence both mean values significantly.

For example, figure 7.5.2 shows the mean values of the EVI estimates for portfolio 2, where $n = 500$. The graph shows that when a small number of excesses are used to estimate the EVI, $\bar{\gamma}_0^{MLE}$ is just below the benchmark estimator and $\bar{\gamma}_{EB2}$ is just above the benchmark estimator. The difference between $\bar{\gamma}_0^{MLE}$ and the benchmark estimator is -0.027 when k is fixed at 5% of n. The difference between $\bar{\gamma}_{EB2}$ and the benchmark estimator is 0.064 when $k = 5\%$ of n.

When a large number of excesses are used to estimate the EVI, both estimators lie just below the benchmark. If k is fixed at 50% of n, the difference between $\bar{\gamma}_0^{MLE}$ and the benchmark is even smaller (-0.055). The difference between $\bar{\gamma}_{EB2}$ and the benchmark is also smaller (-0.015).

Figure 7.5.3 shows the mean values of the EVI estimates for portfolio 3, where $n = 500$. The graph shows that, when k is fixed at 5% of n, the differences between the two estimates and the benchmark are in the same range as in the case with figure 7.5.2 (-0.013 and 0.109, respectively). However, when k is fixed at 50% of n, the difference between $\bar{\gamma}_0^{MLE}$ and the benchmark is considerably bigger (0.363). This is also true for the difference between $\bar{\gamma}_{EB2}$ and the benchmark (0.396).

In summary, if a high threshold (small k) is used to estimate the EVI, the estimates obtained for portfolio 1 and 3 are poor. However, a high threshold leads to sufficient EVI estimators for portfolio 2, 4 and 5.

If a low threshold (large k) is used to estimate the EVI, relatively good estimates are obtained for all portfolios when the PPD is fitted to the relative excesses. For large k, the second order EB estimation method leads to estimates that are too high on average.

7.6 Conclusion

In this chapter a case study was used to compare the performance of the first order EB estimator to the performance of the Hill estimator. The data of the case study were also used to compare the performance of the second order EB estimator to the performance of the benchmark estimator, namely to fit the PPD to the relative multiplicative excesses by means of MLE.

In section 7.2 a general description of the insurance claims data was given. Graphical and numerical techniques were used in section 7.3 to obtain a benchmark EVI estimate for each insurance portfolio. The Hill plots indicated that the data of portfolio 1 and 3 will not lead to reliable estimates of the EVI. For portfolio 2, 4 and 5, it will be possible to obtain a reliable estimate of the EVI.

Section 7.4 presents the first order estimation procedure that was followed in order to make inferences of a specific insurance portfolio. It was found that the first order EB estimator performed better overall than the Hill estimator.

Section 7.5 provides the second order estimation procedure that was followed. The results indicated that the MLE estimator performed better overall than the second order EB estimator.

With regards to second order estimation of the EVI, alternative methods of threshold selection were considered. This was done in order to determine if the choice of threshold has a significant influence on the EVI estimates. It was found that, for portfolio 1 and 3, the EVI estimates depend greatly on the choice of threshold. Especially a too low threshold (large k) leads to poor results. The EVI estimates obtained for portfolio 2, 4 and 5 were more robust with regards to the method of threshold selection. These estimates were stable for a wide range of k.

Overall, the second order EB estimation method never led to improved estimates of the EVI, when compared to the first order EB estimation method. With regards to the individual insurance portfolios, it was found that the second order MLE estimator resulted in an accurate EVI estimate for portfolios 1 and 3, with $k = 2\sqrt{n}$. The second order MLE estimator with $k = 2n^{2/3}$ resulted in an accurate EVI estimate for portfolio 5. Considering portfolios 2 and 4, the first order EB estimation method resulted in accurate estimates of the EVI, with $k = 2\sqrt{n}$.

CHAPTER 8

Conclusion

The main purpose of this thesis was to obtain an improved estimator of the extreme value index that is applicable to an ANOVA setting. Empirical Bayes techniques were considered and a first order and a second order EB estimator were derived.

A case study of five insurance claims portfolios was used to determine whether the two EB estimators improve the accuracy of estimating the EVI, when compared to similar estimators in the literature, not taking info from other treatments (portfolios) into account.

8.1 Summary of findings

In chapter 2 an overview of extreme value theory was given. Assumptions that were made in this thesis were listed and motivated.

