Extreme Values. High Order Quantiles and Applications.

Laura Luciana Cavalcante de Sousa

Programa de Doutoramento em Matemática Aplicada
Departamento de Matemática
2014

Orientador
Ana Cristina Moreira Freitas, Professora Auxiliar, FEP

Coorientador
Margarida Maria Araújo Brito, Professora Associada, FCUP

Happiness is when what you think, what you say, and what you do are in harmony.

Mahatma Gandhi

Il est impossible que l'improbable n'arrive jamais.

Emil Julius Gumbel

Agradecimentos

“Quando tudo parece dar errado, acontecem coisas boas que não teriam acontecido se tudo tivesse dado certo.”

E algumas das melhores coisas boas que me aconteceram ao longo deste percurso foram:

O apoio constante dedicado pelas minhas orientadoras Ana Cristina Moreira Freitas e Margarida Brito e a amizade entre nós desenvolvida. Quando tudo parecia dar errado (e não só), encorajaram, motivaram e inspiraram a superar as dificuldades, incentivando-me a ultrapassar as barreiras do conhecimento.

A amizade especial, carinho e paciência dos meus amigos mais queridos, nos quais estão incluídos os familiares que mais estimo. Quando tudo parecia dar errado, não desistiram até me desencaminharem rumo a momentos mais agradáveis, cedendo muito do seu tempo em prol desta amizade.

A todas as pessoas aqui mencionadas e ainda a todas as outras que interagiram comigo durante esta aventura, agradeço por terem partilhado um pouco delas e de alguma forma me terem proporcionado o equilíbrio necessário para que este caminho pudesse ser percorrido com alegria e felicidade.

Gostaria ainda de agradecer:

Ao Centro de Matemática da Universidade do Porto pelo apoio facultado para participação em conferências e à Faculdade de Ciências da Universidade do Porto pelas condições de trabalho proporcionadas.

À Fundação para a Ciência e a Tecnologia (FCT) pelo apoio financeiro proporcionado relativo à bolsa de doutoramento com a referência SFRH/BD/60642/2009, co-financiada pelo POPH - Programa Operacional do Potencial Humano.

À McGraw-Hill Education LLC e a Horst Bergmeier, por terem permitido a utilização de algumas imagens ilustrativas neste documento.

Por último, manifesto com muito carinho a minha gratidão à professora Maria Adelaide Bessa por me ter feito sentir a Matemática. A confiança por ela transmitida foi sem dúvida o primeiro grande impulso para toda uma vida académica a desenvolver nesta deslumbrante área.

Resumo

Neste trabalho debruçamo-nos em torno de problemas que se destacam em Teoria de Valores Extremos pela sua relevância em aplicações reais, como é o caso da estimação de parâmetros relacionados com a cauda das distribuições. Neste contexto, estamos particularmente interessados na estimação adequada do índice de cauda e de quantis de ordem elevada.

Numa primeira etapa abordamos o problema da estimação do índice de cauda e de quantis de ordem elevada. Consideramos o estimador geométrico introduzido por Brito e Freitas (2003) e apresentamos uma prova alternativa para a normalidade assintótica deste estimador de modo a facilitar a aplicação à estimação de quantis de ordem elevada e a redução do seu viés. Propomos também um estimador de quantis de ordem elevada resultante da aplicação do estimador do tipo geométrico para o índice de cauda e, sob condições apropriadas, provamos que este é assintoticamente normal.

No seguimento, investigamos cuidadosamente a questão da redução do viés dos estimadores propostos e, tendo em conta as propriedades específicas do estimador geométrico, introduzimos duas novas versões deste estimador com viés reduzido, e a respetiva normalidade assintótica é estabelecida. Uma vez que nestas versões o desempenho do estimador geométrico melhorou notoriamente, procedemos à aplicação destas na estimação de quantis e a normalidade assintótica dos estimadores resultantes é estabelecida.

Apresentamos vários estudos de simulação, através dos quais ilustramos o comportamento amostral finito de todos os estimadores propostos e efetuamos a comparação

destes com os estimadores clássicos. O estudo relativo aos intervalos de confiança assintóticos também é aqui considerado.

A aplicação na área da sismologia suscitou o nosso interesse por se ter mostrado promissora tanto a nível de facilidade de aquisição de dados como de forte impacto humanitário. Assim, motivados pela aplicação do trabalho científico realizado a casos práticos reais e a fim de realizar uma análise ao comportamento dos estimadores propostos, concentramo-nos no estudo das caudas das distribuições dos momentos sísmicos. Nesta sequência combinamos diferentes abordagens, tendo sido fornecida especial atenção à questão da independência e ao estudo inferencial.

Palavras-chave: Caudas pesadas, distribuição generalizada de Pareto, redução do viés, estimação de quantis elevados, estimação do índice de cauda, momentos sísmicos dos terramotos, teoria de valores extremos.

Abstract

In this work we are concerned with problems that stand out in Extreme Value Theory as having high relevance in real applications, such as the estimation of parameters related to the tail of the distributions. In this context, we are particularly interested in the proper estimation of the tail index and high order quantiles.

As a first step we address the problem of estimating the tail index and high order quantiles. We consider the geometric-type estimator introduced by Brito and Freitas (2003) and present an alternative proof of the asymptotic normality of this estimator in order to facilitate the application to the estimation of high order quantiles and the reduction of its bias. We also propose an estimator of high order quantiles resulting from the application of the geometric-type estimator for the tail index and, under appropriate conditions, we prove its asymptotically normal.

We then carefully investigate the issue of reducing the bias of the proposed estimators and, taking into account the specific properties of the geometric-type estimator, we introduce two new versions of this estimator with reduced bias, and the corresponding asymptotic normality is established. Since in these versions the performance of the geometric-type estimator was improved markedly, we apply it on the quantiles estimation and the asymptotic normality of the resulting estimators is established.

Several simulation studies are presented by which we illustrate the finite sample behaviour of all the proposed estimators and compare them with the classical estimators. The relative study on the asymptotic confidence intervals is also considered here.

The application in the field of seismology provoked our interest since it has shown promise both in the ease of data acquisition and as a strong humanitarian impact. Motivated by the use of the scientific work that has been accomplished in real practical cases and in order to perform an analysis on the behaviour of the proposed estimators, we therefore concentrate on the study of the tails of the seismic moment distributions. In this sequence we combine different approaches, special attention having been paid to the question of independence and to the inferential study.

Keywords: Bias reduction, earthquake seismic moments, extreme value theory, generalised Pareto distribution, heavy tails, high quantiles estimation, tail index estimation.

Contents

List of Notation and Abbreviations

F^{-1}	left continuous inverse function of F
$\overline{F}$	tail of the distribution function F
χ_q	quantile of order q
$\lfloor x \rfloor$	integer part of x
$\stackrel{D}{=}$	equality in distribution
$\xrightarrow{D}$	convergence in distribution
$\xrightarrow{P}$	convergence in probability
O	$f = O(g)$ iff $\lim_{t\to\infty} \frac{f(t)}{g(t)} = \lambda$, $\lambda \in \mathbb{R} \setminus \{0\}$
o	$f = o(g)$ iff $\lim_{t\to\infty} \frac{f(t)}{g(t)} = 0$
O_p	$V_n = O_p(a_n)$ iff $\forall \delta > 0\ \exists M_\delta > 0 : P\left[\frac{\lvert V_n \rvert}{a_n} \leq M_\delta\right] \geq 1 - \delta$, $\forall n \in \mathbb{N}$
o_p	$V_n = o_p(Y_n)$ iff $\frac{V_n}{Y_n} \xrightarrow{P} 0$, $n \to \infty$
RV_α	regularly varying function at infinity with index α
$U(x)$	reciprocal quantile function, ie, $U(x) = F^{-1}\left(1 - \frac{1}{x}\right)$
$X_{(i,n)}$	i-th order statistic
BM	Block Maxima
d.f.	distribution function or distribution functions

DA	domain of attraction
EVT	Extreme Value Theory
GEV	Generalised Extreme Value
GPD	Generalised Pareto Distribution
GT	geometric-type
i.i.d.	independent and identically distributed
iif	if and only if
ML	maximum likelihood
MM	method of moments
o.s.	order statistic or order statistics
POT	Peaks Over Thresholds
PWM	Probability Weighted Moments
r.v.	random variable or random variables

List of Figures

Chapter 3 - Geometric-type tail index estimation

Chapter 5 - Simulation study results and comparisons

Chapter 6 - Modelling extremal earthquakes

* Figures from The Good Earth: Introduction to Earth Science. McConnell et al. (2007) (Courtesy of David McConnell, David Steer, Catherine Knight, Katharine Owens and Lisa Park, with permission of McGraw-Hill Education LLC, Copyright 2008, McGraw-Hill).

List of Tables

Chapter 5 - Simulation study results and comparisons

Chapter 1

Introduction

The extreme events are present in our daily lives. Most of the time they are associated with catastrophic events and naturally its study is very important. There is a general consensus that a probabilistic approach to define most of the events gives appropriate and useful results. However, when we are dealing with extreme events, the classical statistical models are inappropriate for an adequate statistical modelling. In classical data analysis, extreme values are often called outliers and ignored for the construction of a statistical model. This happens because extreme events may not be well described by the same distribution as the central events and, due to this, they should be studied separately. Thus, in this kind of study we seek to analyse the tail of the distributions rather than the whole of the distributions.

The Extreme Value Theory (EVT) is a probabilistic theory that deals with extreme events in order to quantify and model them. Since extreme events are also rare, the EVT offers a skill to remove statistical information from the limited amount of data that characterise such phenomena. The interest in the extreme values study is shared by several areas of knowledge, since from its application result possible answers to some important problems with which they deal daily.

These areas regard the EVT as a valuable tool in solving problems of prediction of rare situations, more specifically in the prediction of the quantities related to rare events that

occur with a very low probability, such as floods, payment of high claim by insurers, materials corrosion or earthquakes. Due to the important applications in a variety of fields of scientific knowledge, the EVT is considered one of the most important study areas and has received strong attention in the recent decades. Several applications of this theory can be found in literature (see, for example, Coles (2001), Embrechts et al. (1997) and Reiss and Thomas (2007)).

When we are facing a problem related to extreme events, our main focus lies in the characterisation of the tail of the distribution that models the sample in which these events are based. Among the several problems with which the EVT deals (cf. e.g. Beirlant et al. (2012)), the adequate estimation of the tail parameters can be considered the most relevant topic given its great importance.

In this thesis, we concentrate on the particular problem of the appropriate estimation of parameters related to the tail of the distributions, more specifically the tail index and high quantiles, as well as applying it to earthquake data. We use, as a starting point, the geometric-type estimator proposed by Brito and Freitas (2003). We give our contribution presenting alternatives in the parameters estimation that have a good performance in the description and characterisation of rare phenomena occurrence. In this context we deal mainly with questions related with high quantiles, bias correction and asymptotic properties of the proposed estimators. Part of our study is devoted to describe some peculiarities of seismic phenomenon that, despite its occurrence is considered rare, often has catastrophic consequences.

The thesis is partitioned into six chapters. The Chapter 2 briefly reviews the fundamentals of EVT. An overview of the literature is presented and some basic concepts are introduced in order to reach the main limiting results of the EVT. We also consider the choice of the most appropriate modelling approach to follow and discuss some important points that one should take into account when dealing with the parameter estimation.

In Chapter 3 we introduce the geometric-type estimator and the context under which the work is developed. We study some properties of the geometric-type estimator and

use them to estimate high order quantiles. The asymptotic normality of the geometric-type estimator is shown using a method that proves to be very useful for statistical inference. The asymptotic normality of the resulting high order quantiles estimator is also established.

In order to improve the performance of the geometric-type estimator, in Chapter 4 we propose two asymptotic equivalent bias corrected estimators and study their asymptotic behaviour. Then, these two corrected estimators are used for high quantiles estimation yielding the corresponding high quantiles estimators and an asymptotic study is performed.

Chapter 5 is dedicated to the evaluation of the finite sample behaviour of the proposed estimators through some simulation studies. The suggested estimators are used to construct asymptotic confidence intervals. The results are compared with the corresponding Hill version estimators.

Chapter 6 is fully devoted to the modelling of extremal earthquakes and includes a detailed study of the used data. To approach this statistical problem we apply the Peaks Over Thresholds methodology to some real earthquake data sets in the Harvard Seismic Catalog in order to estimate the parameters quantifying the tails of the distribution of the large earthquakes considered. Then we consider the observed seismic moments of Philippines and Vanuatu Islands during 35 years. The validity of the assumptions required to use the results are investigated and both suggested geometric-type and Hill estimators are used for the estimation of the tail index and are employed on Peaks Over Threshold estimator for the quantiles estimation. A comparison between the suggested estimators is carried out and their performance is carefully discussed.

We resort to the use of R software to perform all the analysis, which are supported by graphical tools that show in a clear way the features of the data that are regarded as relevant to the study that is addressed here.

Chapter 2

Overview of Extreme Value Theory

2.1 Brief historical review

The expression “distribution of rare events” dates back to the beginning of the twentieth century. According to Falk et al. (2010), in 1922, Von Bortkiewicz introduced the concept of a maximum values distribution and also concluded that the Poisson distribution fits well to rare events. Among certain bit macabre examples, his most popular one is on the number of Prussian cavalrymen killed by friendly horse-kick over a period of 20 years, in which the annual number of fatalities are described by a Poisson distribution. Tiago de Oliveira (1990) refer that also a great contribution was given by the work of Dodd, in 1923, where the exact extreme distributions (maximums and minimums) are obtained.

After that, the most important publication was given in 1927 by Fréchet, who introduced the asymptotic distribution of the sample maximum. One year later, Fisher and Tippett (1928) showed that the distribution of the maximum (or minimum), if non-degenerate, could only belong to one of the three types: Gumbel, Fréchet or Weibull. The sufficient conditions for the weak convergence to each of this three limiting distributions was presented by von Mises in 1936. Gumbel (1941) stands out as a pioneer in calling attention to the possible application of EVT and proposing a methodology for statistical

analysis involving the EVT.

Gnedenko (1943) established the first complete result including the three types of distributions, ie, based on the type convergence theorem of Khintchine, completed the Fisher and Tippett result, yielding to one of the main limiting results in EVT, the extremal types theorem. He also showed the necessary and sufficient conditions for the weak convergence in distribution of the extreme order statistics, and with this work finalised the basic theory of the asymptotic extremal behaviour for independent and identically distributed observations.

The above mentioned results were explored and gave rise to several studies oriented to practical applications, more related to natural phenomena, in which the statistic of extreme values can assume an important role. This initial development of EVT was enhanced by the parameterisation due to von Mises (1936) and Jenkinson (1955), from which is derived that the limiting distributions Gumbel, Fréchet and Weibull, are particular cases of the Generalised Extreme Value distribution.

At the same time a new methodology, called Peaks-Over-Thresholds, was being developed. In 1970, Todorovic and Zelenhasic provided one of the reference works. An important result in this context is that the asymptotic distribution of the excesses above a threshold value can be approximated by the Generalised Pareto Distribution, a family whose special cases are the exponential, Weibull and Pareto distributions (Balkema and de Haan (1974) and Pickands (1975)). In this context, among several studies that have been carried, we may refer Hosking and Wallis (1987), Martins and Stedinger (2000), Davison and Smith (1990) and Coles (2001).

The study of EVT has received an increasingly interest in a quite diverse domains of application and important work has been developed (see e.g. Leadbetter et al. (1983), Embrechts et al. (1997), Kotz and Nadarajah (2000) and McNeil et al. (2005)).

2.2 Main theoretical results

The main purpose of the extreme value theory is to model the real extreme data set to which we have access. Since we are interested only in extreme data, our aim lies in modelling the tail of the distribution by which the data are governed.

Here we will concentrate our attention in models with heavy tails, ie, the underlying distribution being in the Fréchet domain of attraction having a positive shape parameter γ. This kind of models plays an important role in the extreme value theory and has received considerable attention in the most researches in EVT (see e.g. Csörgő and Viharos (1998)).

The results presented in this section show that if we only consider the values of the most extreme observations, then their distribution behaviour can be approximated in an asymptotic way by its limiting distribution function. Putting in another way, in EVT one tries to make an inference about the limiting behaviour of the extreme values in a dataset.

To go toward the theoretical results, we shall assume that we have a sample $X_1, X_2, \ldots, X_n$ of n independent and identically distributed (i.i.d.) random variables (r.v.) from a distribution function (d.f.) F given by

$$F(x) = P(X_i \leq x).$$

Let $X_{(1,n)} \leq X_{(2,n)} \leq \cdots \leq X_{(n,n)}$ be the corresponding ascending order statistics (o.s.) where $X_{(i,n)}$, $i = 1, \ldots, n$, denotes the i-th order statistic.

Using the regular variation theory it is possible to characterise the tail function of the d.f. F, $\overline{F} = 1 - F$, and some related functions such as the reciprocal quantile function, $U(x) = F^{-1}(1 - 1/x)$, $x \in [1, \infty]$ in which F^{-1} denotes the left continuous inverse of F, $F^{-1}(s) = inf\{y : F(y) \geq s\}$.

The function $f : \mathbb{R}^+ \to \mathbb{R}^+$ is called a **regularly varying function** at infinity with index

α, and we write $f \in RV_\alpha$, if for some $\alpha \in \mathbb{R}$ the following limit holds:

$$\lim_{t\to\infty} \frac{f(tx)}{f(t)} = x^\alpha, \qquad \text{for all } x > 0.$$

In particular, a function $f \in RV_\alpha$ is called a **slowly varying function** at infinity if $\alpha = 0$ in the above expression. If $f \in RV_\alpha$ then we can rewrite it in the following representation form

$$f(x) = x^\alpha l(x), \quad \text{for } x > 0,$$

where $l(x)$ is a slowly varying function at infinity. For a more detailed discussion about the theory of regular variation we refer to Bingham (1987).

Since we are interested in estimating the parameters of heavy-tailed distributions, we consider here that a model F is heavy-tailed if the tail function $\overline{F} \in RV_{-1/\gamma}$, that is,

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-1/\gamma}, \quad \text{for all } x > 0,$$

or equivalently,

$$1 - F(x) = x^{-1/\gamma} l(x), \quad \text{for } x > 0.$$

Once the observations that interest are the extreme ones, we will concentrate in the maximum or in the minimum of the sample. We represent the maximum of a sample with size n by $M_n = X_{(n,n)} = \max(X_1, X_2, \ldots, X_n)$ and the minimum by $m_n = X_{(1,n)} = \min(X_1, X_2, \ldots, X_n)$.

From the relation $\min(X_1, \ldots, X_n) = -\max(-X_1, \ldots, -X_n)$ between maxima and minima, the results for maxima (or minima) can be immediately derived for minima (or maxima). For this reason, we only consider the results for the maxima.

The exact distribution of the maximum M_n can be obtained by F to the power n,

$$\begin{aligned} P(M_n \le x) &= P(X_1 \le x, X_2 \le x, \cdots, X_n \le x) \\ &= F^n(x), \end{aligned}$$

for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$.

However, since the d.f. F is unknown, this result is not useful and we have to search for a distribution that serves as an approximation of F^n.

To achieve this, as we seek the limiting behaviour of the maximum, we suppose that the sample is embedded in a sequence of n r.v., with n increasing toward infinity. Since the maximum values are located near of the right endpoint of the distribution of X, this last one is related with the asymptotic behaviour of M_n. Then, if we define the left endpoint $x_F = inf\{x \in \mathbb{R} : F(x) > 1\}$ and the right endpoint $x^F = sup\{x \in \mathbb{R} : F(x) < 1\}$ of the underlying d.f. F, as a consequence of the above expression we have that $m_n \xrightarrow{P} x_F$ and $M_n \xrightarrow{P} x^F$ as $n \to \infty$, as it was expected.

Consequently, whatever the value x take, the limiting d.f. of the maximum threshold will be degenerate, ie,

$$F^n(x) \xrightarrow[n\to\infty]{} \begin{cases} 0, & \text{if } x < x^F, \\ 1, & \text{if } x \geq x^F. \end{cases}$$

In order to find the possible non-degenerate limiting distributions of the maximum M_n, and in a similar way to what is done in the Central Limit Theorem for sums of r.v., we look for appropriate normalising sequences $a_n > 0$ and b_n real, $n \geq 1$, such that $(M_n - b_n)/a_n$ converges in distribution to a non-degenerate law G, ie,

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x), \tag{2.1}$$

for each continuity point x of G.

The following result is the so-called *Extremal Types Theorem*, stated by Fisher and Tippett (1928) and established by Gnedenko (1943), which provides the three possible limiting forms for the distribution of M_n under linear normalisations.

Theorem 2.2.1 (Fisher-Tippett-Gnedenko theorem)**.** *Let $X_1, X_2, \ldots, X_n$ be i.i.d. r.v. with d.f. F and $M_n = \max(X_1, X_2, \ldots, X_n)$ denote the maximum of the n observations. If a sequence of real numbers $a_n > 0$ and b_n exists such that*

$$\lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = G(x),$$

then if G is a non degenerate d.f., it belongs to one of the following types

$$Type\ I\ :\quad \Lambda(x) = \exp\{-\exp(-x)\},\ x \in \mathbb{R};$$

$$Type\ II\ :\quad \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0, \\ \exp(-x^{-\alpha}), & x > 0; \end{cases}$$

$$Type\ III\ :\quad \Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\}, & x > 0, \\ 1, & x \geq 0; \end{cases}$$

for all continuity points of G, where $\alpha > 0$ is the shape parameter of the distribution describing the tail's behaviour of the underlying d.f. F.

The extreme value distributions of Types I, II and III are often known as Gumbel, Fréchet and Weibull families, respectively.

This theorem is the cornerstone of EVT since it derived all possible limiting non-degenerate distributions G that can appear as a limit in (2.1), solving the *extremal limiting problem*. Nevertheless, it remains to characterise the distributions F for which there exist sequences $a_n > 0$ and b_n, $n \geq 1$, such that (2.1) holds for any such specific limiting distribution, called the *domain of attraction problem*. More precisely, the problem lies in determine the necessary and sufficient conditions that must hold on the d.f. of X in order to get each one of the possible limiting forms that G can take. These conditions were established by Gnedenko (1943).

By definition, if $F^n(a_n x + b_n)$ tends to $G(x)$, it is said that F is attracted, to maxima, by G and that $a_n > 0$ and b_n are the coefficients of attraction. The *domain of attraction of G*, denoted by $DA(G)$ is then the set of d.f. that are attracted to the limiting distribution G. In this way, there are three distinct domains of attraction, one for each different limiting distribution.

The characterisation of domains of attraction using the theory of functions of regular variation allows to represent the necessary and sufficient conditions for

$F \in DA(G_\gamma)$. In particular, for heavy tailed models, ie, models in the Fréchet domain of attraction, we may write

$$F \in DA(\Phi_\alpha) \Longleftrightarrow \overline{F} \in RV_{-\alpha}.$$

The three models presented in the above theorem are distinct but related. It is possible to unify the three corresponding d.f., considering a unique shape parameter, in a global family known as von Mises-Jenkinson form or Generalised Extreme Value (GEV) distribution

$$G_\gamma(x) = \begin{cases} \exp\left(-(1+\gamma x)^{-1/\gamma}\right), & \text{for } 1+\gamma x > 0,\ \gamma \neq 0, \\ \exp\left(-\exp(-x)\right), & \text{for } x \in \mathbb{R},\ \gamma = 0, \end{cases}$$

where γ is the shape parameter, known as tail index, determining the weight of the right tail of the underlying d.f. F.

The Gumbel, Fréchet and Weibull families are particular cases and the only members of the GEV distribution: for $\gamma > 0$, G_γ is the Fréchet d.f. with $\alpha = 1/\gamma$; for $\gamma < 0$, G_γ is the Weibull d.f. with $\alpha = -1/\gamma$; and, for $\gamma = 0$, taken as a continuity limit for $\gamma \to 0$, G_γ is the Gumbel d.f..

Using the concept of domain of attraction, if the Theorem 2.2.1 holds, it is said that F belongs to the domain of attraction of the d.f. GEV, denoted by $F \in D(G_\gamma)$, $\gamma \in \mathbb{R}$. The tail behaviour of the distributions influences the shape of the distribution: the light right tailed distributions are contained in Weibull domain of attraction; the heavy right tailed distributions belong to the Fréchet domain of attraction; and the Gumbel domain of attraction contains the exponential right tailed distributions.

The three types of extremal distributions are *max-stable* distributions, ie, each one belongs to its own domain of attraction. However, there are distributions that do not belong to any domain of attraction, ie, it is not possible to found normalising sequences $a_n > 0$ and b_n in order to obtain a non degenerate d.f. G.

A great contribution to the introduction of GEV distribution is that it becomes possible to make the inference directly on the shape parameter γ, instead of having to assume the validity of one of the models (Weibull, Fréchet or Gumbel) initially.

The GEV distribution is a general limiting distribution that can be used as an approximation of F^n, as well as the normal distribution can be used as an approximation of the distribution of the sum of r.v.. In this sense, Extreme Value Theory is analogous to the Central Limit Theory, the study of partial maxima replacing that of partial sums.

Apart from this result that describes the asymptotic distribution of an o.s., such as maximum and minimum, of a sequence of r.v. that arise from an unknown distribution, it is also possible to specify the asymptotic distribution of the excesses over high thresholds.

The interest of this approach is to work with the probability that an observation exceeds the threshold u by no more than an amount x, given that this threshold is exceeded, represented by $F_{X-u|X>u}(x)$.

Given a random sample $X_1, \ldots, X_n$, we define the excesses over a threshold u as

$$\{R_i\}_{i=1}^{N_u} = \{Y_i : Y_i = X_i - u, i = 1, \ldots, N_u\},$$

where $N_u = \sharp\{i : X_i > u, i = 1, \ldots, N_u\}$ is the number of X_i which exceed the threshold u. Then, the conditional d.f. of excesses $X_i - u$ over a threshold u given that u is exceeded is defined by

$$F_{X-u|X>u}(x) = P(X - u \leq x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)},$$

for $0 \leq x < x^F - u$.

In this context, another very important result in EVT is the following theorem, due to Balkema and de Haan (1974) and Pickands (1975), which specifies the form of the limiting distribution of the excesses over a high threshold, as the threshold tends to the right endpoint.

Theorem 2.2.2 (Pickands-Balkema-de Haan theorem)**.** *Let $X_1, X_2, \ldots, X_n$ be a sample of n i.i.d. r.v. with d.f. F, x^F the right endpoint of F and $F_{X-u|X>u}(x) = P(X - u \leq x \mid X > u)$ the excess d.f. over a (high) threshold u. Then,*

$$F \in DA(G_\gamma) \text{ iff } \lim_{u \to x^F} \sup_{0 \leq x < x^F - u} \left| F_{X-u|X>u}(x) - H_{\gamma,\sigma_u}(x) \right| = 0,$$

where $H_{\gamma,\sigma_u}(x)$ represents the Generalised Pareto Distribution (GPD), given by:

$$H_{\gamma,\sigma_u}(x) = \begin{cases} 1 - \left(1 + \gamma\frac{x-u}{\sigma_u}\right)^{-1/\gamma}, & \text{for } 1 + \gamma\frac{x-u}{\sigma_u} > 0,\ \gamma \neq 0, \\ 1 - \exp(-\frac{x-u}{\sigma_u}), & \text{for } x \geq u,\ \gamma = 0, \end{cases}$$

where γ, u, $\sigma_u > 0$ are the shape, location, and scale parameter depending on threshold u, respectively.

This theorem shows that the approximation of the excesses over a threshold by a GPD holds if and only if the d.f. F belongs to the domain of attraction of some extreme value d.f., showing a relation between the distributions GEV and GPD

$$H_\gamma(x) = 1 + \log G_\gamma(x),$$

for all $x \in \mathbb{R}$ such that $1 + \log G_\gamma(x) > 0$.

The GEV and GPD d.f. have equivalent asymptotic tails and, in particular, the tail index γ is the same for both of them and independent of the selected threshold u. As suggested by this relation, the GPD also incorporates three types of distributions depending on the value of γ. For $\gamma < 0$, $\gamma = 0$ and $\gamma > 0$, the GPD, H_γ, is reduced to Beta, Exponential and (type II) Pareto d.f., respectively, corresponding to the Weibull, Gumbel and Fréchet domains of attraction to maxima.

Summarising, it turns out that if the maxima have the GEV distribution as a limiting distribution then the excesses over a high threshold are asymptotically distributed according to the GPD.

2.3 Modelling approaches

From the practical point of view, when confronted with a problem related to extreme events, whose distribution is unknown, the ultimate objective of the study will be the characterisation of the tail of the distribution that models the sample in which these events are based since it is a starting point for statistical inference. Consequently, the adequate estimation of the distribution parameters is considered very important. The

first issue one is faced with is the choice of what approach to follow in order to estimate the quantities of interest. The EVT provides several different techniques upon which one can rely on, which can be basically divided into two approaches, a parametric approach and a semi-parametric approach.

The parametric approach assumes that the sample data comes from the exact limiting distribution (GEV or GPD) of the sample extremes, ie, assumes this distribution as the exact distribution while it is only an approximation. The use of this assumption has generated some doubts among the researchers. Thus, this approach is concerned with the parameter estimation of that limit model by point estimation methods.

The semi-parametric approach makes only partly assumptions about the underlying d.f. F since only supposes that it belongs to the domain of attraction of an extreme value distribution. As described in the previous section, the domain of attraction is governed by the right tail of the underlying d.f. F and as such, is in this part of the distribution that we should focus our attention only taking into account the behaviour of the high o.s.. Thus, this approach focuses primarily on the direct estimation of the shape parameter that characterises the behaviour of the tail of the distribution in order to describe the behaviour of extreme values. Currently, the estimation of parameters of extreme events is often developed under this framework.

In order to perform a correct inference about extreme events from the accessible data, it is necessary to properly select the extreme observations, following some criterion, to which the distributions should be fitted. Within EVT framework, there are two primary methods to define such extreme observations which arise from the two main theorems of the EVT in the previous section: the Block Maxima method, also known as Gumbel's approach, and the Peaks Over Threshold method.

The **Block Maxima** (BM) method consists in dividing the data in equal size blocks with a previous determined amplitude and the maximum observation of each block is collected; the interest lies in the asymptotic study of maxima, to which the GEV distribution is fitted. In the **Peaks Over Threshold** (POT) method one select the observations that exceed a certain high threshold; the interest lies in the asymptotic

behaviour of the excesses over a high threshold, to which the GPD is fitted.

Accordingly with the data set under study, one must choose which is the most appropriate approach to adopt being aware that both methods have disadvantages. One major drawback of the BM method is that only one observation in a block is used to make an inference about the limiting distribution of the maximum, resulting in a small final sample size. That is why, in most of the cases, and whenever a sufficient large number of observations from a given sample is available, the POT approach presents more advantages. However, the POT approach also has a drawback, since one of the assumptions made in the theory is the fact that the observations need to be independent and in many natural processes there is a time dependence. This encompasses a problem, since probably the excesses are clustered, that is usually countered by using methods to identify these clusters and taking the largest value in each cluster as one observation. On other hand, the selection of an appropriate level to be considered to a given sample is another important issue to take into account. An inadequate choice of this level may seriously compromise the tail estimation.

The modelling involves the determination of the extreme value distribution that best suits each case, ie, the only issue that remains to be resolved is the parameter estimation.

2.4 Tail parameter estimation

The adequate estimation of the tail parameters is one of the most relevant topics in the EVT. Although we concentrate our investigation under a semi-parametric framework, in this section we first present some considerations about the behaviour of the most popular parametric estimators in a EVT context, and after we introduce some important estimators used in a semi-parametric approach.

2.4.1 Parametric approach

In order to estimate the unknown parameters of the extreme values models, different parametric methodologies have been proposed and all of them have advantages and disadvantages. Since the estimators are functions of the sample, their effectiveness varies depending on the sample used. In this way it is useful to know some techniques that give reasonable estimators to consider in each case.

The classical and most popular methods of estimation are the method of moments (MM) (see e.g. Hosking et al. (1985)), introduced by Karl Pearson at the end of the 19th century, and the maximum likelihood (ML) method (see e.g. Coles and Dixon (1999) and Katz et al. (2002)) developed by Fisher in 1922. However, the least squares method, which is a special case of the estimation techniques based on MM, had already been described by Gauss around 1794 and is believed to be the oldest method of estimation.

The MM consists in considering the moments of the sample equal to the corresponding moments of the population and solving the resulting system of equations in order to get the parameters to be estimated. The ML method, which maximises the probability, is the most common technique for finding estimators. An estimate obtained by this method is the most likely value for the parameter given the observed sample and is therefore, intuitively, a good choice for an estimator. For more details see e.g. Casella and Berger (1990).

Trying to culminate some problems of these methods, other methods have emerged. The Probability Weighted Moments (PWM) method was introduced by Landwehr et al. (1979) and Greenwood et al. (1979) and its application has generated quite a discussion to present a good alternative to ML method. The PWM method is a generalisation of the MM and was developed by Hosking et al. (1985) with the purpose of enabling the application of the MM for distributions with moments missing, assigning a greater weight to values of the tail of the distribution and thus solving a problem identified for ML method which weight each value of the distribution equally. The effect egalitarian

weighting of the ML method can also be solved through the Penalised Maximum Likelihood method (see e.g. Coles and Dixon (1999)). Other methods used in the parameters estimation include the Elemental Percentil Method introduced by Castillo and Hadi (1997), which is not restricted on the tail index, as well as Bayesian approaches (see e.g. Coles and Powell (1996)), which have gained more importance with the increasing of computational capacity.

When estimating the parameters of the extreme values models, one should take into consideration some important results about the behaviour of the estimators. For instance, it is known that the asymptotic normality of the ML estimate was established for $\gamma > -1/2$, the asymptotic normality of the MM estimate was showed for $\gamma < 1/2$, and that the Penalized Maximum Likelihood method will never be able to provide a shape estimate greater than one, while the PWM estimators can be calculated even when $\gamma > 1$ (cf. e.g. Diebolt et al. (2008)).

In a general way, the methods based on moments are pointed out as having better performance than the ML estimators for small samples. However, once is necessary to impose certain restrictions to its use, is considered by some authors that imposing the same restrictions to the ML estimators, they have equal or better performance. Furthermore, the extension to models including covariables is more direct in the ML method. We recall that several extensions of these methods have been proposed.

Through the use of methods such as those presented, can be inferred parametric estimators for the parameters of interest.

2.4.2 Semi-parametric approach

The semi-parametric estimators are motivated from the conditions imposed by the domains of attraction, which are naturally related to the tails of the distributions. The semi-parametric estimators of the extreme events parameters, such as the tail index and high quantiles, are then based in the upper o.s. of the associated sample. For more details see e.g. Beirlant et al. (2004) and de Haan and Ferreira (2006).

Several estimators have been developed in this context. The estimators presented in this section were chosen for historical reasons or because they are relevant for this study.

Tail index estimation

As mentioned before, the tail index γ is one of the most important parameters in EVT. This real valued parameter determinates the tail behaviour of a distribution and is directly related with the heaviness of the tail of the underlying model. Deciding the right tail weight for the distribution underlying the sample data, through a proper estimation of γ, constitutes a very important starting task in statistical inference for extreme values allowing to understand and describe the behaviour of the extreme values of a population.

Therefore, the accurate estimation of the tail index is very important not only by itself but also because of its great influence on the estimation of other relevant parameters of rare events, such as the right endpoint of the underlying d.f. F, or high quantiles. Since we can obtain estimates of the tail parameters of interest based on the estimation of the tail index, the main question lies in how to estimate γ from a finite sample $X_1, \ldots, X_n$.

For a general $\gamma \in \mathbb{R}$, Pickands (1975) proposed the following estimator

$$\widehat{P}\,(k) = \frac{1}{\log 2} \cdot \log \frac{X_{(n-\lfloor k/4 \rfloor+1,n)} - X_{(n-\lfloor k/2 \rfloor+1,n)}}{X_{(n-\lfloor k/2 \rfloor+1,n)} - X_{(n-k+1,n)}},$$

where $\lfloor x \rfloor$ denotes the integer part of x.

The asymptotic properties of this estimator are discussed in Dekkers and de Haan (1989). This estimator depends heavily on the number of o.s. used. Therefore the estimators are rather unworkable in practice for small and moderate sample sizes. Drees (1996) introduces refined Pickands estimators that suffers less from instability.

Another popular estimator for a general tail index is the moment estimator proposed by

Dekkers et al. (1989) and is given by

$$\widehat{M}\left(k\right)=M_n^{(1)}+1-\frac{1}{2}\left(1-\frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}}\right)^{-1},$$

where

$$M_n^{(j)}=\frac{1}{k}\sum_{i=1}^{k}\left(\log X_{(n-i+1,n)}-\log X_{(n-k,n)}\right)^j,\ \ j>0.$$

Its consistency and asymptotic normality was proved by Dekkers et al. (1989). This estimator generalises the classical Hill estimator, proposed for the case $\gamma > 0$ by Hill (1975), defined by

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

The asymptotic properties of Hill estimator have been much studied. Its weak consistency was proved by Mason (1982) and Deheuvels et al. (1988) showed the strong consistency. The asymptotic normality was investigated by several authors, for instance Haeusler and Teugels (1985), Csörgő and Mason (1985) and Csörgő and Viharos (1995), and proved under certain extra conditions.

The Hill estimator is also included as a particular case in the so-called kernel class of estimators for $\gamma > 0$ derived by Csörgő et al. (1985). Many other estimators have been proposed.

In this study we concentrate our attention on the geometric-type estimator for γ proposed in Brito and Freitas (2003), given by

$$\widehat{GT}\left(k\right)=\sqrt{\frac{M_n^{(2)}-\left[M_n^{(1)}\right]^2}{i_n(k)}},$$

where

$$i_n(k)=\frac{1}{k}\sum_{i=1}^{k}\log^2(n/i)-\left(\frac{1}{k}\sum_{i=1}^{k}\log(n/i)\right)^2.$$

This estimator is related to the estimators proposed by Schultze and Steinebach (1996) and arises in a natural way from a geometrical adaptation of the procedure used by these authors.

The asymptotic properties of $\widehat{GT}$ were investigated in Brito and Freitas (2003). In particular, these authors established the consistency of the estimator and proved, under general regular conditions, its asymptotic normality.

High quantiles estimation

As it is well known, the adequate estimation of extreme quantiles is a complex problem and several questions are still open. The quantiles are used in several inferential settings. In more recent years quantiles have received increased attention as a useful tool in data modelling and they have been used in a wide variety of problems in many different scientific areas. Through a quantiles study we are interested in quantifying the size of some extreme event that will only occur with a given small probability, or in the expected time until it happens. Their role is also important in the exploratory data analysis since it allows a robust statistical inference.

High quantiles, as functions of the tail index, are possibly the most important parameters of extreme events, as well as location and scale parameters. We denote by χ_{1-p} the $(1-p)$-order quantile, that is, a value such that the probability of the occurrence of an exceedance is equal to p, small.

The classical quantiles estimator for the case of heavy tails was proposed by Weissman (1978),

$$\widehat{\chi}_{1-p}^{W} = X_{(n-k,n)} \left(\frac{k}{np}\right)^{\widehat{\gamma}},$$

where $\widehat{\gamma}$ is a consistent estimator of γ.

Using general quantile techniques and the POT methodology, the POT estimator for high quantiles above the threshold $X_{(n-k,n)}$ arises naturally and is given by

$$\widehat{\chi}_{1-p}^{P} = \frac{\left(\frac{k}{np}\right)^{\widehat{\gamma}} - 1}{\widehat{\gamma}} \cdot \widehat{a}\left(\frac{n}{k}\right) + \widehat{b}\left(\frac{n}{k}\right), \quad p < \frac{k}{n}, \tag{2.2}$$

where $\widehat{\gamma}$, $\widehat{a}\left(\frac{n}{k}\right)$ and $\widehat{b}\left(\frac{n}{k}\right)$ are, respectively, suitable estimators of the shape, scale and location parameters of the Generalised Pareto Distribution.

Details about the high order quantiles estimation can be found, for example, in de Haan and Rootzén (1993), Matthys and Beirlant (2003) and McNeil and Saladin (1997).

Chapter 3

Geometric-type tail estimation

3.1 Introduction and geometric-type estimator context

We consider the problem of estimating the Pareto-tail index of a distribution function F, with tail function $\overline{F} = 1 - F \in RV_{-1/\gamma}$, $\gamma > 0$, that is,

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-1/\gamma}, \quad \text{for all } x > 0.$$

Equivalently,

$$1 - F(x) = x^{-1/\gamma} l(x) \quad \text{for } x > 0, \tag{3.1}$$

where l is a slowly varying function at infinity, that is, l satisfies the condition $l(tx)/l(x) \to 1$ as $x \to \infty$ for all $t > 0$. The condition (3.1) is equivalent to the regular variation of the quantile function $U(x) = F^{-1}(1 - 1/x)$, ie,

$$U(x) = x^{\gamma} L(x), \tag{3.2}$$

where L is a slowly varying function at infinity. In this way, the question addressed is the estimation of γ from a finite sample $X_1, \ldots, X_n$.

Let us consider $X_1, X_2, \ldots$ i.i.d. r.v. with d.f. F and let $X_{(1,n)} \leq X_{(2,n)} \leq \cdots \leq X_{(n,n)}$ denote the corresponding o.s. based on the n first observations. We also consider intermediate sequences $k = k_n$ of positive integers

$(1 \leq k < n)$, that is

$$k \to \infty, \qquad \frac{k}{n} \to 0 \quad \text{as} \quad n \to \infty. \tag{3.3}$$

We recall that the above problem of tail index estimation of a Pareto-type distribution is equivalent to the estimation of the exponential tail coefficient. Setting $Z_i = \log X_i$, $i = 1, 2, \ldots$, with X_i as above, we have

$$1 - G(z) = P(Z_1 > z) = r(z)\, e^{-Rz}, \quad z > 0, \tag{3.4}$$

where $r(z) = l(e^z)$ is a regularly varying function at infinity and $R = 1/\gamma$ is a positive constant, called exponential tail coefficient. Equivalently we have

$$G^{-1}(1-s) = -\frac{1}{R}\log s + \log \widetilde{L}(s), \quad 0 < s < 1,$$

where $\widetilde{L}(s) = L(1/s)$ is a slowly varying function at 0.

The problem of the estimation of the exponential tail coefficient has applications in a variety of domains and an overview of the existing literature is given in Schultze and Steinebach (1996).

We focus this work in the problem of estimating the tail index using a geometric-type estimator of the exponential tail coefficient R, proposed by Brito and Freitas (2003), given by

$$\widehat{R}(k) = \sqrt{\frac{\sum_{i=1}^{k} \log^2(n/i) - \frac{1}{k}\left(\sum_{i=1}^{k} \log(n/i)\right)^2}{\sum_{i=1}^{k} Z_{(n-i+1,n)}^2 - \frac{1}{k}\left(\sum_{i=1}^{k} Z_{(n-i+1,n)}\right)^2}}. \tag{3.5}$$

This estimator arises from the study of two estimators based on the least squares method introduced by Schultze and Steinebach (1996). One of these estimators was also introduced by Kratz and Resnick (1996) in an independent but equivalent way, and generalised by Beirlant et al. (2005) to the case where γ is real-valued. In general, when compared with other tail index estimators, it is reported that the estimators proposed by Schultze and Steinebach have a very good behaviour, performing better in several circumstances.

One of the interesting characteristics of the least squares estimators is the smoothness of the realisations as a function of k. It should be noted that the high variability that

some tail estimators present is not a welcome feature, since it makes more difficult the proper selection of the number of upper order statistics involved in the estimation. In this sense, the stability presented in almost all examples can be considered a prominent advantage of the least squares estimators over the classical Hill estimator, which plots often exhibit strong trends and a considerable lack of smoothness resulting in different estimates for neighbouring values of k and an extreme sensibility to the choice of the ideal k-value (see e.g. Csörgő and Viharos (1998)). On the other hand, it can be shown that the asymptotic variance of the geometric-type estimator is twice the asymptotic variance of the Hill estimator. However, given the bias presented by the Hill estimator, the asymptotic variance should not be the only criterion to be considered.

The estimators provided by Schultze and Steinebach were motivated by the fact that $-\log(1 - G(z))$, from (3.4), is approximately linear with slope R, for large z, since $z^{-1} \log r(z) \to 0$ as $z \to \infty$. It is then expected that $-\log(1 - G_n(z))$ is also approximately linear for high values of n and z, where G_n denotes the empirical d.f. associated to the random sample $Z_1, \ldots, Z_n$. It was also assumed that $r(z) \equiv c$, $\forall z > 0$, and thus

$$y := -\log(1 - G(z)) = Rz - \log c = Rz - d,$$

or equivalentely,

$$z = R^{-1}(y + d) = ay + b,$$

where $a = R^{-1}$, $b = R^{-1}d$ and $d = \log c$.