Chapter 3 discussed different methods that can be used to estimate the EVI. The theory of first and second order estimation of the EVI was introduced. The chapter discussed specific aspects regarding the estimation of the EVI, which included methods of determining the optimal choice of threshold and techniques of estimating the second order parameter.

In chapter 4 an overview of Bayesian theory was provided. Approaches of obtaining prior distributions were mentioned. In this chapter Gibbs sampling was discussed in detail. Gibbs sampling was used in upcoming chapters to obtain Bayesian estimates of the PPD parameters.

A simulation study was conducted in chapter 5. The aim of the simulation study was to compare the performance of a number of EVI estimators to each other. The study compared the performance of three first order estimators and two second order estimators.

With regards to first order estimation, fixed threshold and adaptive threshold selection methods were used to determine the optimal number of excesses. It was found that the method of Drees and Kauffmann (1998) performed the best for certain distributions, for example the Burr distribution with $\gamma = -2$ and $\gamma = -0.25$, irrespective of sample size.

With regards to second order estimation, the method followed in this thesis was to externally estimate the second order parameter. Two methods were considered to estimate this parameter. The results indicated that the method proposed by Gomes and Martins (2001) was the preferred method to use. Also, the threshold was fixed at $k = 2n^{2/3}$.

The results obtained from the simulation study showed that the two second order estimation methods performed better overall than the first order estimation methods. Comparing the two second order estimation methods of the EVI, the method based on Bayesian methodology performed better when the sample size was small. In the situation where the sample size was large, the method based on maximum likelihood performed better.

In chapter 6 an overview of EB methodology was given. The chapter included a description of how EB methods can be applied in the general ANOVA setting. A first order EB estimate of the EVI and a second order EB estimate of the EVI were also derived.

Chapter 7 compared the performance of the derived first and second order EB estimators to the first and second order benchmark estimators. This was done by means of a case study. The case study data consisted of the claim sizes of five insurance portfolios. By analysing the data of each portfolio, it was found that for portfolios 1 and 3, it would not be possible to obtain reliable estimates of the EVI.

Considering first order estimation of the EVI, the results indicated that the first order EB estimator outperformed the Hill estimator, which was used as benchmark estimator.

Considering second order estimation of the EVI, the results indicated that the benchmark estimator, namely to fit the PPD to the multiplicative excesses using MLE, outperformed the second order EB estimator.

Comparing the first and second order EB estimation methods, the second order EB method never resulted in improved estimates of the EVI. For portfolio 2, 4 and 5, the first order EB estimation method were superior to the rest. For portfolio 1 and 3, the second order MLE estimator resulted in the most accurate estimates of the EVI.

Alternative methods of threshold selection for the second order estimation of the EVI were also investigated, choosing $k = 2n^{2/3}$ and $k = 2\sqrt{n}$. It was found that for portfolios that did not show "good" behaviour (portfolios 1 and 3) the estimates of the EVI depended significantly on the choice of threshold. For portfolios that showed "good" behaviour (portfolios 2, 4 and 5), the choice of threshold was not critical. The estimates obtained for these portfolios were stable for a wide range of k.

8.2 Future research recommendations

The following areas can be considered for future research:

1. The alternative EVI estimators mentioned in section 3.4 can be included in the simulation study.

2. Only two estimators were considered for the external estimator of the second order parameter of the PPD. The performance of other estimators can be investigated.

3. The results of the case study indicated that the choice of threshold has a significant influence on the estimates of the EVI. The method that was used in this thesis can be improved upon by investigating alternative ways of determining the optimal threshold.

4. The second order EB estimation method did not perform better than second order maximum likelihood estimation. Alternative prior distributions can be considered to determine if improved estimates of the EVI can be obtained.

References

Aragones, J.R., Blanco, C. & Dowd, K. 2000. *Learning curve extreme value var.* Derivatives Week.

Ashour, S.K. & El-Adl, Y.M. 1980. *Bayesian estimation of the parameters of the extreme value distribution*. Eqyptian Statistical Journal, vol. 24, p.140-152.

Beirlant, J. & Goegebeur, Y. 2004. *Simultaneous tail index estimation*. REVSTAT, vol.2, p.15-39.

Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J.L. 2004. *Statistics of extremes.* Chichester: John Wiley & Sons.

Bernardo, J.M. 2003. *Bayesian statistics*. Encyclopedia of Life Support Systems. Probability and Statistics. Oxford, UK.

Berning, T.L. 2013. *Quantification of Estimation Instability and its Application to Threshold Selection in Extremes.* Working paper.