Denoting by $z_i := z_{(n-i+1,n)}$, $i = 1, \ldots, k \leq n$, the k upper o.s. of the sample $Z_1, \ldots, Z_n$, Schultze and Steinebach approximate $-\log(1 - G(z_i))$ by $y_i := -\log(1 - G_n(z_i^-)) = -\log(1 - (n - i)/n) = \log(n/i)$, obtaining that y_i is "close" to $Rz_i - d$, or equivalently, z_i is "close" to $ay_i + b$. Following this approach, one of the estimators was obtained by minimising the function $f_1(a, b) = \sum_{i=1}^{k}(z_i - ay_i - b)^2$ and the other one by minimising the function $f_2(R, d) = \sum_{i=1}^{k}(y_i - Rz_i + d)^2$, which corresponds to determining the inverse of the slope of the line by minimising the sum of the distances between the points $\{(z_i, y_i), i = 1, \ldots, k\}$ and the line, measured in horizontal or vertical, respectively.

The $\widehat{R}$ estimator is obtained through a geometrical adaptation of these two perspectives, minimising the sum of the areas of the rectangles whose sides are the horizontal and vertical segments between the points $\{(z_i, y_i), i = 1, \ldots, k\}$ and the line, in Figure 3.1, which is equivalent to minimise the function

$$f(R, d) = \sum_{i=1}^{k} (y_i - Rz_i + d)(R^{-1}y_i + R^{-1}d - z_i).$$

In this way both horizontal and vertical distances between the points $\{(z_i, y_i), i = 1, \ldots, k\}$ and the line are minimised.

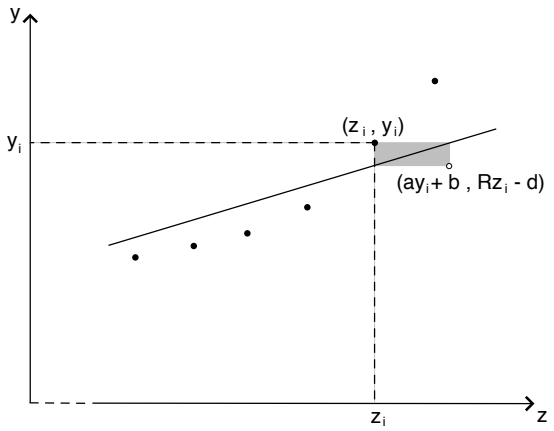


Figure 3.1: Geometric representation of the rectangles whose areas will be minimised to obtain the estimator of R.

The asymptotic properties of $\widehat{R}$ were investigated in Brito and Freitas (2003). In particular, these authors established the consistency of the estimator and proved that, under general regularity conditions, the distribution of $k^{1/2}\left(\widehat{R}\,(k) - R\right)$ is asymptotically normal. This estimator also enjoys of certain properties that makes its use specially attractive for the case where R is expected to be small (see e.g. Csörgő and Viharos (1998) and Brito and Freitas (2006)).

In the context of estimating the tail index, we will consider the following geometric-type (GT) estimator for γ:

$$\widehat{GT}\,(k) = \frac{1}{\widehat{R}\,(k)} = \sqrt{\frac{M_n^{(2)} - \left[M_n^{(1)}\right]^2}{i_n(k)}}\,. \tag{3.6}$$

where

$$i_n(k) = \frac{1}{k}\sum_{i=1}^{k}\log^2(n/i) - \left(\frac{1}{k}\sum_{i=1}^{k}\log(n/i)\right)^2 \tag{3.7}$$

and

$$M_n^{(j)}(k) = \frac{1}{k}\sum_{i=1}^{k}\left(\log X_{(n-i+1,n)} - \log X_{(n-k,n)}\right)^j. \tag{3.8}$$

The asymptotic properties of $\widehat{GT}$ arise naturally from the corresponding properties of $\widehat{R}$ studied in Brito and Freitas (2003).

To deal with the suggested problems, the procedures are formulated under second order conditions. We begin by assuming there exists a positive function a such that, for all $x > 0$,

$$\lim_{t\to\infty}\frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}. \tag{3.9}$$

From (3.2), we can choose $a(t) = \gamma U(t)$. We also suppose that there exists a function $A(t)$, tending to zero as $t \to \infty$, such that

$$\lim_{t\to\infty}\frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma\frac{x^\rho - 1}{\rho}, \tag{3.10}$$

for all $x > 0$, where $\rho < 0$ is the shape parameter governing the rate of convergence of $U(tx)/U(t)$ to x^γ and the function $|A(t)| \in RV_\rho$ (see e.g. Geluk and de Haan (1987)).

3.2 Asymptotic properties of the geometric-type estimator

Here the asymptotic normality of the geometric-type estimator is shown using a method that proves to be very useful for statistical inference, in particular for bias treatment. We first derive the asymptotic distributional representation of the geometric-type estimator.

Since $i_n(k) \to 1$ as $n \to \infty$, we begin by considering the asymptotic normality of the following tail index estimator of γ

$$\widetilde{\gamma}(k) = \sqrt{M_n^{(2)} - \left[M_n^{(1)}\right]^2}.$$

Theorem 3.2.1. *Assume* (3.10) *holds. For sequences* k *such that* (3.3) *holds, we have the following asymptotic distributional representation*

$$\widetilde{\gamma}(k) \stackrel{D}{=} \gamma + \frac{\gamma}{2\sqrt{k}} Q_n - \frac{\gamma}{\sqrt{k}} P_n + \frac{A\left(\frac{n}{k}\right)}{(1-\rho)^2} + o_p\left(A\left(\frac{n}{k}\right)\right) + O_p\left(\frac{1}{k}\right),$$

where $P_n = \sqrt{k}\left(\sum_{i=1}^{k} Z_i/k - 1\right)$ *and* $Q_n = \sqrt{k}\left(\sum_{i=1}^{k} Z_i^2/k - 2\right)$, (P_n, Q_n) *is asymptotically normal with mean equal to* $\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$ *and covariance matrix* $\left(\begin{smallmatrix} 1 & 4 \\ 4 & 20 \end{smallmatrix}\right)$*, and* $\{Z_i\}$ *denote i.i.d. standard exponential r.v..*

For the proof of Theorem 3.2.1 we need the following Lemma.

Lemma 3.2.2 (Dekkers and de Haan (1993), Lemma 3.1)**.** *Let* $Y_{(1,n)} \leq Y_{(2,n)} \leq \cdots \leq Y_{(n,n)}$ *denote the o.s. based on the* n *first observations of the sequence* $Y_1, \ldots, Y_n$ *of i.i.d. r.v. with common d.f.* $1 - 1/x$ $(x > 1)$*. Let* k *be such that* (3.3) *holds. For* $\gamma > 0$*, define*

$$T_n = \sqrt{k}\left\{\frac{1}{k}\sum_{i=1}^{k} \log Y_{(n-i+1,n)} - \log Y_{(n-k,n)} - 1\right\},$$

$$V_n = \sqrt{k}\left\{\frac{1}{k}\sum_{i=1}^{k} \left(\log Y_{(n-i+1,n)} - \log Y_{(n-k,n)}\right)^2 - 2\right\}.$$

Then (T_n, V_n) *is asymptotically normal with mean equal to* $\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$ *and covariance matrix* $\left(\begin{smallmatrix} 1 & 4 \\ 4 & 20 \end{smallmatrix}\right)$.

Proof of Theorem 3.2.1. Note that the condition (3.10) is equivalent to

$$\lim_{t\to\infty} \frac{\log U(tx) - \log U(t) - \gamma \log x}{A(t)} = \frac{x^\rho - 1}{\rho}.$$

Consequently we have

$$\log U(tx) - \log U(t) = \gamma \log x + A(t)\frac{x^\rho - 1}{\rho}(1 + o(1))$$

and

$$(\log U(tx) - \log U(t))^2 = (\gamma \log x)^2 + 2\gamma \frac{x^\rho - 1}{\rho}(\log x) A(t) + o(A(t)),$$

as $t \to \infty$.

Let us consider the variables presented in Lemma 3.2.2.

Since $\left(X_{(1,n)}, X_{(2,n)}, \cdots, X_{(n,n)}\right) \stackrel{D}{=} \left(U\left(Y_{(1,n)}\right), U\left(Y_{(2,n)}\right), \cdots, U\left(Y_{(n,n)}\right)\right)$, without loss of generality we can write $X_{(i,n)} = U\left(Y_{(i,n)}\right)$.

Then,

$$\begin{aligned} M_n^{(1)} &= \frac{1}{k}\sum_{i=1}^{k} \log X_{(n-i+1,n)} - \log X_{(n-k,n)} \\ &= \frac{1}{k}\sum_{i=1}^{k} \log U\left(\frac{Y_{(n-i+1,n)}}{Y_{(n-k,n)}} Y_{(n-k,n)}\right) - \log U\left(Y_{(n-k,n)}\right) \\ &= \gamma + \frac{\gamma}{\sqrt{k}} T_n + \frac{A\left(Y_{(n-k,n)}\right)}{1-\rho} + o_p\left(A\left(Y_{(n-k,n)}\right)\right). \end{aligned}$$

The last equality follows from regular variation properties and Potter bounds (cf. the proof of Theorem 3.2.5 of de Haan and Ferreira (2006)), using in particular that

$$\frac{1}{k}\sum_{i=1}^{k}\left(\frac{\left(\frac{Y_{(n-i+1,n)}}{Y_{(n-k,n)}}\right)^{\rho} - 1}{\rho}\right) \stackrel{D}{=} \frac{1}{k}\sum_{i=1}^{k}\left(\frac{Y_i^{\rho} - 1}{\rho}\right),$$

which tends to $E\left(\frac{Y_1^{\rho}-1}{\rho}\right) = \frac{1}{1-\rho}$.

We have also

$$\begin{aligned} M_n^{(2)} &= \frac{1}{k}\sum_{i=1}^{k}\left[\log X_{(n-i+1,n)} - \log X_{(n-k,n)}\right]^2 \\ &= 2\gamma^2 + \frac{\gamma^2}{\sqrt{k}} V_n + A\left(Y_{(n-k,n)}\right)\frac{2\gamma\left(2-\rho\right)}{\left(1-\rho\right)^2} + o_p\left(A\left(Y_{(n-k,n)}\right)\right), \end{aligned}$$

using in particular that

$$\frac{1}{k}\sum_{i=1}^{k}\left(\frac{\left(\frac{Y_{(n-i+1,n)}}{Y_{(n-k,n)}}\right)^{\rho} - 1}{\rho}\log\frac{Y_{(n-i+1,n)}}{Y_{(n-k,n)}}\right) \stackrel{D}{=} \frac{1}{k}\sum_{i=1}^{k}\left(\frac{Y_i^{\rho} - 1}{\rho}\log Y_i\right),$$

which tends to $E\left(\frac{Y_1^{\rho}-1}{\rho}\log Y_1\right) = \frac{2-\rho}{(1-\rho)^2}$.

Considering $h(x) = x^2$ and the Taylor expansion of $h\left(M_n^{(1)}\right)$ around γ we obtain

$$\left[M_n^{(1)}\right]^2 = \gamma^2 + \frac{2\gamma^2}{\sqrt{k}} T_n + \frac{2\gamma}{1-\rho} A\left(Y_{(n-k,n)}\right) + o_p\left(A\left(Y_{(n-k,n)}\right)\right) + O_p\left(\frac{1}{k}\right).$$

Using the above representations we obtain

$$\begin{aligned}\widetilde{\gamma}^2(k) &= M_n^{(2)} - \left[M_n^{(1)}\right]^2 \\ &= \gamma^2 + \frac{\gamma^2}{\sqrt{k}}(V_n - 2T_n) + d\, A\left(Y_{(n-k,n)}\right) + o_p\left(A\left(Y_{(n-k,n)}\right)\right) + O_p\left(\frac{1}{k}\right),\end{aligned}$$

where $d = 2\gamma/(1-\rho)^2$.

Since $A(t) \in RV_\rho$, then $A(tx) = x^\rho A(t)(1 + o(1))$.

Noting that $(k/n)\, Y_{(n-k,n)} = 1 + o_p(1)$, we have

$$\begin{aligned}A\left(Y_{(n-k,n)}\right) &= A\left(\frac{n}{k}(1 + o_p(1))\right) \\ &= A\left(\frac{n}{k}\right) + o_p\left(A\left(\frac{n}{k}\right)\right).\end{aligned}$$

Therefore, we may write

$$\widetilde{\gamma}^2(k) = \gamma^2 + \frac{\gamma^2}{\sqrt{k}}(V_n - 2T_n) + d\, A\left(\frac{n}{k}\right) + o_p\left(A\left(\frac{n}{k}\right)\right) + O_p\left(\frac{1}{k}\right).$$

Considering $g(x) = \sqrt{x}$ and the Taylor expansion of $g\left(\widetilde{\gamma}^2(k)\right)$ around γ^2 we obtain

$$\begin{aligned}\widetilde{\gamma}(k) &= \gamma + \frac{1}{2\gamma}\left[\frac{\gamma^2}{\sqrt{k}}(V_n - 2T_n) + d\, A\left(\frac{n}{k}\right) + o_p\left(A\left(\frac{n}{k}\right)\right) + O_p\left(\frac{1}{k}\right)\right] \\ &= \gamma + \frac{\gamma}{2\sqrt{k}}V_n - \frac{\gamma}{\sqrt{k}}T_n + \frac{A\left(\frac{n}{k}\right)}{(1-\rho)^2} + o_p\left(A\left(\frac{n}{k}\right)\right) + O_p\left(\frac{1}{k}\right).\end{aligned} \tag{3.11}$$

Recall that $\log\left(Y_{(n-i+1,n)}/Y_{(n-k,n)}\right)$ are exponential standard r.v., $Exp(1)$. Using Lemma 3.2.2, from (3.11) we can write

$$\widetilde{\gamma}(k) \overset{D}{=} \gamma + \frac{\gamma}{2\sqrt{k}}Q_n - \frac{\gamma}{\sqrt{k}}P_n + \frac{A\left(\frac{n}{k}\right)}{(1-\rho)^2} + o_p\left(A\left(\frac{n}{k}\right)\right) + O_p\left(\frac{1}{k}\right),$$

where $P_n = \sqrt{k}\left(\sum_{i=1}^k Z_i/k - 1\right)$, $Q_n = \sqrt{k}\left(\sum_{i=1}^k Z_i^2/k - 2\right)$, with $\{Z_i\}$ i.i.d. exponential standard r.v., are jointly asymptotic normal.

This completes the proof. □

Corollary 3.2.3. *Assume the conditions of Theorem 3.2.1 hold. If k is such that $\sqrt{k}A(n/k) \to \lambda$ finite, then*

$$\sqrt{k}\left(\widetilde{\gamma}(k) - \gamma\right)$$

is asymptotically normal distributed as $n \to \infty$ *with variance* $2\gamma^2$ *and a non-null mean value given by* $\lambda/(1-\rho)^2$*.*

Proof of Corollary 3.2.3. As $(P_n, Q_n) \xrightarrow{D} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 4 \\ 4 & 20 \end{pmatrix}\right]$,

$$V\left(\sqrt{k}\left(\widetilde{\gamma}\left(k\right)-\gamma\right)\right) \approx V\left(\frac{\gamma}{2}Q_n\right) + V\left(\gamma P_n\right) - 2Cov\left(\frac{\gamma}{2}Q_n, \gamma P_n\right) \xrightarrow[n\to\infty]{} 2\gamma^2.$$

The result follows from the proof of Theorem 3.2.1. □

Theorem 3.2.4. *Assume* (3.10) *holds. For sequences* k *such that* (3.3) *holds, we have the following asymptotic distributional representation*

$$\widehat{GT}\left(k\right) \overset{D}{=} \gamma + \frac{\gamma}{2\sqrt{k}}Q_n - \frac{\gamma}{\sqrt{k}}P_n + \frac{A\left(\frac{n}{k}\right)}{\left(1-\rho\right)^2} + o_p\left(A\left(\frac{n}{k}\right)\right) + O_p\left(\frac{\log^2 k}{k}\right),$$

where $P_n = \sqrt{k}\left(\sum_{i=1}^{k} Z_i/k - 1\right)$ *and* $Q_n = \sqrt{k}\left(\sum_{i=1}^{k} Z_i^2/k - 2\right)$*,* (P_n, Q_n) *is asymptotically normal with mean equal to* $\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$ *and covariance matrix* $\left(\begin{smallmatrix} 1 & 4 \\ 4 & 20 \end{smallmatrix}\right)$*, and* $\{Z_i\}$ *denote i.i.d. standard exponential r.v..*

For proving Theorem 3.2.4 we use the following auxiliary Lemma.

Lemma 3.2.5 (Brito and Freitas (2003), Lemma 2)**.** *Let* k *be a sequence of positive integers such that* $1 \le k \le n$*. For the sequence* $i_n\left(k\right)$ *defined in* (3.7) *we have*

$$i_n\left(k\right) = 1 + O\left(\frac{\log^2 k}{k}\right).$$

Proof of Theorem 3.2.4. We recall that

$$\widehat{GT}\left(k\right) = \frac{\widetilde{\gamma}\left(k\right)}{\sqrt{i_n\left(k\right)}}.$$

Note now that we can write

$$\sqrt{k}\left(\widehat{GT}\left(k\right)-\gamma\right) = \sqrt{k}\left(\widetilde{\gamma}\left(k\right)-\gamma\right) + \sqrt{k}\widetilde{\gamma}\left(k\right)\left(\frac{1}{\sqrt{i_n\left(k\right)}} - 1\right).$$

As $\widetilde{\gamma}\left(k\right) \xrightarrow[n\to\infty]{P} \gamma$, from Lemma 3.2.5 we get

$$\sqrt{k}\widetilde{\gamma}\left(k\right)\left(\frac{1}{\sqrt{i_n\left(k\right)}} - 1\right) = O_p\left(\frac{\log^2 k}{\sqrt{k}}\right).$$

So, from the proof of Theorem 3.2.1 and Lemma 3.2.5, we have

$$\widehat{GT}\,(k) = \gamma + \frac{\gamma}{2\sqrt{k}}V_n - \frac{\gamma}{\sqrt{k}}T_n + \frac{A\left(\frac{n}{k}\right)}{(1-\rho)^2} + o_p\left(A\left(\frac{n}{k}\right)\right) + O_p\left(\frac{\log^2 k}{k}\right),$$

where T_n and V_n are the same as in proof of Theorem 3.2.1 and the result follows. □

Corollary 3.2.6. *Assume the conditions of Theorem 3.2.4 hold. If k is such that $\sqrt{k}A(n/k) \to \lambda$ finite, then*

$$\sqrt{k}\left(\widehat{GT}\,(k) - \gamma\right)$$

is asymptotically normal distributed as $n \to \infty$ with variance $2\gamma^2$ and a non-null mean value given by $\lambda/(1-\rho)^2$.

Proof of Corollary 3.2.6. By Theorem 3.2.4, the result is established in a similar way to the proof of Corollary 3.2.3. □

3.3 High order quantiles estimation using the geometric-type estimator

The information given by a proper study of quantiles is very important and useful for the many different areas of knowledge. Such a study enables to use a probabilistic approach to characterise events with an extremely rare occurrence. Then, the adequate estimation of high order quantiles is one of the most important problems in statistics. In this way, we are interested in the estimation of a high quantile $\chi_{1-p} = U(1/p)$, a value exceeded with a small probability, ie, such that $F(\chi_{1-p}) = 1 - p$. The estimation of high quantiles has been considered by several authors (see e.g. de Haan and Rootzén (1993)). Here, we use (2.2) applying the geometric-type estimator as an estimator of γ, $X_{(n-k,n)}M_n^{(1)}$ and $X_{(n-k,n)}$ as suitable estimators of the scale and location parameters, and the asymptotic normality of the resulting estimator is established.

We consider the following estimator for high quantiles

$$\widehat{\chi}_{1-p}^{\widehat{GT}} = \frac{\left(\frac{k}{np}\right)^{\widehat{GT}(k)} - 1}{\widehat{GT}\,(k)} \cdot X_{(n-k,n)}M_n^{(1)} + X_{(n-k,n)}, \tag{3.12}$$

for some $p = p_n \to 0$ as $n \to \infty$, with np small. We begin by showing the following theorem.

Note that $a(n/k)$, γ and $U(n/k)$, in (3.9), are the scale, shape and location parameters of the Generalised Pareto Distribution.

Theorem 3.3.1. *Suppose that for some function A with $A(t) \to 0$ as $t \to \infty$, the condition* (3.10) *holds and $\sqrt{k}A(n/k) \to \lambda$ finite, as $n \to \infty$. Let k be such that* (3.3) *holds. Then*

$$\sqrt{k}\left(\frac{X_{(n-k,n)}M_n^{(1)}}{a\left(\frac{n}{k}\right)} - 1, \widehat{GT}\left(k\right) - \gamma, \frac{X_{(n-k,n)} - U\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)}\right) \xrightarrow[n\to\infty]{D} \left(\Lambda, \Gamma, B\right),$$

where (Λ, Γ, B) are jointly normal r.v. with mean vector $\left(\lambda/\left(\gamma - \gamma\rho\right), \lambda/\left(1-\rho\right)^2, 0\right)^\top$ and covariance matrix

$$\begin{pmatrix} 1+\gamma^2 & \gamma & \gamma \\ \gamma & 2\gamma^2 & 0 \\ \gamma & 0 & 1 \end{pmatrix}. \tag{3.13}$$

Proof of Theorem 3.3.1. The result follows by using the classic methodology of Central Limit Theory. By Cramér-Wold Device, it is enough to prove that every linear combination of its three components is asymptotically normal, ie,

$$t_1\sqrt{k}\left(\frac{X_{(n-k,n)}M_n^{(1)}}{a\left(\frac{n}{k}\right)} - 1\right) + t_2\sqrt{k}\left(\widehat{GT}\left(k\right) - \gamma\right) + t_3\sqrt{k}\left(\frac{X_{(n-k,n)} - U\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)}\right) \tag{3.14}$$

has an asymptotically normal distribution with $t_1, t_2, t_3 \in \mathbb{R}$.

Let us consider the r.v. $Y_{(i,n)}$, V_n and T_n as in Lemma 3.2.2.