Berning, T.L. 2010. *Improved estimation procedures for a positive extreme value index.* PhD thesis. Stellenbosch University.

Carlin, B.P. & Louis, T.A. 1996. *Bayes and empirical Bayes methods for data analysis*. Chapman & Hall.

Casella, G. 1992. *Illustrating empirical Bayes methods*. Chemometrics and Intelligent Laboratory Systems, vol. 16, p.107-125. Elsevier Science Publishers, Amsterdam.

Casella, G. & George, E.I. 1992. *Explaining the Gibbs sampler.* The American Statistician, vol. 4, p.167-174.

Coles, S.G. 2001. *An introduction to statistical modelling of extreme values*. London: Springer.

Coles, S. & Tawn, J.A. 1996. *Modelling extremes: a Bayesian approach*. Applied Statistics, vol. 45, p.463-478.

Csörgó, S., Deheuvels, P. & Mason, D. 1985. *Kernel estimates of the tail index of a distribution*. The Annals of Statistics, vol. 13, p.1050-1077.

Cui, G., Wong, M.L. & Lui, H. 2006. *Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming*. Management Science, vol.52, p.597-612.

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.

De Haan, L. & Stadtmüller, U. 1996. *Generalized regular variation of second order*. Journal of the Australian Mathematical Society, vol. 61, p.381-395.

De Wet, T. & Berning, T.L. 2013. *Empirical Bayes estimation of the EVI*. Working paper. Stellenbosch University.

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, p.1795-1832. [Online] Available: http://www.jstor.org/stable/2241667. Accessed: 07/01/2012 06:41.

Drees, H. 1995. *Refined Pickands estimators of the extreme value index*. The Annals of Statistics, vol. 23, p.2059-2080. Published by: Institute of Mathematical Statistics. [Online] Available: http://www.jstor.org/stable/2242785. Accessed: 07/01/2012 06:33.

Drees, H. & Kaufmann, E. 1998. *Selecting the optimal sampling fraction in univariate extreme value estimation*. Stochastic Processes and their Applications, vol. 75, p.149-172.

Eddy, S.R. 2004. *What is Bayesian statistics?* Nature Biotechnology, vol.22, p.1177-1178.

Edwards, A.W.F. 1972. *Likelihood*. Cambridge University Press.

Embrechts, P. 2000. *Extreme Value Theory: Potential and Limitations as an integrated risk management tool*. Derivatives Use, Trading & Regulation, vol.6.

Engelund, S. & Rackwitz, R. 1992. *On predictive distribution functions for the three asymptotic extreme value distributions*. Structural Safety, vol. 11, p.255-258.

Ferson, S. *Bayesian methods in risk assessment*. Prepared for: D. Guyonnet. [Online] Available:
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.87.4577&rep=rep1&type=pdf.
Accessed: 16/02/2013 10:43.

Feuerverger, A. & Hall, P. 1999. *Estimating a tail exponent by modelling departure from a Pareto distribution*. Annals of Statistics, vol. 27, p.760-781.

Fisher, R.A. & Tippet, L.H.C. 1928. *Limiting forms of the frequency distribution of the largest or smallest member of a sample*. Cambridge Philosophical Society, vol. 24, p.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*. Portugaliae Mathematica, vol. 60.

Fraga Alves, M.I., Gomes, M.I., de Haan, L. & Neves, C. 2009. *Mixed Moment Estimator and Location Invariant Alternatives.* Extremes, vol. 12, p.149-185.

Frechét, M. 1927. *Sur la loi de probabilité de l'écart maximum*. Ann. Soc. Polon. Math. Cracovie, vol. 6, p.93-116.

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, p.139-159.

Geyer, C.J. 1992. *Practical Markov Chain Monte Carlo*. Statistical Science, vol. 7, p.473-483. Published by: Institute of Mathematical Statistics.
[Online] Available: http://www.jstor.org/stable/2246094. Accessed: 14/12/2012 03:15.

Gnedenko, B. 1943. *Sur la distribution limite du terme maximum d'une série aléatoire.* Translated and reprinted in: *Breakthroughs in Statistics*, 1992. Springer-Verlag, vol. 1, p.195-225.

Gomes, M.I., Caero, 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, p.497-510.

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, p.5-31.

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. & Neves, M. 2000. *Alternatives to a semi-parametric estimator of parameters of rare events: the Jackknife methodology*. Extremes, vol. 3, p.207-229.