Setting $a\left(\frac{n}{k}\right) = \gamma U\left(\frac{n}{k}\right)$ without loss of generality, then we have that (3.14) is equal to

$$t_1\sqrt{k}\left(\frac{X_{(n-k,n)}M_n^{(1)}}{\gamma U\left(\frac{n}{k}\right)} - 1\right) + t_2\sqrt{k}\left(\widehat{GT}\left(k\right) - \gamma\right) + t_3\sqrt{k}\left(\frac{X_{(n-k,n)} - U\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)}\right).$$

Considering now the condition (3.10), we have

$$\frac{\frac{U(tx)}{U(t)} - x^\gamma}{A\left(t\right)} = x^\gamma \frac{x^\rho - 1}{\rho}\left(1 + o\left(1\right)\right),$$

that we rewrite as

$$\frac{\frac{U\left(Y_{(n-k,n)}\right)}{U\left(\frac{n}{k}\right)}-\left(\frac{k}{n}Y_{(n-k,n)}\right)^{\gamma}}{A\left(\frac{n}{k}\right)}=\left(\frac{k}{n}Y_{(n-k,n)}\right)^{\gamma}\frac{\left(\frac{k}{n}Y_{(n-k,n)}\right)^{\rho}-1}{\rho}\left(1+o_p\left(1\right)\right).$$

We write again without loss of generality $X_{(i,n)}=U\left(Y_{(i,n)}\right)$. Then

$$\frac{X_{(n-k,n)}}{U\left(\frac{n}{k}\right)}=\left(\frac{k}{n}Y_{(n-k,n)}\right)^{\gamma}\left[1+\frac{\left(\frac{k}{n}Y_{(n-k,n)}\right)^{\rho}-1}{\rho}A\left(\frac{n}{k}\right)\left(1+o_p\left(1\right)\right)\right].$$

Once $x^{\alpha}=1+\alpha\left(x-1\right)+o\left(x-1\right)$ as $x\to 1$ and $\sqrt{k}A(n/k)$ converges to λ as $n\to\infty$, we have

$$\frac{X_{(n-k,n)}}{U\left(\frac{n}{k}\right)}=1+\gamma\left(\frac{k}{n}Y_{(n-k,n)}-1\right)+o_p\left(A\left(\frac{n}{k}\right)\right).$$

Thus, we have the following property

$$\frac{X_{(n-k,n)}}{U\left(\frac{n}{k}\right)}=1+\frac{\gamma}{\sqrt{k}}B_n+o_p\left(A\left(\frac{n}{k}\right)\right),$$

where $B_n=\sqrt{k}\left(\frac{k}{n}Y_{(n-k,n)}-1\right)$ is an asymptotically standard normal r.v. (cf. e.g. Smirnov (1967)).

Moreover, by the proof of Theorem 3.2.1, and since $\sqrt{k}A(n/k)\to\lambda$ as $n\to\infty$,

$$M_n^{(1)}=\gamma+\frac{\gamma}{\sqrt{k}}T_n+\frac{A\left(Y_{(n-k,n)}\right)}{1-\rho}+o_p\left(A\left(Y_{(n-k,n)}\right)\right)$$

and

$$\sqrt{k}\left(\widetilde{\gamma}\left(k\right)-\gamma\right)=\frac{\gamma}{2}V_n-\gamma T_n+\frac{\lambda}{\left(1-\rho\right)^2}+o_p\left(1\right).$$

Then we can write

$$\begin{aligned}\sqrt{k}\left(\frac{X_{(n-k,n)}M_n^{(1)}}{a\left(\frac{n}{k}\right)}-1\right)&=\sqrt{k}\left(\frac{X_{(n-k,n)}}{U\left(\frac{n}{k}\right)}\cdot\frac{M_n^{(1)}}{\gamma}-1\right)\\&=\sqrt{k}\left(\frac{M_n^{(1)}}{\gamma}-1\right)+\sqrt{k}\left(\frac{M_n^{(1)}}{\gamma}\frac{\gamma}{\sqrt{k}}B_n+\frac{M_n^{(1)}}{\gamma}o_p\left(A\left(\frac{n}{k}\right)\right)\right)\\&=T_n+\gamma B_n+\frac{\lambda}{\gamma\left(1-\rho\right)}+o_p\left(1\right),\end{aligned}\tag{3.15}$$

since $\sqrt{k}A\left(\frac{n}{k}\right) \to \lambda$ as $n \to \infty$ and $M_n^{(1)} = \gamma + o_p(1)$.

From the proof of Theorem 3.2.4 we can write

$$\sqrt{k}\left(\widehat{GT}(k) - \gamma\right) = \frac{\gamma}{2}V_n - \gamma T_n + \frac{\lambda}{(1-\rho)^2} + o_p(1). \tag{3.16}$$

Consider now the third component. We have

$$\begin{aligned}\sqrt{k}\left(\frac{X_{(n-k,n)} - U\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)}\right) &= \frac{\sqrt{k}}{\gamma}\left(\frac{X_{(n-k,n)}}{U\left(\frac{n}{k}\right)} - 1\right) \\ &= \frac{\sqrt{k}}{\gamma}\left[\frac{\gamma}{\sqrt{k}}B_n + o_p\left(A\left(\frac{n}{k}\right)\right)\right] \\ &= B_n + o_p(1).\end{aligned} \tag{3.17}$$

By (3.15), (3.16) and (3.17), we may now rewrite (3.14) as

$$\begin{aligned}&t_1\sqrt{k}\left(\frac{X_{(n-k,n)}M_n^{(1)}}{a\left(\frac{n}{k}\right)} - 1\right) + t_2\sqrt{k}\left(\widehat{GT}(k) - \gamma\right) + t_3\sqrt{k}\left(\frac{X_{(n-k,n)} - U\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)}\right) \\ &= (t_1 - t_2\gamma)T_n + t_2\frac{\gamma}{2}V_n + (t_1\gamma + t_3)B_n + t_1\frac{\lambda}{\gamma(1-\rho)} + t_2\frac{\lambda}{(1-\rho)^2} + o_p(1).\end{aligned}$$

Since T_n and V_n are as in Lemma 3.2.2, $Z_i := \log Y_{(n-i+1,n)} - \log Y_{(n-k,n)}$ are i.i.d exponential standard r.v. and $B_n = \sqrt{k}\left(\frac{k}{n}Y_{(n-k,n)} - 1\right)$ is asymptotically normal and independent of T_n and V_n, we have that (3.14) is asymptotically normal.

Now we are going to compute the asymptotic mean and variance of

$$\sqrt{k}\left(\frac{X_{(n-k,n)}M_n^{(1)}}{a\left(\frac{n}{k}\right)} - 1, \widehat{GT}(k) - \gamma, \frac{X_{(n-k,n)} - U\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)}\right).$$

By (3.15) we have

$$E(T_n) + \gamma E(B_n) + E\left(\frac{\lambda}{\gamma(1-\rho)}\right) \xrightarrow[n\to\infty]{} \frac{\lambda}{\gamma(1-\rho)}$$

and

$$V(T_n) + \gamma^2 V(B_n) - 2Cov(T_n, \gamma B_n) \xrightarrow[n\to\infty]{} 1 + \gamma^2.$$

By (3.16) we have

$$\frac{\gamma}{2}E(V_n) - \gamma E(T_n) + E\left(\frac{\lambda}{(1-\rho)^2}\right) \xrightarrow[n\to\infty]{} \frac{\lambda}{(1-\rho)^2}$$

and

$$\frac{\gamma^2}{4}V\left(V_n\right)+\gamma^2 V\left(T_n\right)-2Cov\left(\frac{\gamma}{2}V_n,\gamma T_n\right)\xrightarrow[n\to\infty]{}2\gamma^2.$$

By (3.17) we have

$$E\left(B_n\right)\xrightarrow[n\to\infty]{}0$$

and

$$V\left(B_n\right)\xrightarrow[n\to\infty]{}1.$$

By (3.15) and (3.16) and using the fact that B_n is independent of T_n and V_n, we have

$$\begin{aligned}&E\left[\left(T_n+\gamma B_n+\frac{\lambda}{\gamma\left(1-\rho\right)}\right)\cdot\left(\frac{\gamma}{2}V_n-\gamma T_n+\frac{\lambda}{\left(1-\rho\right)^2}\right)\right]=\\&=E\left(\frac{\gamma}{2}T_nV_n-\gamma T_n^2+\frac{\gamma^2}{2}B_nV_n-\gamma^2B_nT_n\right)\\&=\frac{\gamma}{2}Cov\left(T_n,V_n\right)-\gamma V\left(T_n\right)\xrightarrow[n\to\infty]{}2\gamma-\gamma=\gamma.\end{aligned}$$

By (3.15) and (3.17) and using the fact that B_n is independent of T_n, we obtain

$$E\left[\left(T_n+\gamma B_n+\frac{\lambda}{\gamma\left(1-\rho\right)}\right)\cdot\left(B_n\right)\right]=E\left(T_nB_n+\gamma B_n^2\right)\xrightarrow[n\to\infty]{}\gamma.$$

By (3.16) and (3.17) and using the fact that B_n is independent of T_n and V_n, we have

$$E\left[\left(\frac{\gamma}{2}V_n-\gamma T_n+\frac{\lambda}{\left(1-\rho\right)^2}\right)\cdot\left(B_n\right)\right]=E\left(\frac{\gamma}{2}V_nB_n-\gamma T_nB_n\right)\xrightarrow[n\to\infty]{}0.$$

□

Using the previous result, the asymptotic normality of the quantiles estimator can be established.

Theorem 3.3.2. *Suppose that for some function A with $A(t)\to 0$ as $t\to\infty$ such that $\sqrt{k}A(n/k)\to\lambda$ finite, the conditions* (3.10) *and* (3.3) *holds, $np=o(k)$ and $\log(np)=o(\sqrt{k})$ as $n\to\infty$. Then,*

$$\sqrt{k}\frac{\widehat{\chi}_{1-p}^{\widehat{GT}}-\chi_{1-p}}{a\left(\frac{n}{k}\right)\int_1^{\frac{k}{np}}s^{\gamma-1}\log s\,ds}\xrightarrow[n\to\infty]{D}N\left(\frac{\lambda}{\left(1-\rho\right)^2},2\gamma^2\right),\tag{3.18}$$

where $\widehat{\chi}_{1-p}^{\widehat{GT}}$ is the quantiles estimator based on the geometric-type estimator defined in (3.12).

Proof of Theorem 3.3.2. First we begin to write

$$\begin{aligned}
\widehat{\chi}_{1-p}^{\widehat{GT}} - \chi_{1-p} &= \frac{\left(\frac{k}{np}\right)^{\widehat{GT}} - 1}{\widehat{GT}} \cdot X_{(n-k,n)} M_n^{(1)} + X_{(n-k,n)} - \chi_{1-p} \\
&= \frac{\left(\frac{k}{np}\right)^{\widehat{GT}} - 1}{\widehat{GT}} \cdot X_{(n-k,n)} M_n^{(1)} + X_{(n-k,n)} - U\left(\frac{1}{p}\right) - U\left(\frac{n}{k}\right) + U\left(\frac{n}{k}\right) \\
&= X_{(n-k,n)} - U\left(\frac{n}{k}\right) + \frac{\left(\frac{k}{np}\right)^{\widehat{GT}} - 1}{\widehat{GT}} \cdot X_{(n-k,n)} M_n^{(1)} + \frac{\left(\frac{k}{np}\right)^{\gamma} - 1}{\gamma} \cdot X_{(n-k,n)} M_n^{(1)} \\
&- \frac{\left(\frac{k}{np}\right)^{\gamma} - 1}{\gamma} \cdot X_{(n-k,n)} M_n^{(1)} + \frac{\left(\frac{k}{np}\right)^{\gamma} - 1}{\gamma} \cdot a\left(\frac{n}{k}\right) - \frac{\left(\frac{k}{np}\right)^{\gamma} - 1}{\gamma} \cdot a\left(\frac{n}{k}\right) - \left[U\left(\frac{1}{p}\right) - U\left(\frac{n}{k}\right)\right] \\
&= X_{(n-k,n)} - U\left(\frac{n}{k}\right) + X_{(n-k,n)} M_n^{(1)} \left[\frac{\left(\frac{k}{np}\right)^{\widehat{GT}} - 1}{\widehat{GT}} - \frac{\left(\frac{k}{np}\right)^{\gamma} - 1}{\gamma}\right] \\
&+ \frac{\left(\frac{k}{np}\right)^{\gamma} - 1}{\gamma} \left[X_{(n-k,n)} M_n^{(1)} - a\left(\frac{n}{k}\right)\right] - \left[U\left(\frac{1}{p}\right) - U\left(\frac{n}{k}\right) - \frac{\left(\frac{k}{np}\right)^{\gamma} - 1}{\gamma} \cdot a\left(\frac{n}{k}\right)\right].
\end{aligned}$$

Hence we obtain

$$\begin{aligned}
&\sqrt{k} \frac{\widehat{\chi}_{1-p}^{\widehat{GT}} - \chi_{1-p}}{a\left(\frac{n}{k}\right) \int_1^{\frac{k}{np}} s^{\gamma-1} \log s \, ds} = \\
&= \sqrt{k} \frac{X_{(n-k,n)} - U\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right) \int_1^{\frac{k}{np}} s^{\gamma-1} \log s \, ds} + \sqrt{k} \frac{X_{(n-k,n)} M_n^{(1)} \left[\frac{\left(\frac{k}{np}\right)^{\widehat{GT}} - 1}{\widehat{GT}} - \frac{\left(\frac{k}{np}\right)^{\gamma} - 1}{\gamma}\right]}{a\left(\frac{n}{k}\right) \int_1^{\frac{k}{np}} s^{\gamma-1} \log s \, ds} \\
&+ \sqrt{k} \frac{\frac{\left(\frac{k}{np}\right)^{\gamma} - 1}{\gamma} \left[X_{(n-k,n)} M_n^{(1)} - a\left(\frac{n}{k}\right)\right]}{a\left(\frac{n}{k}\right) \int_1^{\frac{k}{np}} s^{\gamma-1} \log s \, ds} - \sqrt{k} \frac{U\left(\frac{1}{p}\right) - U\left(\frac{n}{k}\right) - \frac{\left(\frac{k}{np}\right)^{\gamma} - 1}{\gamma} a\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right) \int_1^{\frac{k}{np}} s^{\gamma-1} \log s \, ds} \\
&= \underbrace{\sqrt{k} \frac{X_{(n-k,n)} - U\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)} \cdot \frac{1}{\int_1^{\frac{k}{np}} s^{\gamma-1} \log s \, ds}}_{\text{PART I}} + \underbrace{\frac{X_{(n-k,n)} M_n^{(1)}}{a\left(\frac{n}{k}\right)} \left[\frac{\sqrt{k}}{\int_1^{\frac{k}{np}} s^{\gamma-1} \log s \, ds} \cdot \left(\frac{\left(\frac{k}{np}\right)^{\widehat{GT}} - 1}{\widehat{GT}} - \frac{\left(\frac{k}{np}\right)^{\gamma} - 1}{\gamma}\right)\right]}_{\text{PART II}} \\
&\underbrace{+ \sqrt{k} \left(\frac{X_{(n-k,n)} M_n^{(1)}}{a\left(\frac{n}{k}\right)} - 1\right) \frac{\left(\frac{k}{np}\right)^{\gamma} - 1}{\gamma \int_1^{\frac{k}{np}} s^{\gamma-1} \log s \, ds}}_{\text{PART III}} \underbrace{- \frac{\sqrt{k}}{\int_1^{\frac{k}{np}} s^{\gamma-1} \log s \, ds} \left[\frac{U\left(\frac{1}{p}\right) - U\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)} - \frac{\left(\frac{k}{np}\right)^{\gamma} - 1}{\gamma}\right]}_{\text{PART IV}}.
\end{aligned}$$

Now we are going to show that $\text{PART I} \xrightarrow{P} 0$, $\text{PART II} \xrightarrow{D} \Gamma$ (where Γ is from Theorem

3.3.1), $\text{PART III} \xrightarrow{P} 0$ and $\text{PART IV} \xrightarrow{P} 0$. First we consider the first part

$$\text{PART I} = \sqrt{k}\frac{X_{(n-k,n)} - U\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)} \cdot \frac{1}{\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\, ds}.$$

We have that

$$\begin{aligned}\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\, ds &= \left[\frac{s^\gamma}{\gamma}\log s\right]_1^{\frac{k}{np}} - \frac{1}{\gamma}\int_1^{\frac{k}{np}} s^\gamma \frac{1}{s}\, ds \\ &= \frac{\left(\frac{k}{np}\right)^\gamma \log\frac{k}{np}}{\gamma} - \left[\frac{1}{\gamma^2}s^\gamma\right]_1^{\frac{k}{np}} \\ &= \frac{1}{\gamma}\left[\left(\frac{k}{np}\right)^\gamma \log\frac{k}{np} - \frac{1}{\gamma}\left(\frac{k}{np}\right)^\gamma + \frac{1}{\gamma}\right],\end{aligned}$$

and, from Theorem 3.3.1, we obtain

$$\sqrt{k}\frac{X_{(n-k,n)} - U\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)} \xrightarrow{D} B.$$

It remains to prove that

$$\frac{1}{\frac{1}{\gamma}\left[\left(\frac{k}{np}\right)^\gamma \log\frac{k}{np} - \frac{1}{\gamma}\left(\frac{k}{np}\right)^\gamma + \frac{1}{\gamma}\right]} \to 0.$$

Since, from initial conditions, $np = o(k)$ as $n \to \infty$, then $\left(\frac{k}{np}\right)^\gamma \to \infty$ (with $\gamma > 0$). Thus, $\log\left(\frac{k}{np}\right) \to \infty$ and

$$\frac{1}{\frac{1}{\gamma}\left[\left(\frac{k}{np}\right)^\gamma \log\frac{k}{np} - \frac{1}{\gamma}\left(\frac{k}{np}\right)^\gamma + \frac{1}{\gamma}\right]} = \frac{\gamma\left(\frac{k}{np}\right)^{-\gamma}}{\log\frac{k}{np} - \frac{1}{\gamma} + \frac{1}{\gamma}\left(\frac{k}{np}\right)^{-\gamma}} \to 0.$$

Hence,

$$\text{PART I} = \sqrt{k}\frac{X_{(n-k,n)} - U\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)} \cdot \frac{1}{\frac{1}{\gamma}\left[\left(\frac{k}{np}\right)^\gamma \log\frac{k}{np} - \frac{1}{\gamma}\left(\frac{k}{np}\right)^\gamma + \frac{1}{\gamma}\right]} \to 0.$$

Consider now the second part

$$\text{PART II} = \frac{X_{(n-k,n)}M_n^{(1)}}{a\left(\frac{n}{k}\right)}\left[\frac{\sqrt{k}}{\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\, ds} \cdot \left(\frac{\left(\frac{k}{np}\right)^{\widehat{GT}} - 1}{\widehat{GT}} - \frac{\left(\frac{k}{np}\right)^\gamma - 1}{\gamma}\right)\right].$$

Since from Theorem 3.3.1

$$\sqrt{k}\left(\frac{X_{(n-k,n)}M_n^{(1)}}{a\left(\frac{n}{k}\right)}-1\right)\xrightarrow{D}\Lambda,$$

then,

$$\frac{X_{(n-k,n)}M_n^{(1)}}{a\left(\frac{n}{k}\right)}-1\xrightarrow{P}0,$$

ie,

$$\frac{X_{(n-k,n)}M_n^{(1)}}{a\left(\frac{n}{k}\right)}=1+o_p\left(1\right).$$

We also know that

$$\int_1^{\frac{k}{np}}s^{\gamma-1}\,ds=\left[\frac{s^\gamma}{\gamma}\right]_1^{\frac{k}{np}}=\frac{\left(\frac{k}{np}\right)^\gamma}{\gamma}-\frac{1}{\gamma}=\frac{\left(\frac{k}{np}\right)^\gamma-1}{\gamma}.$$

Then,

$$\begin{aligned}\frac{\left(\frac{k}{np}\right)^{\widehat{GT}}-1}{\widehat{GT}}-\frac{\left(\frac{k}{np}\right)^\gamma-1}{\gamma}&=\int_1^{\frac{k}{np}}s^{\widehat{GT}-1}\,ds-\int_1^{\frac{k}{np}}s^{\gamma-1}\,ds\\&=\int_1^{\frac{k}{np}}\left(s^{\widehat{GT}-1}-s^{\gamma-1}\right)\,ds\\&=\int_1^{\frac{k}{np}}s^{\gamma-1}\left(s^{\widehat{GT}-\gamma}-1\right)\,ds.\end{aligned}$$

Thus, we may write

$$\frac{\sqrt{k}}{\int_1^{\frac{k}{np}}s^{\gamma-1}\log s\,ds}\cdot\left(\frac{\left(\frac{k}{np}\right)^{\widehat{GT}}-1}{\widehat{GT}}-\frac{\left(\frac{k}{np}\right)^\gamma-1}{\gamma}\right)=\frac{\sqrt{k}\int_1^{\frac{k}{np}}s^{\gamma-1}\left(s^{\widehat{GT}-\gamma}-1\right)\,ds}{\int_1^{\frac{k}{np}}s^{\gamma-1}\log s\,ds}$$

$$=\frac{\sqrt{k}}{\int_1^{\frac{k}{np}}s^{\gamma-1}\log s\,ds}\int_1^{\frac{k}{np}}s^{\gamma-1}\frac{e^{(\widehat{GT}-\gamma)\log s}-1}{(\widehat{GT}-\gamma)\log s}(\widehat{GT}-\gamma)\log s\,ds$$

$$=\frac{\sqrt{k}(\widehat{GT}-\gamma)}{\int_1^{\frac{k}{np}}s^{\gamma-1}\log s\,ds}\int_1^{\frac{k}{np}}s^{\gamma-1}\log s\frac{e^{(\widehat{GT}-\gamma)\log s}-1}{(\widehat{GT}-\gamma)\log s}\,ds.$$

From Theorem 3.3.1,

$$\sqrt{k}(\widehat{GT}-\gamma)\xrightarrow{D}\Gamma.$$

Then, in order to obtain the pretended result, we may write

$$\begin{aligned}
&\frac{\sqrt{k}(\widehat{GT}-\gamma)}{\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\,ds}\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\frac{e^{(\widehat{GT}-\gamma)\log s}-1}{(\widehat{GT}-\gamma)\log s}\,ds-\sqrt{k}(\widehat{GT}-\gamma)\\
&=\frac{\sqrt{k}(\widehat{GT}-\gamma)}{\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\,ds}\left[\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\frac{e^{(\widehat{GT}-\gamma)\log s}-1}{(\widehat{GT}-\gamma)\log s}\,ds-\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\,ds\right]\\
&=\frac{\sqrt{k}(\widehat{GT}-\gamma)}{\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\,ds}\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\left(\frac{e^{(\widehat{GT}-\gamma)\log s}-1}{(\widehat{GT}-\gamma)\log s}-1\right)\,ds.
\end{aligned}$$

Moreover, by initial conditions, $\log(np)=o\left(\sqrt{k}\right)$, then we have that, for any $1\le s\le\frac{k}{np}$,

$$\left|(\widehat{GT}-\gamma)\log s\right|\le\left|\sqrt{k}(\widehat{GT}-\gamma)\right|\frac{\log\left(\frac{k}{np}\right)}{\sqrt{k}}\xrightarrow{P}0.$$

From the above result and taking into account the fact that, by Taylor expansion, $\left|\frac{e^x-1}{x}-1\right|\le|x|$ as $x\to0$, we obtain that

$$\begin{aligned}
&\left|\frac{\sqrt{k}(\widehat{GT}-\gamma)}{\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\,ds}\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\left(\frac{e^{(\widehat{GT}-\gamma)\log s}-1}{(\widehat{GT}-\gamma)\log s}-1\right)\,ds\right|\le\\
&\le\frac{\sqrt{k}\left|\widehat{GT}-\gamma\right|}{\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\,ds}\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\left|(\widehat{GT}-\gamma)\log s\right|\,ds\\
&\le\frac{\sqrt{k}\left|\widehat{GT}-\gamma\right|}{\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\,ds}\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\left|(\widehat{GT}-\gamma)\log\frac{k}{np}\right|\,ds\\
&=\left[\sqrt{k}\left(\widehat{GT}-\gamma\right)\right]^2\frac{\log\frac{k}{np}}{\sqrt{k}}=O_p(1)\frac{\log\frac{k}{np}}{\sqrt{k}}\xrightarrow{P}0.
\end{aligned}$$

Consequently we have,

$$\text{PART II}=\frac{X_{(n-k,n)}M_n^{(1)}}{a\left(\frac{n}{k}\right)}\left[\frac{\sqrt{k}}{\int_1^{\frac{k}{np}} s^{\gamma-1}\log s\,ds}\cdot\left(\frac{\left(\frac{k}{np}\right)^{\widehat{GT}}-1}{\widehat{GT}}-\frac{\left(\frac{k}{np}\right)^{\gamma}-1}{\gamma}\right)\right]\xrightarrow{D}\Gamma.$$

Next we write the third part as

$$\begin{aligned}\text{PART III} &= \sqrt{k}\left(\frac{X_{(n-k,n)}M_n^{(1)}}{a\left(\frac{n}{k}\right)}-1\right)\frac{\left(\frac{k}{np}\right)^{\gamma}-1}{\gamma\int_1^{\frac{k}{np}}s^{\gamma-1}\log s\;ds}= \\ &= \sqrt{k}\left(\frac{X_{(n-k,n)}M_n^{(1)}}{a\left(\frac{n}{k}\right)}-1\right)\frac{\left(\frac{k}{np}\right)^{\gamma}-1}{\gamma\frac{1}{\gamma}\left[\left(\frac{k}{np}\right)^{\gamma}\log\frac{k}{np}-\frac{1}{\gamma}\left(\frac{k}{np}\right)^{\gamma}+\frac{1}{\gamma}\right]} \\ &= \sqrt{k}\left(\frac{X_{(n-k,n)}M_n^{(1)}}{a\left(\frac{n}{k}\right)}-1\right)\frac{1-\left(\frac{k}{np}\right)^{-\gamma}}{\log\frac{k}{np}-\frac{1}{\gamma}+\frac{1}{\gamma}\left(\frac{k}{np}\right)^{-\gamma}} \\ &\xrightarrow{P} 0,\end{aligned}$$

since $\sqrt{k}\left(\frac{X_{(n-k,n)}M_n^{(1)}}{a\left(\frac{n}{k}\right)}-1\right)\xrightarrow{D}\Lambda$ by Theorem 3.3.1.

Finally we focus on the fourth part. Since $\gamma>0$, we have $a(n/k)=\gamma U(n/k)$, and then

$$\begin{aligned}\text{PART IV} &= -\frac{\sqrt{k}}{\int_1^{\frac{k}{np}}s^{\gamma-1}\log s\;ds}\left[\frac{U\left(\frac{1}{p}\right)-U\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)}-\frac{\left(\frac{k}{np}\right)^{\gamma}-1}{\gamma}\right] \\ &= -\frac{\sqrt{k}}{\int_1^{\frac{k}{np}}s^{\gamma-1}\log s\;ds}\frac{1}{\gamma}\left[\frac{U\left(\frac{1}{p}\right)}{U\left(\frac{n}{k}\right)}-1-\left(\frac{k}{np}\right)^{\gamma}+1\right] \\ &= -\frac{\sqrt{k}A\left(\frac{n}{k}\right)}{\gamma\int_1^{\frac{k}{np}}s^{\gamma-1}\log s\;ds}\left[\frac{\frac{U\left(\frac{n}{k}\frac{k}{np}\right)}{U\left(\frac{n}{k}\right)}-\left(\frac{k}{np}\right)^{\gamma}}{A\left(\frac{n}{k}\right)}\right].\end{aligned}$$

Recall that by the condition (3.10) we have that

$$\frac{\frac{U(tx)}{U(t)}-x^{\gamma}}{A(t)}=x^{\gamma}\frac{x^{\rho}-1}{\rho}\left(1+o(1)\right).$$

Thus,

$$\begin{aligned}\text{PART IV} &= -\frac{\sqrt{k}A\left(\frac{n}{k}\right)}{\gamma\frac{1}{\gamma}\left[\left(\frac{k}{np}\right)^{\gamma}\log\left(\frac{k}{np}\right)-\frac{1}{\gamma}\left(\frac{k}{np}\right)^{\gamma}+\frac{1}{\gamma}\right]}\left(\frac{k}{np}\right)^{\gamma}\frac{\left(\frac{k}{np}\right)^{\rho}-1}{\rho}\left(1+o(1)\right) \\ &= -\frac{\sqrt{k}A\left(\frac{n}{k}\right)\left(\frac{k}{np}\right)^{\gamma}\left[\left(\frac{k}{np}\right)^{\rho}-1\right]\left(1+o(1)\right)}{\rho\left(\frac{k}{np}\right)^{\gamma}\left[\log\left(\frac{k}{np}\right)-\frac{1}{\gamma}+\frac{1}{\gamma}\left(\frac{np}{k}\right)^{\gamma}\right]} \\ &\to 0,\end{aligned}$$

since $\rho < 0$ and $\sqrt{k}A(n/k) \to \lambda$.

□

Chapter 4

Bias corrected geometric-type estimation for tail parameters

In order to obtain information about the upper tail of F, most of the estimators are constructed as functions of the upper k o.s. of a sample of size n (see e.g. Pickands (1975) and Dekkers et al. (1989)). When the number of upper o.s. used in the estimation of γ increases, the bias in the estimation becomes larger. This considerable bias that appears in several estimators reveals a difficult problem to go beyond the applications and there are several papers trying to deal with. Once this is such an important research theme, the bias reduction has become popular and received considerable attention in extreme value statistics. Some estimators were built in order to deal with the bias term in an appropriate way (see e.g. Peng (1998), Beirlant et al. (1999), Feuerverger and Hall (1999), Gomes et al. (2000), Gomes and Pestana (2007) and Beirlant et al. (2008)). One of the procedures commonly used to approach with this problem was formulated under second order properties of the d.f. and gave rise to the second order reduced-bias estimators.

In this chapter we improve the geometric-type estimator in the sense of reducing its bias. For this we propose two asymptotic equivalent bias corrected estimators for both tail index and high quantiles, and study the corresponding asymptotic behaviour.

It is convenient to assume that the underlying models belong to Hall's class (Hall (1982)), given by

$$U(t) = Ct^{\gamma}\left(1 + \frac{A(t)}{\rho}(1 + o(1))\right), \quad \text{as } t \to \infty,$$

where

$$A(t) = \gamma\beta t^{\rho}, \tag{4.1}$$

with $\gamma > 0$, and $C > 0$, $\rho < 0$ and $\beta \neq 0$ are, respectively, the shape and scale parameters. This is a very important family with several applications.

4.1 Bias corrected geometric-type estimators

In order to achieve the improvement of the geometric-type estimator behaviour presented in (3.6), and following some suggestions in the literature (see e.g. Caeiro et al. (2005)), we derive corrected geometric-type estimators by removing its bias dominant component.

For this we use the asymptotic representation of the geometric-type estimator presented in Theorem 3.2.4,

$$\widehat{GT}(k) \overset{D}{=} \gamma + \frac{\gamma}{2\sqrt{k}}Q_n - \frac{\gamma}{\sqrt{k}}P_n + \frac{A\left(\frac{n}{k}\right)}{(1-\rho)^2} + o_p\left(A\left(\frac{n}{k}\right)\right) + O_p\left(\frac{\log^2 k}{k}\right),$$

where the bias dominant component can be written as

$$\frac{A\left(\frac{n}{k}\right)}{(1-\rho)^2} = \frac{\gamma\beta\left(\frac{n}{k}\right)^{\rho}}{(1-\rho)^2}.$$

Thus, removing the bias dominant component directly, we obtain a bias corrected estimator of $\widehat{GT}$ given by

$$\overline{\widehat{GT}}(k) = \widehat{GT}(k)\left(1 - \frac{\beta\left(\frac{n}{k}\right)^{\rho}}{(1-\rho)^2}\right). \tag{4.2}$$

Considering now the exponential expansion $e^{-x} = 1 - x + o(x)$ as $x \to 0$, we may get the asymptotically equivalent bias corrected estimator

$$\overline{\overline{\widehat{GT}}}(k) = \widehat{GT}(k)\, exp\left\{-\frac{\beta}{(1-\rho)^2}\left(\frac{n}{k}\right)^{\rho}\right\}. \tag{4.3}$$

We can easily note that the bias dominant component is dependent of the shape ρ and scale β second order parameters. Thus, another challenge of utmost importance to consider is the proper and adequate estimation of the second order parameters, ρ and β, in order to remove the bias dominant component and obtain bias corrected estimators.

We remark that the geometric-type estimator has a lower bias dominant component than the Hill estimator when evaluated at the same threshold, ie, for the same k.

Estimation of the second order parameters

Here, we consider the class of estimators of the parameter ρ (depending on τ) proposed by Fraga Alves et al. (2003)

$$\widehat{\rho}_n^{(\tau)}(k) = -\left| \frac{3\left(T_n^{(\tau)}(k) - 1\right)}{T_n^{(\tau)}(k) - 3} \right|, \tag{4.4}$$

where

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

with M_n^j as in (3.8), and the β estimator obtained in Gomes and Martins (2002)

$$\widehat{\beta}_{\widehat{\rho}}(k) = \left(\frac{k}{n}\right)^{\widehat{\rho}} \frac{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\right)\frac{1}{k}\sum_{i=1}^{k}U_i - \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}U_i}{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\right)\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}U_i - \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-2\widehat{\rho}}U_i}, \tag{4.5}$$

where

$$U_i = i\left(\log\frac{X_{(n-i+1,n)}}{X_{(n-i,n)}}\right),$$

with $1 \leq i \leq k < n$.

We remark that the class of estimators of ρ presented above, and consequently also the β estimators, is dependent on a tuning parameter $\tau \geq 0$. In the literature it has

been suggested the use of the tuning parameter $\tau = 0$ when $\rho \in [-1, 0)$ and $\tau = 1$ when $\rho \in (-\infty, -1)$. This parameter must be chosen appropriately in order to provide a higher stability for the estimator of ρ and as such, a graphical study supporting this choice must always be seen as a relevant tool.

Choice of the k_h level to be used in the second order parameters estimation

It is known that the external estimation of ρ and β at a larger k value than the one used for γ-estimation has clear advantages, allowing the bias reduction without increasing the asymptotic variance (see e.g. Caeiro et al. (2005)).

Through some simulation studies presented in the next chapter we can notice that the estimator of ρ only stabilises at high levels of k, which justifies the suggestion given in some works that ρ must be estimated at a high level k_h (see e.g. Caeiro and Gomes (2008) and Gomes et al. (2004)). Moreover, the number k_h of the top observations to be considered for the estimation of ρ and β should be such as to ensure that $\widehat{\rho} - \rho = o_p(1/\log n)$.

In the lines of other studies, and among some suggestions (see e.g. Gomes et al. (2007)), the level that seemed to be the most appropriate to consider in illustrations is

$$k_h = \left\lfloor n^{1-\epsilon} \right\rfloor, \text{ for some } \epsilon > 0 \text{ small,} \tag{4.6}$$

where $\lfloor x \rfloor$ denotes the integer part of x.

4.2 Asymptotic properties of the geometric-type bias corrected estimators

We begin by assuming that only the tail index parameter γ is unknown and that $\widehat{GT}^*$ is one of the estimators $\widehat{\widehat{GT}}$ or $\widehat{\widehat{\widehat{GT}}}$.

Theorem 4.2.1. *Assume* (3.10) *holds. For sequences* k *such that* (3.3) *holds, and* $A(t)$

as in (4.1)*, we have the following asymptotic distributional representation*

$$\widehat{GT}^{*}(k) \overset{D}{=} \gamma + \frac{\gamma}{2\sqrt{k}} Q_n - \frac{\gamma}{\sqrt{k}} P_n + o_p\left(A\left(\frac{n}{k}\right)\right) + O_p\left(\frac{\log^2 k}{k}\right),$$

where $P_n = \sqrt{k}\left(\sum_{i=1}^{k} Z_i/k - 1\right)$ *and* $Q_n = \sqrt{k}\left(\sum_{i=1}^{k} Z_i^2/k - 2\right)$*,* (P_n, Q_n) *is asymptotically normal with mean equal to* $\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$ *and covariance matrix* $\left(\begin{smallmatrix} 1 & 4 \\ 4 & 20 \end{smallmatrix}\right)$*, and* $\{Z_i\}$ *denote i.i.d. standard exponential r.v..*

Proof of Theorem 4.2.1. Recall that $\widehat{GT}(k) \xrightarrow{P} \gamma$ as $n \to \infty$. If all parameters are known, except the tail index γ, we get

$$\begin{aligned}
\widehat{\widehat{GT}}(k) &= \widehat{GT}(k)\left(1 - \frac{\beta\left(\frac{n}{k}\right)^{\rho}}{(1-\rho)^2}\right) \\
&= \widehat{GT}(k) - \frac{A\left(\frac{n}{k}\right)}{(1-\rho)^2}\left(1 + o_p(1)\right) \\
&= \gamma + \frac{\gamma}{2\sqrt{k}} V_n - \frac{\gamma}{\sqrt{k}} T_n + o_p\left(A\left(\frac{n}{k}\right)\right) + O_p\left(\frac{\log^2 k}{k}\right).
\end{aligned}$$

With an easy calculation, we also have

$$\begin{aligned}
\widehat{\widehat{\widehat{GT}}}(k) &= \widehat{GT}(k)\, exp\left(-\frac{\beta\left(\frac{n}{k}\right)^{\rho}}{(1-\rho)^2}\right) \\
&= \widehat{GT}(k)\left[1 - \frac{A\left(\frac{n}{k}\right)}{\gamma(1-\rho)^2} + o_p\left(\frac{A\left(\frac{n}{k}\right)}{\gamma(1-\rho)^2}\right)\right] \\
&= \gamma + \frac{\gamma}{2\sqrt{k}} V_n - \frac{\gamma}{\sqrt{k}} T_n + o_p\left(A\left(\frac{n}{k}\right)\right) + O_p\left(\frac{\log^2 k}{k}\right).
\end{aligned}$$

□

Corollary 4.2.2. *Assume the conditions of Theorem 4.2.1 hold. If we choose* k *such that* $\sqrt{k}A(n/k) \to \lambda$ *finite, then*

$$\sqrt{k}\left(\widehat{GT}^{*}(k) - \gamma\right)$$

is asymptotically normal distributed as $n \to \infty$*, with variance* $2\gamma^2$ *and a null mean value.*

Proof of Corollary 4.2.2. From the proof of Theorem 4.2.1 we have

$$\sqrt{k}\left(\widehat{GT}^{*}(k) - \gamma\right) = \frac{\gamma}{2} V_n - \gamma T_n + \sqrt{k} o_p\left(A\left(\frac{n}{k}\right)\right) + O_p\left(\frac{\log^2 k}{\sqrt{k}}\right).$$

Since $\sqrt{k}A(n/k) \to \lambda$ as $n \to \infty$,

$$\sqrt{k}\left(\widehat{GT}^{*}(k) - \gamma\right) = \frac{\gamma}{2}V_n - \gamma T_n + o_p(1).$$

It remains to compute the values of the asymptotic variance and mean:

$$E\left[\sqrt{k}\left(\widehat{GT}^{*}(k) - \gamma\right)\right] = \frac{\gamma}{2}E(V_n) - \gamma E(T_n) \xrightarrow[n\to\infty]{} 0,$$
$$V\left[\sqrt{k}\left(\widehat{GT}^{*}(k) - \gamma\right)\right] = \frac{\gamma^2}{4}V(V_n) + \gamma^2 V(T_n) - 2Cov\left(\frac{\gamma}{2}V_n, \gamma T_n\right) \xrightarrow[n\to\infty]{} 2\gamma^2.$$

□

Assuming now that $\widehat{GT}^{**}$ denotes the version of $\widehat{GT}^{*}$ where the parameters ρ and β are estimated externally, we have the following result

Theorem 4.2.3. *Under the conditions of Theorem 4.2.1 and assuming consistent estimators for ρ and β computed at a level that implies $\widehat{\rho} - \rho = o_p(1/\log n)$, we have the following asymptotic distributional representation*

$$\widehat{GT}^{**}(k) \overset{D}{=} \gamma + \frac{\gamma}{2\sqrt{k}}Q_n - \frac{\gamma}{\sqrt{k}}P_n + o_p\left(A\left(\frac{n}{k}\right)\right) + O_p\left(\frac{\log^2 k}{k}\right),$$

where $P_n = \sqrt{k}\left(\sum_{i=1}^{k} Z_i/k - 1\right)$ and $Q_n = \sqrt{k}\left(\sum_{i=1}^{k} Z_i^2/k - 2\right)$, (P_n, Q_n) is asymptotically normal with mean equal to $\binom{0}{0}$ and covariance matrix $\left(\begin{smallmatrix} 1 & 4 \\ 4 & 20 \end{smallmatrix}\right)$, and $\{Z_i\}$ denote i.i.d. standard exponential r.v..

Proof of Theorem 4.2.3. If ρ and β are estimated consistently, we can use the Taylor's expansion for bivariate functions and get

$$\begin{aligned}\frac{\widehat{\beta}}{(1-\widehat{\rho})^2}\left(\frac{n}{k}\right)^{\widehat{\rho}} &= \frac{\beta}{(1-\rho)^2}\left(\frac{n}{k}\right)^{\rho} + \left(\widehat{\beta} - \beta\right)\frac{1}{(1-\rho)^2}\left(\frac{n}{k}\right)^{\rho}(1 + o_p(1)) \\ &\quad + \frac{\beta}{(1-\rho)^2}(\widehat{\rho} - \rho)\left(\frac{n}{k}\right)^{\rho}\left(\frac{2}{1-\rho} + \log\left(\frac{n}{k}\right)\right)(1 + o_p(1)) \\ &= \frac{A(n/k)}{\gamma(1-\rho)^2}\left(\frac{\widehat{\beta}}{\beta} + \frac{2(\widehat{\rho} - \rho)}{1-\rho} + (\widehat{\rho} - \rho)\log\left(\frac{n}{k}\right)\right)(1 + o_p(1)),\end{aligned}$$

where $\widehat{\beta}$ and $\widehat{\rho}$ are the estimators of β and ρ, respectively.

Therefore we have

$$\widehat{GT}(k)\left(1-\frac{\widehat{\beta}\left(\frac{n}{k}\right)^{\widehat{\rho}}}{(1-\widehat{\rho})^2}\right)=\widehat{GT}(k)-\frac{A\left(\frac{n}{k}\right)}{(1-\rho)^2}+o_p\left(A\left(\frac{n}{k}\right)\right)$$
$$=\gamma+\frac{\gamma}{\sqrt{k}}\left(\frac{V_n}{2}-T_n\right)+o_p\left(A\left(\frac{n}{k}\right)\right)+O_p\left(\frac{log^2k}{k}\right)$$

and

$$\widehat{GT}(k)\,exp\left(-\frac{\widehat{\beta}\left(\frac{n}{k}\right)^{\widehat{\rho}}}{(1-\widehat{\rho})^2}\right)=\gamma+\frac{\gamma}{\sqrt{k}}\left(\frac{V_n}{2}-T_n\right)+o_p\left(A\left(\frac{n}{k}\right)\right)+O_p\left(\frac{\log^2 k}{k}\right),$$

since $\widehat{\rho}$ and $\widehat{\beta}$ are consistent estimators of ρ and β computed at a level such that $\widehat{\rho}-\rho=o_p\left(1/\log n\right)$. The result follows. □

Corollary 4.2.4. *Assume the conditions of Theorem 4.2.3 hold. If we choose k such that $\sqrt{k}A(n/k)\to\lambda$ finite, then*

$$\sqrt{k}\left(\widehat{GT}^{**}(k)-\gamma\right)$$

is asymptotically normal distributed as $n\to\infty$ with variance $2\gamma^2$ and a null mean value.

Proof of Corollary 4.2.4. The result follows using the same approach as in the proof of Corollary 4.2.2. □

4.3 High order quantiles estimation using geometric-type bias corrected estimators

Here, in order to improve the performance of the suggested geometric-type high quantiles estimator, we also consider the form (2.2) based on the geometric-type bias corrected estimators, and its asymptotic normality is established.

More concretely, we apply the bias corrected tail index estimators introduced in this chapter, (4.2) and (4.3), on the POT high quantiles estimator in (2.2), obtaining the following two geometric-type bias corrected high quantiles estimators

$$\widehat{\chi}_{1-p}^{\widehat{\widehat{GT}}}=\frac{\left(\frac{k}{np}\right)^{\widehat{\widehat{GT}}(k)}-1}{\widehat{\widehat{GT}}\,(k)}\cdot X_{(n-k,n)}M_n^{(1)}+X_{(n-k,n)}$$

and

$$\widehat{\chi}_{1-p}^{\widehat{\widehat{\widehat{GT}}}}=\frac{\left(\frac{k}{np}\right)^{\widehat{\widehat{\widehat{GT}}}(k)}-1}{\widehat{\widehat{\widehat{GT}}}\,(k)}\cdot X_{(n-k,n)}M_n^{(1)}+X_{(n-k,n)}.$$

The following result will be used to deduce the asymptotic normality of the quantiles estimator using geometric-type bias corrected estimators.