Gomes, M.I., Martins, M.J. & Neves, M. 2007. *Improving second order reduced bias extreme value index estimation*. REVSTAT Statistical Journal, vol.5, p.117-207.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C. & Wallis, J.R. 1979. *Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form.* Water Resources Research, vol. 15, p.1049-1054.

Guillou, A. & Hall, P. 2001. *A diagnostic for selecting the threshold in extreme value analysis*. Journal of the Royal Statistical Society, vol.63, p.293-305.

Gumbel, E.J. 1941. *The return period of flood flows.* Annual Mathematics Statistics, vol. 12, p.163-190.

Gumbel, E.J. 1944. *On the plotting of flood discharges*. American Geophysical Union, vol. 25, p.699-719.

Gumbel, E.J. 1958. *Statistics of Extremes*. Columbia University Press, New York.

Hill, B.M. 1975. *A simple general approach to inference about the tail of a distribution.* Annals of Statistics, vol. 3, p.1163-1174.

Hosking, J.R.M., Wallis, J.R. & Wood, E.F. 1985. *Estimation of the generalized extreme-value distribution by the method of probability-weighted moments*. Technometrics, vol. 27, p.251-261.

Kass, R.E. & Wasserman, L.A. 1996. *Formal rules for selection prior distributions.* To appear in Journal of the American Statistical Association.

Keller, G. & Warrack, B. 2000. *Statistics for Management and Economics*. 5th edition. Duxbury.

Kotz, S. & Nadarajah, S. 2000. *Extreme value distributions: theory and applications.* Imperial College Press.

Kroese, D.P., Taimre, T. & Botev, Z.I. 2011. *Handbook of Monte Carlo Methods*, Wiley.

Lee, M.D. & Wagenmakers, E.J. 2005. *Bayesian statistical inference in psychology: comment on Trafimow (2003).* Psychological Review, vol. 112, p.662-668.

Matthys, G. & Beirlant, J. 2003. *Estimating the extreme value index and high quantiles with exponential regression models*. Statistica Sinica, vol.13, p.853-880.

Morris, C.N. 1983. *Parametric empirical Bayes inference: theory and applications*. Journal of the American Statistical Association, vol. 78, p.47-65.

Nordquist, J.M. 1945. *Theory of largest values, applied to earthquake magnitudes.* American Geophysical Union, vol. 26, p.29-31.

Novick, M.R., Jackson, P.H., Thayer, D.T. & Cole, N.S. 1972. *Estimating multiple regressions in m-groups: a cross-validation study*. British Journal of Mathematical and Statistical Psychology, vol. 75, p.33-50.

O'Hagan, A. 2003. *Bayesian statistics: principles and benefits*, chapter 3. In Bogers, R.J., (ed.) Wageningen UR Frontis Series: Bayesian Statistics and quality modelling in the agro-food production chain, vol. 3, p.31-45.

Peng, L. & Qi, Y. 2004. *Estimation the first- and second-order parameters of a heavy-tailed distribution.* Australian & New Zealand Journal of Statistics, vol. 46 (2), p.205-312.

Pickands, J. 1975. *Statistical inference using extreme order statistics*. Annals of Statistics, vol.3, p.119-131.

Potter, W.D. 1949. *Normalcy tests of precipitation and frequency studies of runoff on small watersheds.* U.S. Department of Agriculture Technical Bulletin, no.985.

Robbins, H. 1983. *Some thoughts on empirical Bayes estimation*. Annals of Statistics, vol. 11, p.713-723.

Spiegelhalter, D. & Rice, K. 2009. *Bayesian statistics*. Scholarpedia, 4(8):5230.

Von Bortkiewicz, L. 1922. *Variantionsbreite und mittlerer Fehler*. Sitzungsber. Berli. Math. Ges, vol. 21, p.3-11.

Von Mises, R. 1936. *La distribution de la plus grande de n valeurs*. Reproduced in *Selected Papers of Richard von Mises, II* (1954), American Mathematical Society, p.271-294.

Wackerly, D., Mendenhall, W. & Scheaffer, R. 2008. *Mathematical statistics with applications*. Seventh ed. Thomson Learning Inc.

Zellner, A. 1971. *An introduction to Bayesian inference in econometrics*. New York: John Wiley & Sons.

Zipf, G.K. 1949. *Human behaviour and the principle of least effort: an introduction to human ecology*. Reading, Massachusetts: Addison-Wesley.