Theorem 4.3.1. *Assume the conditions of Theorem 3.3.1 hold. For sequences k such that* (3.3) *holds and $\sqrt{k}A(n/k)\to\lambda$ finite, then*

$$\sqrt{k}\left(\frac{X_{(n-k,n)}M_n^{(1)}}{a\left(\frac{n}{k}\right)}-1,\widehat{GT}^{*}\,(k)-\gamma,\frac{X_{(n-k,n)}-U\left(\frac{n}{k}\right)}{a\left(\frac{n}{k}\right)}\right)\xrightarrow[n\to\infty]{D}(\Lambda,\Gamma^{*},B)\,,$$

where (Λ,Γ^{*},B) are jointly normal r.v. with mean vector $(\lambda/(\gamma-\gamma\rho),0,0)^{\top}$ and covariance matrix

$$\begin{pmatrix}1+\gamma^2 & \gamma & \gamma\\ \gamma & 2\gamma^2 & 0\\ \gamma & 0 & 1\end{pmatrix}. \tag{4.7}$$

Proof of Theorem 4.3.1. The result follows combining the proof of Theorem 3.3.1 and Theorem 4.2.1. □

Now, the asymptotic normality of the quantiles estimator may be easily deduced from the previous result jointly with Theorem 3.3.2.

Theorem 4.3.2. *Under the conditions of Theorem 3.3.2,*

$$\sqrt{k}\frac{\widehat{\chi}_{1-p}^{\widehat{GT}^{*}}-\chi_{1-p}}{a\left(\frac{n}{k}\right)q_\gamma\left(d_n\right)}\xrightarrow[n\to\infty]{D}N\left(0,2\gamma^2\right).$$

where $d_n=k/(np)$, $q_\gamma(t)=\int_1^t s^{\gamma-1}\log s\,ds$ for $t>1$ and

$$\widehat{\chi}_{1-p}^{\widehat{GT}^*} = \frac{\left(\frac{k}{np}\right)^{\widehat{GT}^*(k)} - 1}{\widehat{GT}^*(k)} \cdot X_{(n-k,n)} M_n^{(1)} + X_{(n-k,n)},$$

for some $p \in (0, 1/n]$.

Proof of Theorem 4.3.2. The result follows using Theorem 4.3.1 and the same approach as in the proof of Theorem 3.3.2. □

The two results above still hold for the corresponding $\widehat{GT}^{**}$ based quantile estimators.

Chapter 5

Simulation study results and comparisons

In this chapter we present some simulations in order to examine the finite sample behaviour of the proposed tail index and high quantiles estimators. We have generated s=2000 independent replicates of sample size 1000 from the Generalised Pareto Distribution with d.f.

$$F(x) = 1 - (1 + \gamma x)^{-1/\gamma}, \quad x \geq 0, \quad \gamma = 1,$$

and from the Burr distribution with d.f.

$$F(x) = 1 - \left(1 + x^{-\rho/\gamma}\right)^{1/\rho}, \quad x \geq 0, \quad \gamma = 1 \text{ and } \rho = -2.$$

Remark that $\beta = 1$ for both families, and for GPD $\rho = -\gamma$.

The results are compared using mean values of the estimates and through relative root mean square error (RRMSE), with the expression

$$\widehat{RRMSE}\left(\widehat{\theta}\right) = \frac{\sqrt{\frac{1}{s}\sum_{i=1}^{s}\left(\widehat{\theta}_i - \theta\right)^2}}{\theta},$$

where θ is the value we want to estimate.

5.1 Tail index estimation

The main purpose of the simulations performed in this section is to provide a general insight into the distributional behaviour of the new geometric-type bias corrected tail index estimators proposed, (4.2) and (4.3). Once the evaluation of their behaviour encompasses the comparison with similar corrections of Hill estimator, we start by presenting in Figures 5.1 and 5.2 the behaviour of both original estimators for the chosen distributions.

To illustrate the behaviour of the corrected estimators we consider the suitable estimators of the parameter ρ proposed by Fraga Alves et al. (2003), in (4.4), and the β estimator obtained in Gomes and Martins (2002), in (4.5). Firstly we need to choose the tuning parameter τ, in which we will support the estimation of the second order parameters ρ and β. To achieve this, we draw in Figure 5.3 the behaviour of $\widehat{\rho}_\tau$ for the values of the tuning parameter $\tau \in \{0, 0.5, 1\}$ for both distributions and analyse the variations that it causes in their behaviour.

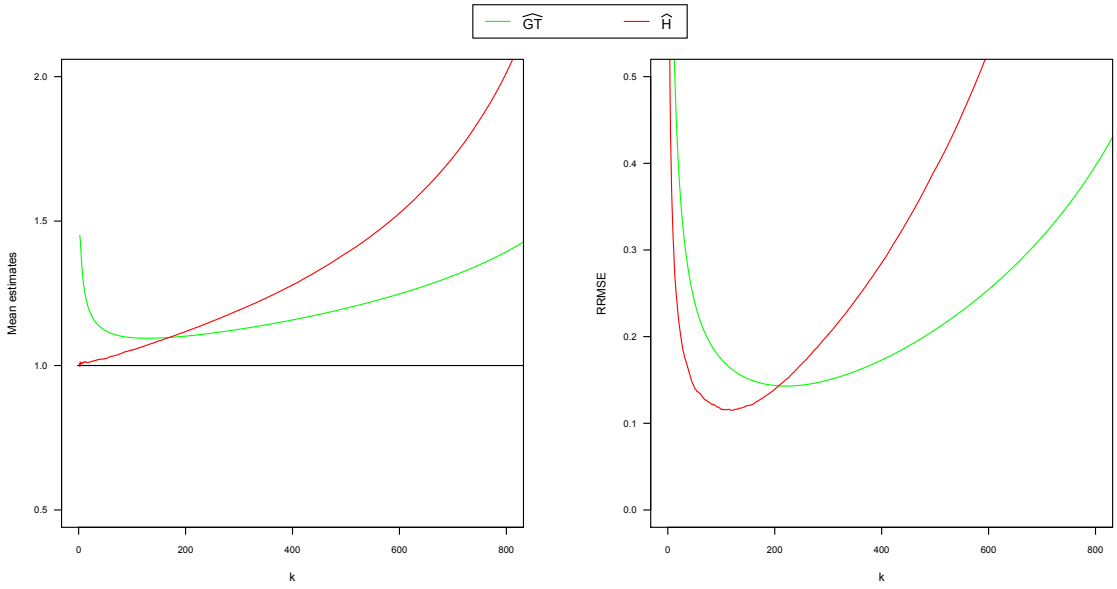


Figure 5.1: Mean estimates (left) and RRMSE (right) of $\widehat{GT}$ and $\widehat{H}$, for a sample size n=1000 (and 2000 replicates), as a function of k, from a GPD given by $F(x) = 1 - (1 + \gamma x)^{-1/\gamma}$, $x \geq 0$ with $\gamma = 1$.

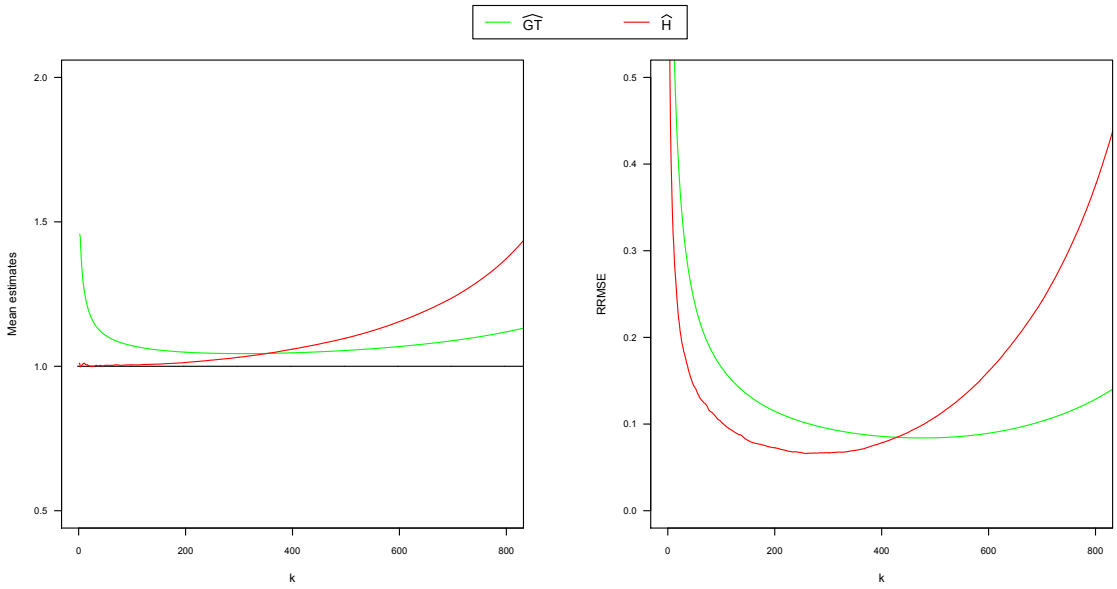


Figure 5.2: Mean estimates (left) and RRMSE (right) of $\widehat{GT}$ and $\widehat{H}$, for a sample size n=1000 (and 2000 replicates), as a function of k, from a Burr distribution given by $F(x) = 1 - \left(1 + x^{-\rho/\gamma}\right)^{1/\rho}$, $x \geq 0$, with $\gamma = 1$ and $\rho = -2$.

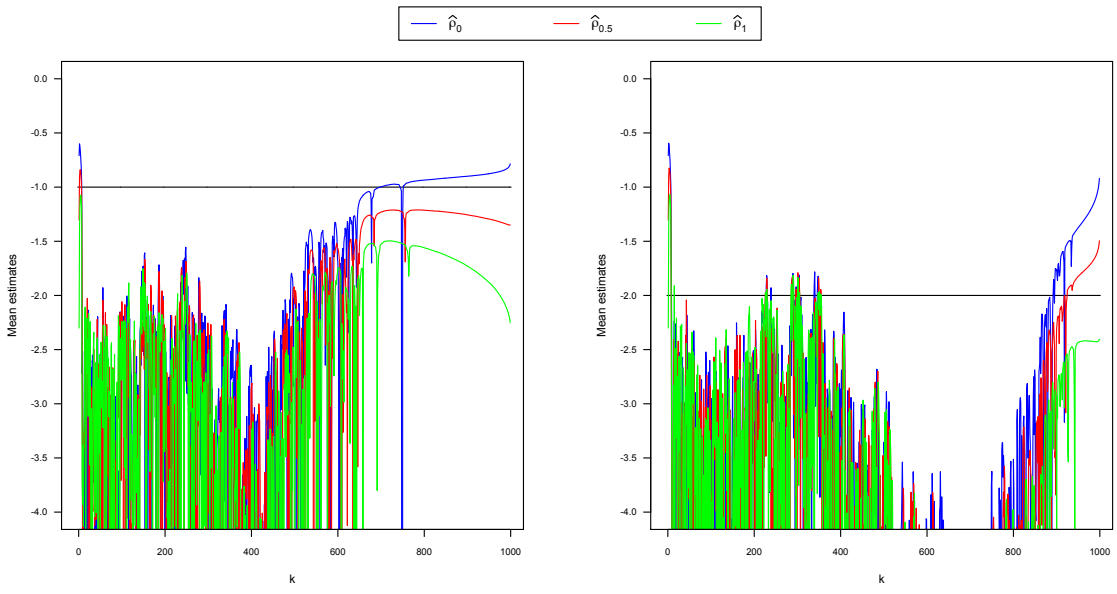


Figure 5.3: Mean estimates of $\widehat{\rho}_\tau$, $\tau = \{0, 0.5, 1\}$, for GPD (left) and Burr (right) distributions. GPD given by $F(x) = 1 - (1 + \gamma x)^{-1/\gamma}$, $x \geq 0$ ($\rho = -1$), and Burr distribution given by $F(x) = 1 - \left(1 + x^{-\rho/\gamma}\right)^{1/\rho}$, $x \geq 0$ and $\rho = -2$, both with $\gamma = 1$ ($\beta = 1$).

It is suggested in some works the use of $\tau = 0$ when $\rho \in [-1, 0)$ and $\tau = 1$ when $\rho \in (-\infty, -1)$ (see e.g. Fraga Alves et al. (2003)). This leads to the choice of $\tau = 0$ for the GPD $(\rho = -1)$ and $\tau = 1$ for Burr distribution $(\rho = -2)$. The Figure 5.3 confirms the prevalent choice of $\tau = 0$ for GPD but suggests that perhaps the choice of $\tau = 0.5$ instead of $\tau = 1$ seems to be more suitable for Burr distribution, leading to better estimates of β and ρ.

We also remark that the estimator of ρ presents a high variation in the majority of k values, stabilising only at very high levels of k, for which the estimates gets closer to the true value of the parameter. This fact reaffirm that estimation of ρ at a high level is favourable and highly recommended.

For exploring the results we consider in (4.6) $\epsilon = 0.005$ and $\epsilon = 0.001$, ie, we use the following k_h levels:

$$k_{h1} = \left\lfloor n^{0.995} \right\rfloor \qquad \text{and} \qquad k_{h2} = \left\lfloor n^{0.999} \right\rfloor . \tag{5.1}$$

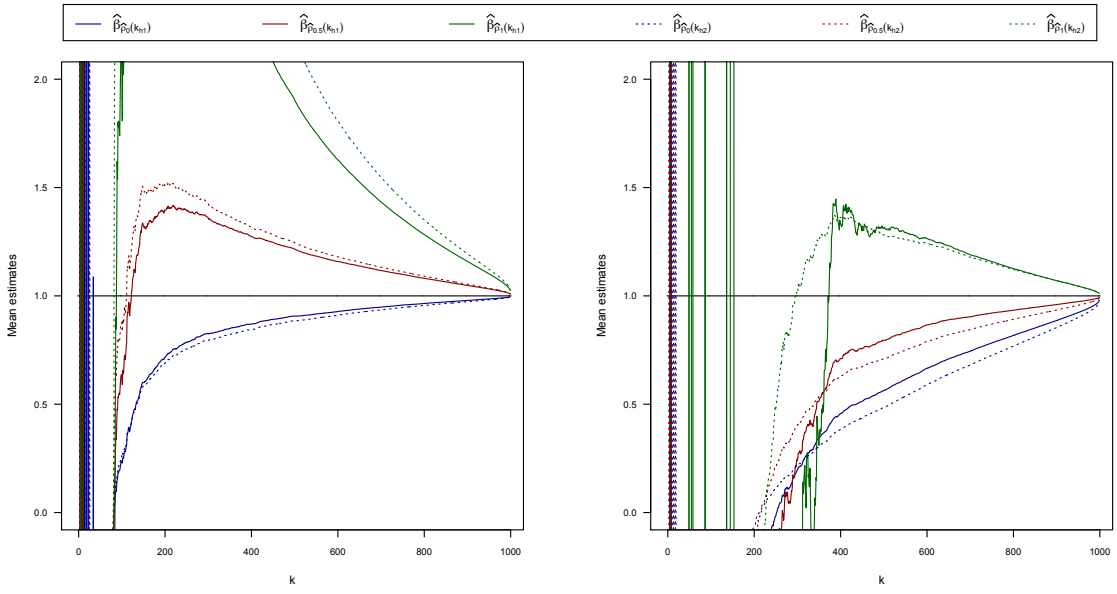


Figure 5.4: Mean estimates of $\widehat{\beta}_{\widehat{\rho}_\tau(k_{h1})}$ and $\widehat{\beta}_{\widehat{\rho}_\tau(k_{h2})}$, $\tau = \{0, 0.5, 1\}$, for GPD (left) and Burr (right) distributions. GPD given by $F(x) = 1 - (1 + \gamma x)^{-1/\gamma}$, $x \geq 0$ $(\rho = -1)$, and Burr distribution given by $F(x) = 1 - \left(1 + x^{-\rho/\gamma}\right)^{1/\rho}$, $x \geq 0$ and $\rho = -2$, both with $\gamma = 1$ $(\beta = 1)$.

To give an idea about the behaviour of $\widehat{\beta}$ according to the choice of τ and the level k_h, we present in Figure 5.4 the estimates of β computed with $\widehat{\rho}_\tau(k_{h1})$ and $\widehat{\rho}_\tau(k_{h2})$, $\tau \in \{0, 0.5, 1\}$, for both distributions. One aspect that stands out in this figure is that estimates of β are more favourable the higher the k-value used for its calculation.

Following what seems to be graphically more propitious, we chose to estimate ρ and β using $\tau = 0$ for GPD and $\tau = 0.5$ for Burr distribution, both computed at the same level k_{h1} or k_{h2}. The correct estimation of these parameters is crucial in order to get better estimates of the tail index using corrected estimators.

Now we have the necessary tools to estimate the tail index using the bias corrected tail index estimators. In this way, the illustrations that follow contain a graphical representation of the behaviour of the corrected estimators according to the choices on τ made for each distribution.

From the asymptotic normality we construct confidence intervals for the tail index, with $(1-\alpha)$-level, in the usual way:

$$I_{\widehat{GT}}(k,\alpha) = \left\{ \gamma : \frac{1}{\sqrt{2}\gamma}\, k^{1/2} |\widehat{GT} - \gamma| \leq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right\}.$$

The confidence bounds for the corresponding geometric-type bias corrected estimators are similar to the previous ones.

From the Figures 5.5 and 5.6, in which the geometric-type estimator is confronted with its new corrected versions, we observe that using both GPD and Burr distribution, the performance of the geometric-type estimator was improved by bias correction and the resulting geometric-type bias corrected estimators shows a very good behaviour.

We also note that the performance of the corrected estimators are slightly better when we calculate the second order parameters using the level k_{h2} instead of using the k_{h1} level. The corresponding 95% confidence bounds of the geometric-type estimator and of the corresponding bias corrected estimators are reported in Tables 5.1 and 5.2. We present three values of k for the illustration of the influence of the choice of k.

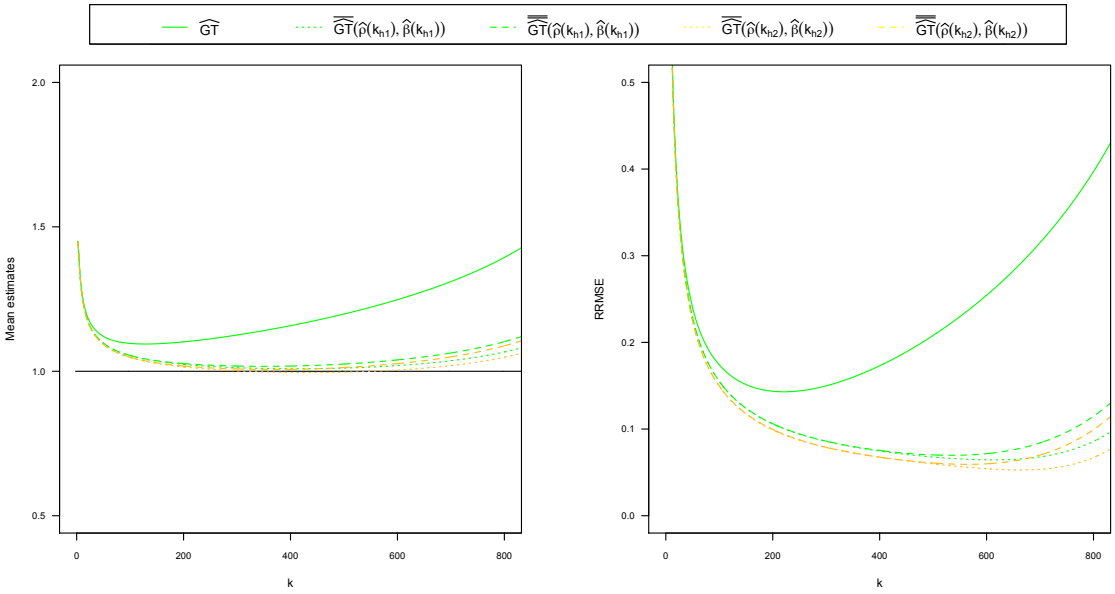


Figure 5.5: Mean estimates (left) and RRMSE (right) of $\widehat{GT}$, $\widehat{\widehat{GT}}$ and $\widehat{\widehat{\widehat{GT}}}$, with $\widehat{\rho}$ and $\widehat{\beta}$ computed at the levels $k_{h1} = \lfloor n^{0.995} \rfloor$ and $k_{h2} = \lfloor n^{0.999} \rfloor$, for a sample size n=1000 (and 2000 replicates), as a function of k, from a GPD given by $F(x) = 1 - (1+\gamma x)^{-1/\gamma}$, $x \geq 0$ with $\gamma = 1$, ($\rho = -1$, $\beta = 1$; $\tau = 0$).

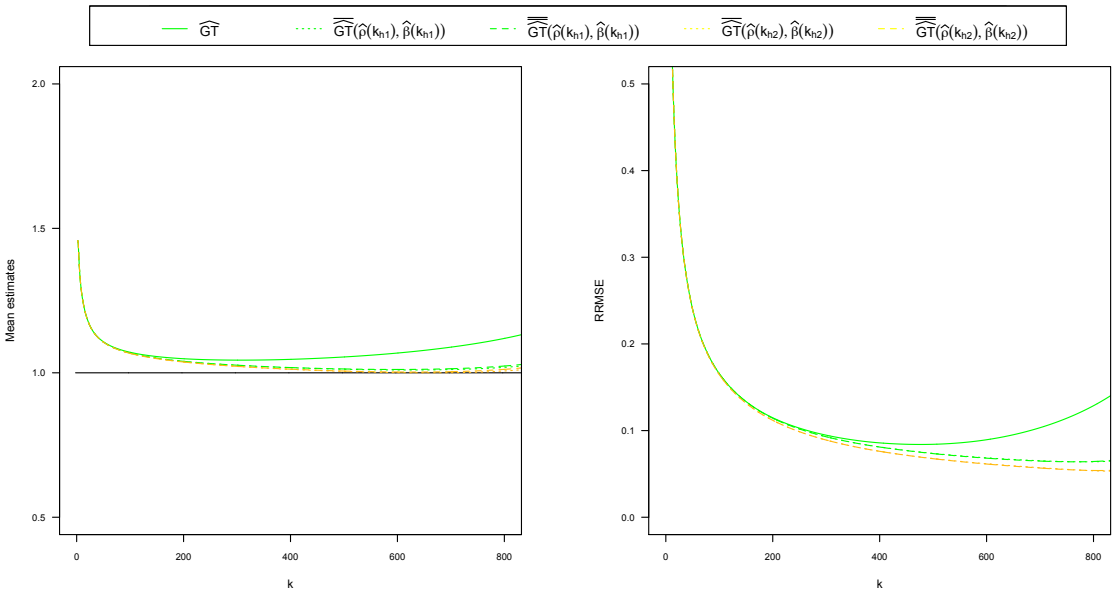


Figure 5.6: Mean estimates (left) and RRMSE (right) of $\widehat{GT}$, $\widehat{\widehat{GT}}$ and $\widehat{\widehat{\widehat{GT}}}$, with $\widehat{\rho}$ and $\widehat{\beta}$ computed at the levels $k_{h1} = \lfloor n^{0.995} \rfloor$ and $k_{h2} = \lfloor n^{0.999} \rfloor$, for a sample size n=1000 (and 2000 replicates), as a function of k, from a Burr distribution given by $F(x) = 1 - \left(1 + x^{-\rho/\gamma}\right)^{1/\rho}$, $x \geq 0$, with $\gamma = 1$, $\rho = -2$ ($\beta = 1$; $\tau = 0.5$).

Table 5.1: Confidence bounds ($\alpha = 0.05$) using the geometric-type estimator and the corresponding bias corrected estimators, with $\widehat{\rho}$ and $\widehat{\beta}$ computed at the levels $k_{h1} = \lfloor n^{0.995} \rfloor$ and $k_{h2} = \lfloor n^{0.999} \rfloor$, for a sample size n=1000 (and 2000 replicates), as a function of k. GPD $F(x) = 1 - (1+\gamma x)^{-1/\gamma}$, $x \geq 0$ with $\gamma = 1$ ($\rho = -1$, $\beta = 1$; $\tau = 0$).

k	$\widehat{GT}$	$\overline{\widehat{GT}}_{\widehat{\rho}(k_{h1}),\widehat{\beta}(k_{h1})}$	$\overline{\widehat{GT}}_{\widehat{\rho}(k_{h2}),\widehat{\beta}(k_{h2})}$	$\overline{\overline{\widehat{GT}}}_{\widehat{\rho}(k_{h1}),\widehat{\beta}(k_{h1})}$	$\overline{\overline{\widehat{GT}}}_{\widehat{\rho}(k_{h2}),\widehat{\beta}(k_{h2})}$
300	1.125 $\pm$ 0.180	1.013 $\pm$ 0.162	1.002 $\pm$ 0.160	1.018 $\pm$ 0.163	1.008 $\pm$ 0.161
500	1.198 $\pm$ 0.149	1.011 $\pm$ 0.125	0.997 $\pm$ 0.124	1.025 $\pm$ 0.127	1.013 $\pm$ 0.126
700	1.310 $\pm$ 0.137	1.037 $\pm$ 0.109	1.020 $\pm$ 0.107	1.063 $\pm$ 0.111	1.050 $\pm$ 0.110

Table 5.2: Confidence bounds ($\alpha = 0.05$) using the geometric-type estimator and the corresponding bias corrected estimators, with $\widehat{\rho}$ and $\widehat{\beta}$ computed at the levels $k_{h1} = \lfloor n^{0.995} \rfloor$ and $k_{h2} = \lfloor n^{0.999} \rfloor$, for a sample size n=1000 (and 2000 replicates), as a function of k. Burr distribution $F(x) = 1 - \left(1 + x^{-\rho/\gamma}\right)^{1/\rho}$, $x \geq 0$, with $\gamma = 1$ and $\rho = -2$ ($\beta = 1$; $\tau = 0.5$).

k	$\widehat{GT}$	$\overline{\widehat{GT}}_{\widehat{\rho}(k_{h1}),\widehat{\beta}(k_{h1})}$	$\overline{\widehat{GT}}_{\widehat{\rho}(k_{h2}),\widehat{\beta}(k_{h2})}$	$\overline{\overline{\widehat{GT}}}_{\widehat{\rho}(k_{h1}),\widehat{\beta}(k_{h1})}$	$\overline{\overline{\widehat{GT}}}_{\widehat{\rho}(k_{h2}),\widehat{\beta}(k_{h2})}$
300	1.044 $\pm$ 0.167	1.026 $\pm$ 0.164	1.022 $\pm$ 0.164	1.026 $\pm$ 0.164	1.022 $\pm$ 0.164
500	1.055 $\pm$ 0.131	1.012 $\pm$ 0.125	1.005 $\pm$ 0.125	1.013 $\pm$ 0.126	1.006 $\pm$ 0.125
700	1.089 $\pm$ 0.114	1.011 $\pm$ 0.106	1.000 $\pm$ 0.105	1.014 $\pm$ 0.106	1.004 $\pm$ 0.105

We note that when using corrected estimators, the amplitude of the asymptotic confidence intervals is smaller.

In order to have an idea of the good behaviour of the geometric-type bias corrected estimators, we compare them with the corresponding Hill bias corrected estimators (see e.g. Caeiro et al. (2005)), given by

$$\overline{\widehat{H}}(k) = \widehat{H}(k)\left(1 - \frac{\widehat{\beta}\left(\frac{n}{k}\right)^{\widehat{\rho}}}{1-\widehat{\rho}}\right)$$

and

$$\overline{\overline{\widehat{H}}}(k) = \widehat{H}(k)\, exp\left\{-\frac{\widehat{\beta}}{1-\widehat{\rho}}\left(\frac{n}{k}\right)^{\widehat{\rho}}\right\},$$

where $\widehat{\rho}$ and $\widehat{\beta}$ are the estimators of the shape and scale parameters, respectively.

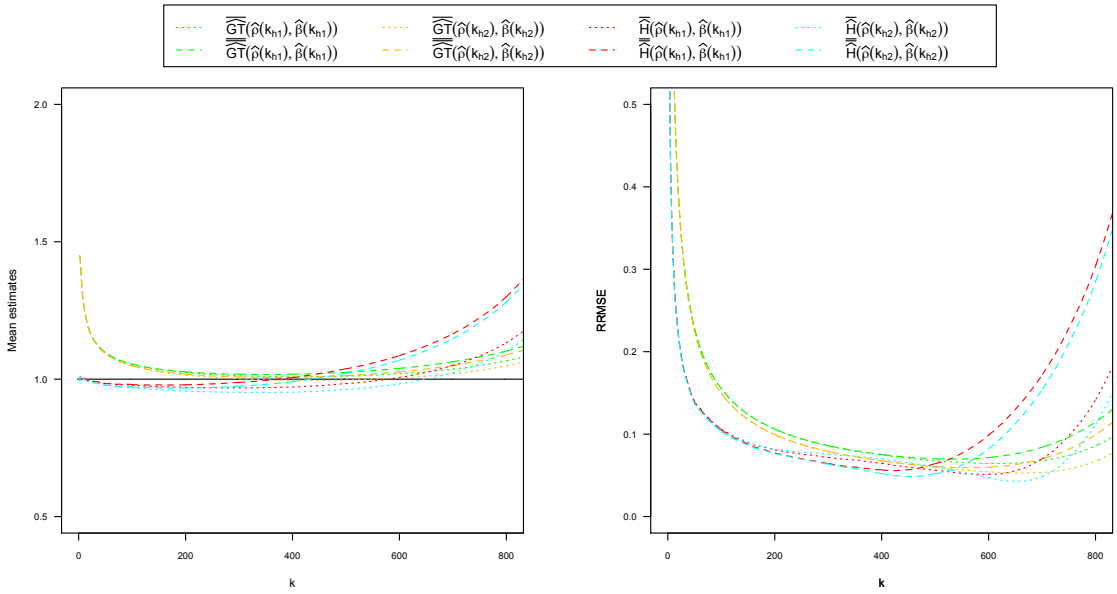


Figure 5.7: Mean estimates (left) and RRMSE (right) of $\overline{\widehat{GT}}$, $\overline{\overline{\widehat{GT}}}$, $\overline{\widehat{H}}$ and $\overline{\overline{\widehat{H}}}$, with $\widehat{\rho}$ and $\widehat{\beta}$ computed at the levels $k_{h1} = \lfloor n^{0.995} \rfloor$ and $k_{h2} = \lfloor n^{0.999} \rfloor$, for a sample size n=1000 (and 2000 replicates), as a function of k, from a GPD given by $F(x) = 1 - (1 + \gamma x)^{-1/\gamma}$, $x \geq 0$ with $\gamma = 1$ ($\rho = -1$, $\beta = 1$; $\tau = 0$).

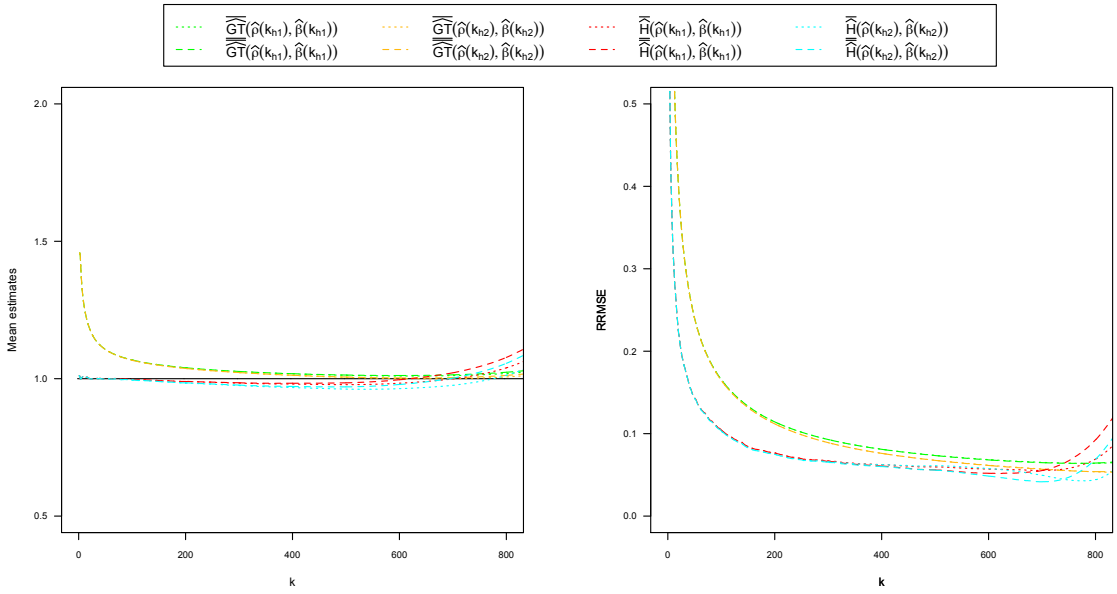


Figure 5.8: Mean estimates (left) and RRMSE (right) of $\overline{\widehat{GT}}$, $\overline{\overline{\widehat{GT}}}$, $\overline{\widehat{H}}$ and $\overline{\overline{\widehat{H}}}$, with $\widehat{\rho}$ and $\widehat{\beta}$ computed at the levels $k_{h1} = \lfloor n^{0.995} \rfloor$ and $k_{h2} = \lfloor n^{0.999} \rfloor$, for a sample size n=1000 (and 2000 replicates), as a function of k, from a Burr distribution given by $F(x) = 1 - \left(1 + x^{-\rho/\gamma}\right)^{1/\rho}$, $x \geq 0$, with $\gamma = 1$ and $\rho = -2$ ($\beta = 1$; $\tau = 0.5$).

From Figures 5.7 and 5.8, we observe that using GPD and Burr distribution, both the geometric-type and the Hill bias corrected estimators present a good performance. Particularly, we note that for GPD the geometric-type estimator has a better posture for intermediate k-values, while the best behaviour of Hill estimator takes place at low values of k. In the case of Burr distribution, a greater distance from the target value is notable at low k-values for the geometric-type estimators, whereas for the Hill estimators the same it is visible for high k-values.

The Hill estimator exhibits in general a lower RRMSE than the geometric-type estimator, which can be understood considering that the asymptotic variance of the Hill estimator is half of the one of the geometric-type estimator.

In addition, for GPD and for large k, the estimates based on $\overline{\widehat{H}}$ clearly show far better results than those conducted with $\overline{\overline{\widehat{H}}}$. Unlike what happens with the corrected geometric-type estimators, the corrected Hill ones have the best estimates when the second order parameters are computed using the level k_{h1} instead of using the k_{h2} level, except for very high k-values in which prevails the use of k_{h2}.

We may conclude that the behaviour of the geometric-type estimator is improved by bias correction. The corrected versions show a good performance and for some cases it is even highlighted.

5.2 High quantiles estimation

In this section we perform a simulation study in order to examine and compare the finite sample behaviour of the different quantiles estimators arising from the use of both the geometric-type and Hill estimators. In a similar way to the previous section, first we analyse the effects caused by using only the GT and Hill estimators in their original forms and after we compare the result of the application of their corrected versions. The quantiles estimators presented here were computed for $p = 0.001$.

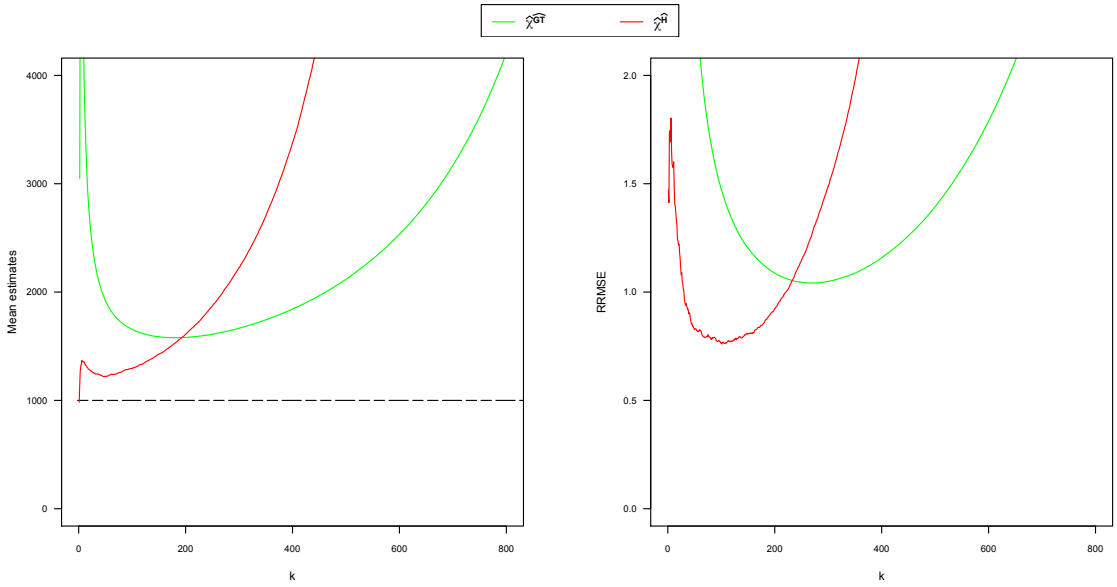


Figure 5.9: Mean estimates (left) and RRMSE (right) of $\widehat{\chi}_{0.999}^{\widehat{GT}}$ and $\widehat{\chi}_{0.999}^{\widehat{H}}$, for a sample size n=1000 (and 2000 replicates), as a function of k, from a GPD given by $F(x) = 1 - (1+\gamma x)^{-1/\gamma}$, $x \geq 0$ with $\gamma = 1$ (empirical quantile $\chi_{0.999} = 999$).

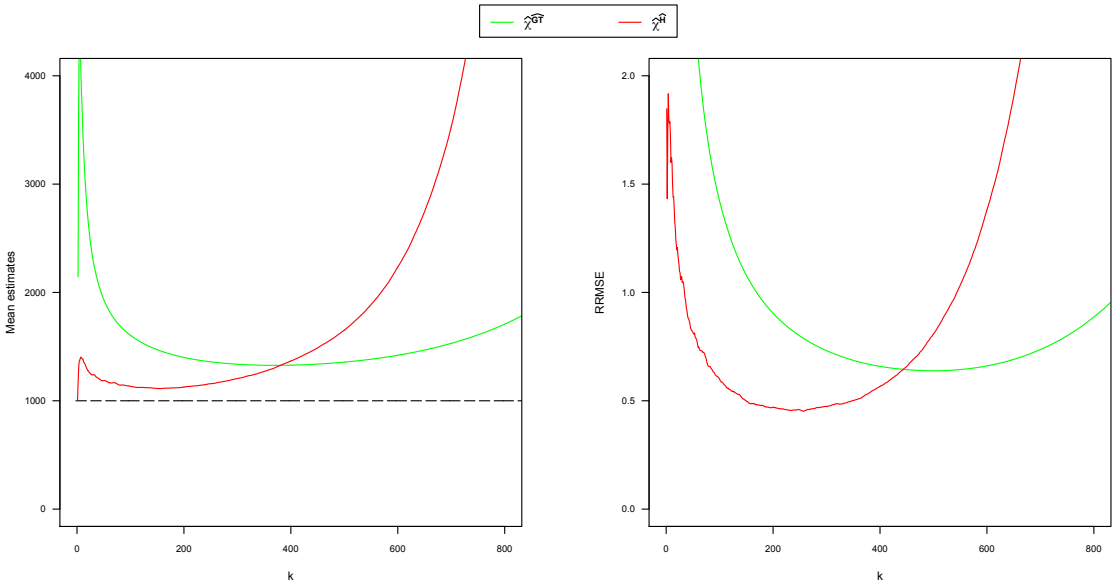


Figure 5.10: Mean estimates (left) and RRMSE (right) of $\widehat{\chi}_{0.999}^{\widehat{GT}}$ and $\widehat{\chi}_{0.999}^{\widehat{H}}$, for a sample size n=1000 (and 2000 replicates), as a function of k, from a Burr distribution given by $F(x) = 1 - \left(1 + x^{-\rho/\gamma}\right)^{1/\rho}$, $x \geq 0$, with $\gamma = 1$ and $\rho = -2$ (empirical quantile $\chi_{0.999} = 1000$).

In Figures 5.9 and 5.10 we illustrate the behaviour of the proposed quantiles estimator using the GT estimator, in (3.12), and the quantiles estimator using the classical Hill estimator, given by

$$\widehat{\chi}_{1-p}^{\widehat{H}} = X_{(n-k,n)} \left(\frac{k}{np} \right)^{\widehat{H}(k)}.$$

We observe that, for GPD and Burr distribution, the quantiles estimator using the geometric-type estimator shows more stability than using the Hill estimator. This last estimator presents a better behaviour only for very small values of k.

In order to get stochastic bounds we consider the results of Theorem 3.3.2, that lead to the following $(1-\alpha)$-level confidence intervals for χ_{1-p}:

$$I_{\widehat{\chi}_{1-p}^{GT}}(k,\alpha) = \left\{ \chi_{1-p} : \frac{k^{1/2}}{\sqrt{2}\gamma^2 U(n/k)\, q_\gamma(d_n)} \left| \widehat{\chi}_{1-p} - \chi_{1-p} \right| \leq \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \right\}, \qquad (5.2)$$

where d_n and $q_\gamma(t)$ are defined in Theorem 4.3.2. For the construction of confidence intervals, the parameters are replaced by their consistent estimators.

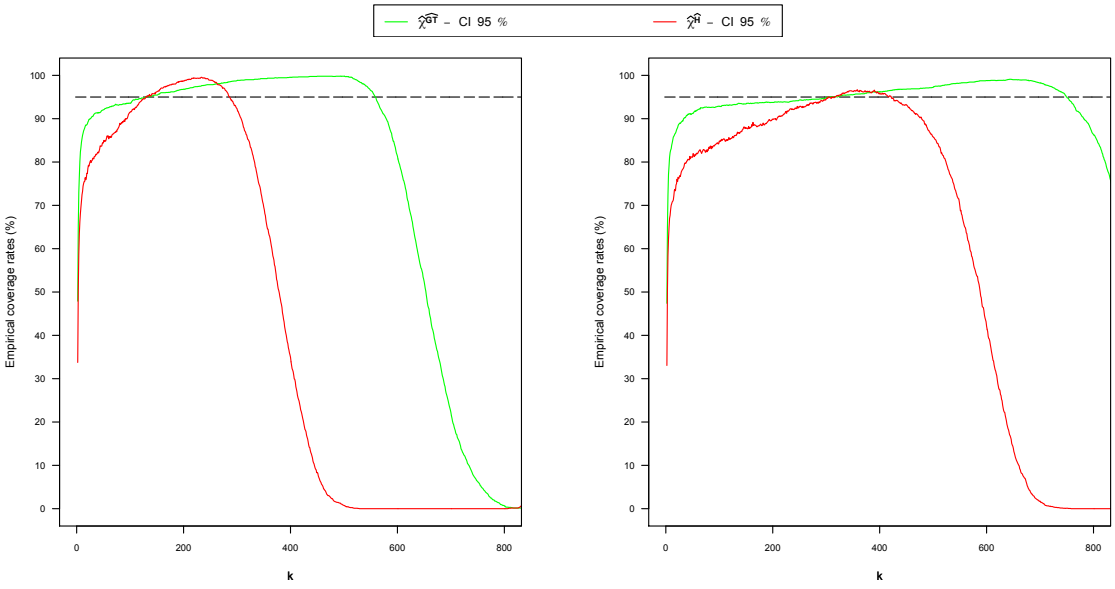


Figure 5.11: Empirical coverage rates of the 95 % confidence bounds for the high quantiles estimator based on the geometric-type (in green) and Hill (in red) estimators for a sample size n=1000 (and 2000 replicates), as a function of k. GPD (left) $F(x) = 1-(1+\gamma x)^{-1/\gamma}$, $x \geq 0$ with $\gamma = 1$, and Burr distribution (right) $F(x) = 1 - \left(1 + x^{-\rho/\gamma}\right)^{1/\rho}$, $x \geq 0$, with $\gamma = 1$ and $\rho = -2$.

The confidence bounds for the corresponding quantiles estimators using geometric-type bias corrected estimators are similar to the previous ones.

The empirical coverage rates of the 95% confidence bounds for both GPD and Burr distribution are illustrated in Figure 5.11. The coverage rates obtained for the geometric-type based estimator are very satisfactory.

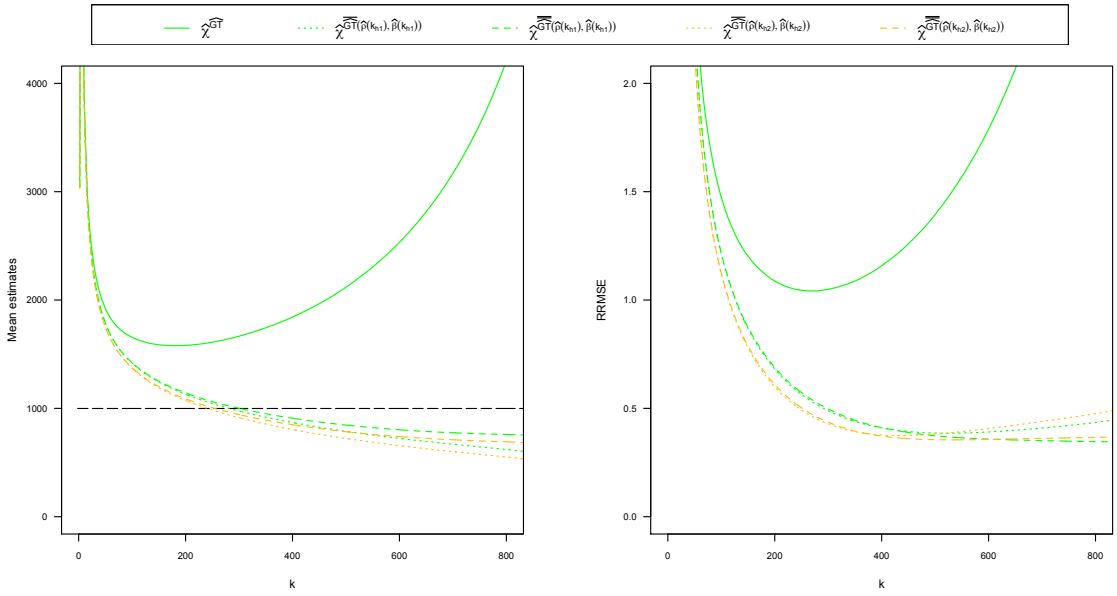


Figure 5.12: Mean estimates (left) and RRMSE (right) of $\widehat{\chi}_{0.999}^{\widehat{GT}}$, $\widehat{\chi}_{0.999}^{\widehat{\overline{GT}}}$ and $\widehat{\chi}_{0.999}^{\widehat{\overline{\overline{GT}}}}$, with $\widehat{\rho}$ and $\widehat{\beta}$ computed at the levels $k_{h1} = \lfloor n^{0.995} \rfloor$ and $k_{h2} = \lfloor n^{0.999} \rfloor$, for a sample size n=1000 (and 2000 replicates), as a function of k, from a GPD with $\gamma = 1$ ($\rho = -1$, $\beta = 1$; $\tau = 0$ and empirical quantile $\chi_{0.999} = 999$).

For comparing the bias corrected estimators we consider again the parameter ρ (depending on τ) proposed by Fraga Alves et al. (2003) and the β estimator obtained in Gomes and Martins (2002). As a consequence of the results in last section, we chose the tuning parameter $\tau = 0$ for GPD and $\tau = 0.5$ for Burr distribution since the estimates using these values show more stability.

In Figures 5.12 and 5.13 we compare the finite sample behaviour of the quantiles estimator using the geometric-type estimator, in (3.12), and the corresponding geometric-type bias corrected estimators, $\widehat{\chi}_{0.999}^{\widehat{\overline{GT}}}$ and $\widehat{\chi}_{0.999}^{\widehat{\overline{\overline{GT}}}}$.

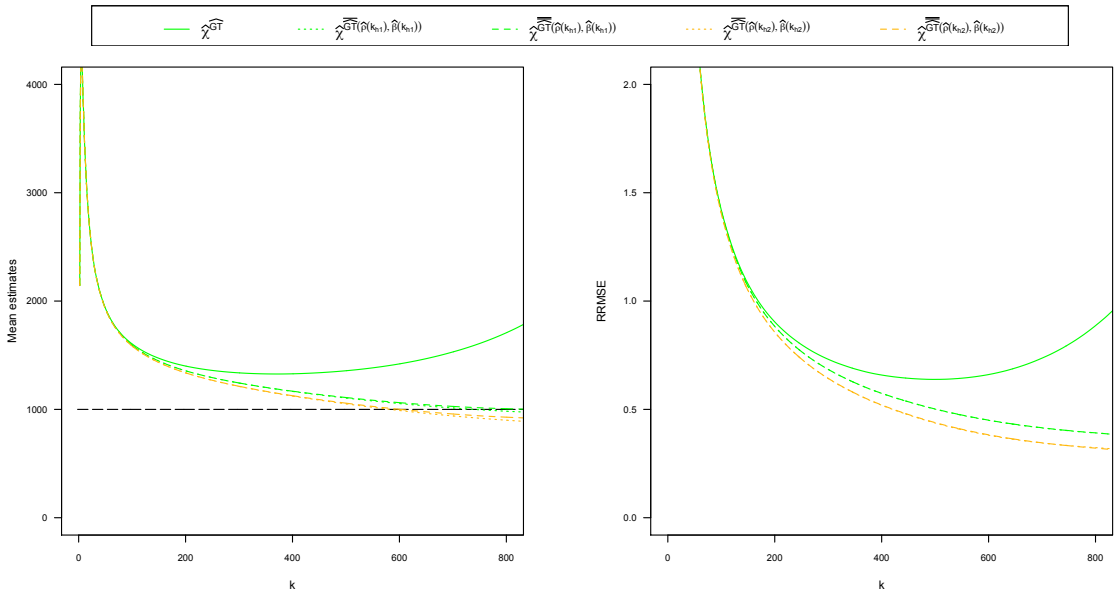


Figure 5.13: Mean estimates (left) and RRMSE (right) of $\widehat{\chi}^{\widehat{GT}}_{0.999}$, $\widehat{\chi}^{\overline{\widehat{GT}}}_{0.999}$ and $\widehat{\chi}^{\overline{\overline{\widehat{GT}}}}_{0.999}$, with $\widehat{\rho}$ and $\widehat{\beta}$ computed at the levels $k_{h1} = \lfloor n^{0.995} \rfloor$ and $k_{h2} = \lfloor n^{0.999} \rfloor$, for a sample size n=1000 (and 2000 replicates), as a function of k, from a Burr distribution with $\gamma = 1$ and $\rho = -2$ ($\beta = 1$; $\tau = 0.5$ and empirical quantile $\chi_{0.999} = 1000$).

We observe that the quantiles estimator using the geometric-type bias corrected tail index estimators presents a better performance than using the standard one. It should be mentioned that as the value of k increases, the corrected estimators have a more stable behaviour and, from a certain value of k, the performance of the estimator is better when the level k_{h1} is used for compute $\widehat{\rho}$ and $\widehat{\beta}$; we note however that the behaviour seems to improve a bit more if the k_h level is lower.

For comparing the performance of high quantiles, we also compare the quantiles estimator using the geometric-type bias corrected estimators, for the same two distributions, with the corresponding quantiles estimator based on the Hill bias corrected tail index estimators.

From Figures 5.14 and 5.15, we may observe that all of the estimators present a good behaviour, more protruding for Burr distribution.

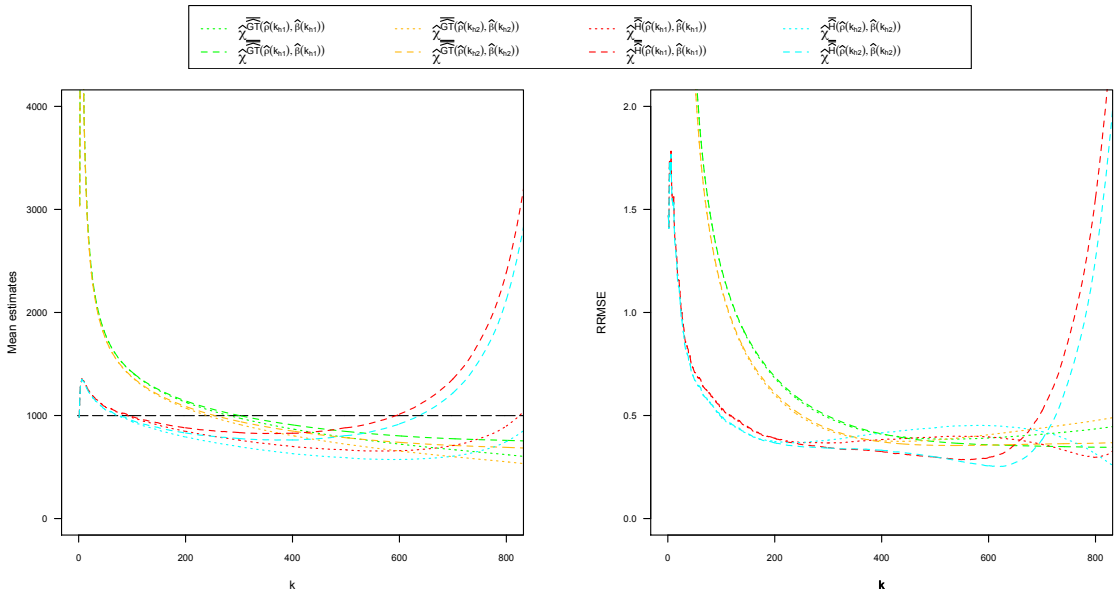


Figure 5.14: Mean estimates (left) and RRMSE (right) of $\widehat{\chi}^{\overline{GT}}_{0.999}$, $\widehat{\chi}^{\overline{\overline{GT}}}_{0.999}$, $\widehat{\chi}^{\overline{H}}_{0.999}$ and $\widehat{\chi}^{\overline{\overline{H}}}_{0.999}$, with $\widehat{\rho}$ and $\widehat{\beta}$ computed at the levels $k_{h1} = \lfloor n^{0.995} \rfloor$ and $k_{h2} = \lfloor n^{0.999} \rfloor$, for a sample size n=1000 (and 2000 replicates), as a function of k, from a GPD with $\gamma = 1$ ($\rho = -1$, $\beta = 1$; $\tau = 0$ and empirical quantile $\chi_{0.999} = 999$).

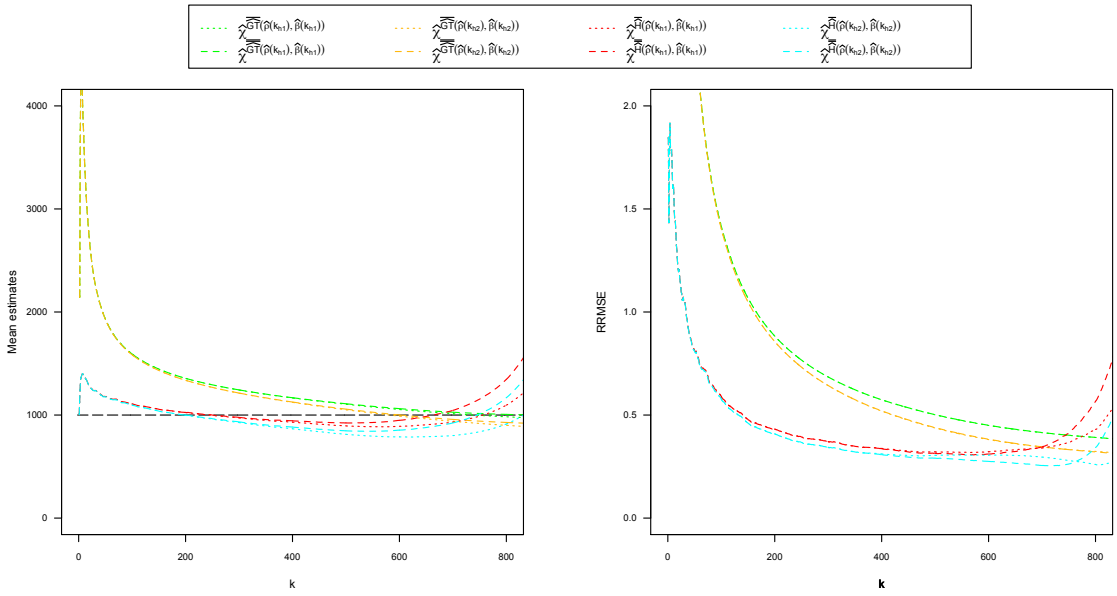


Figure 5.15: Mean estimates (left) and RRMSE (right) of $\widehat{\chi}^{\overline{GT}}_{0.999}$, $\widehat{\chi}^{\overline{\overline{GT}}}_{0.999}$, $\widehat{\chi}^{\overline{H}}_{0.999}$ and $\widehat{\chi}^{\overline{\overline{H}}}_{0.999}$, with $\widehat{\rho}$ and $\widehat{\beta}$ computed at the levels $k_{h1} = \lfloor n^{0.995} \rfloor$ and $k_{h2} = \lfloor n^{0.999} \rfloor$, for a sample size n=1000 (and 2000 replicates), as a function of k, from a Burr distribution with $\gamma = 1$ and $\rho = -2$ ($\beta = 1$; $\tau = 0.5$ and empirical quantile $\chi_{0.999} = 1000$).

We note however that the quantiles estimator using the geometric-type bias corrected estimators is better for certain range of values of k. The last comments about the impact of the level k_h on the performance of the geometric-type corrected estimators, are also valid in the performance of the Hill corrected estimators. It is also clearly visible that for high values of k and mainly in the GPD case, the quantiles estimators when using the Hill corrected estimators $\widehat{\chi}^{\overline{H}}$ show a much more stable behaviour than using the $\widehat{\chi}^{\overline{\overline{H}}}$ estimators.

The comments in Section 5.1 about the lower RRMSE that the Hill estimator generally presents, are also valid for this case.

For smaller sample sizes the estimators show a more irregular behaviour, as expected, but, in general, the relative performance is similar and thus the results are not reported here.

Chapter 6

Modelling extremal earthquakes

6.1 Motivation

The earthquakes are present in everyday life of humanity worldwide. A severe earthquake is one of the most frightening and destructive phenomena of nature. Experiencing an earthquake is certainly the worst experiences anyone can have. The lived moments are reported as full of panic, terror and death. For survivors, the terrible images remain in memory and become part of their daily lives, as well as the constant fear within each based on the possibility of a next big earthquake that can take lives and separate families forever. It is estimated that there are about one million earthquakes per year, however, the vast majority occur in the mid of oceans or in sparsely populated regions and they pass relatively unnoticed by the population. There are annually about 20 earthquakes that cause significant damage and some deaths. On average, only one catastrophic earthquake occurs per year and a highly catastrophic every 5 years.

Since the phenomena that trigger it is still a topic of study and that there are uncontrollable forces of nature that dominate them, they are actually considered unpredictable and mankind will have to learn to live with them. Thus, it is important that their study is oriented to the reduction of the number of deaths and economic losses. It constitutes an important challenge which should be considered a priority among the scientific

community and that can not be successfully tackled without a large multidisciplinary effort.

The EVT may have a relevant role because through it is possible to obtain important information, such as estimating the probability of occurrence of a large earthquake over a long period of time or the time interval until the eventual realisation of such a catastrophe. Hence, the accurate estimation of such quantities turns out to be very relevant to allow the implementation of some adequate prevention measures. Since the POT method has led to satisfactory approximations of the tails, here we use it to appropriately modelling the distribution of extremes, characterising these extremal events through the GPD.

6.2 Earthquake background: basic concepts and definitions

This section provides a basic understanding of earthquakes starting with a brief history of seismology followed by a discussion on the earthquake's causes while defining some commonly used terms, then explaining on how earthquakes can be measured, and finally ends with a discussion about the earthquake's forecast.

6.2.1 A brief history of seismology

The term seismology started to be used around the middle of the nineteenth century, and it is derived from the greek words *seismos*, shaking, and *logos*, science, meaning the science that studies earthquakes.

The early thinking about earthquakes was, as one might expect, superstitious and not very scientific. An earthquake was viewed as an act of God or other supernatural power imposed on mankind as punishment for misbehaviour. According to Ben-Menahem (1995), Aristotle (340 B.C.E.) believed that winds in subterranean caves not only caused earthquakes but produced the large sea waves that sometimes accompanied them. In 1678 A. Kircher related earthquakes and volcanoes to a system of channels of

fire inside the Earth and, in 1703, M. Lister and N. Lesmery proposed explosions of concentrated material that compose the internal fire as the cause of earthquakes. This explanations was widely accepted.

The great Lisbon earthquake of 1 November 1755, which caused widespread destruction in that city and produced a large tsunami, may be considered the starting point of modern seismology. This event changed dramatically man's outlook on the phenomenon of earthquakes. In 1761, John Mitchell, that still held to the *explosive theory of earthquakes*, established that earthquakes are due to the propagation of elastic waves inside the Earth. In the early 1800s the theory of elastic wave propagation began to be developed by Cauchy, Poisson, Stokes, Rayleigh, among others.

The first catalog for the whole world appeared in 1840 and was published by Von Hoff. Mallet's detailed study of the Napolitan earthquake of 1857 constitutes one of the first basic works of modern seismology. Mallet, who funded the instrumental seismology, described the idea that earthquakes radiate seismic waves away from focus point and connected the occurrence of earthquakes with changes in the earth's crust that are often attended by dislocations and fractures. J. Forbes designed the first *seismometer* in 1841 while the first *seismograph* was built by L. Palmieri in 1855. The first useful seismograph was developed by J. Milne in 1880. In 1889 the first teleseismic record, an earthquake from Japan, was identified by E. Paschwitz. This event is considered as the birth of the science of seismology. Since this date several improvements were made in the measure instruments.

A rapid progress was achieved during the following years. In 1895 F. Omori established a law for aftershock time series. After, R. Oldham, A. Mohorovičić, B. Gutenberg and I. Lehmann, among others, had found some evidences about the earth's interior.

In 1928, Kiyoo Wadati reported the first convincing evidence for deep focus earthquakes. In 1935 C. Richter introduced an instrumental magnitude scale. From 1942 to 1956, B. Gutenberg and F. Richter establish the first empirical relations between earthquake magnitude, intensity, energy, acceleration and frequency of occurrence. The results provided by J. Steketee in 1958 leads to the definition of a *source moment*

equal to μAD, where μ is the rigidity over the fault of area A and dislocation D. In the 1960s, seismologists were able to show that the focal mechanisms of most global earthquakes are consistent with that expected from plate tectonic theory, validating the relation between earthquakes and plate boundaries.

The advent of computers in the 1960s changed the nature of terrestrial seismology, by enabling analyses of large data sets and more complicated problems, and led to the routine calculation of earthquake locations. All these advances and the development of computers put seismology in a position where it could exploit the rich information inherent in seismic signals, on both global and local scales. The knowledge of the infrastructure of the Earth's interior and the nature of seismic sources has significantly grown.

However, the progresses that have been made in order to achieve the ultimate goal of seismology, the prediction of earthquakes, is not yet satisfactory. As seismology has always been an interdisciplinary science, the prediction must be first and foremost recognized as a problem at the junction of sciences and it success is highly dependent of a concentrated interdisciplinary research effort. Meanwhile, the prediction of earthquakes would require the unlikely capability of knowing all of its so many factors with great accuracy. Thus, there are a consensus that earthquakes could well be inherently unpredictable in a practical sense. But many scientists are still hopeful about general earthquake forecasting, ie, instead of predicting specific events over short time scales, they hope to forecast the probability of earthquakes over longer periods. The term forecast is more adequate to describe this type of prediction. The probability forecasts associated to earthquakes is the most important target of contemporary seismology and the use of proper mathematical tools still being a need.

For more details, see e.g. Howell (1990), Shearer (2009) and Chen and Scawthorn (2003).

6.2.2 Causes of earthquakes

In general, everything in nature tends to the equilibrium. Due to the thermodynamic equilibrium, the constituents of the Earth's interior are in constant motion. Boosted by this movement that causes friction with its bottom, the tectonic plates move and interchange slowly, thereby contributing to the constant evolution of the terrestrial relief.

The earthquakes mainly arise due to forces within the earth's crust tending to displace one mass of rock relative to another. Each time the plates interact with each other, a large amount of energy is accumulated in its rocks. When its elasticity limit is reached, they will fracture and instantly release all the energy that had been accumulated during the elastic deformation causing vibrations, called seismic waves, which travel outwards in all directions from the fault and give rise to violent motions at the earth's surface, unleashing an earthquake.

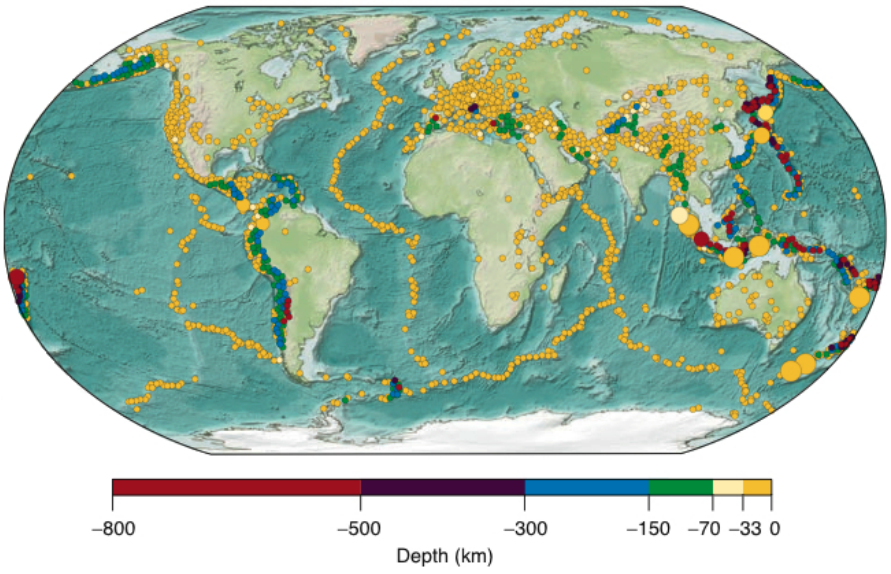


Figure 6.1: Global distribution of earthquakes for 2004 (colors indicate the earthquake depths). (From The Good Earth: Introduction to Earth Science. McConnell et al. (2007). Courtesy of David McConnell, David Steer, Catherine Knight, Katharine Owens and Lisa Park, with permission of McGraw-Hill Education LLC, Copyright 2008, McGraw-Hill).

So, the earthquakes are natural shocks, in which the ground quake strongly in the

matter of seconds to minutes, that occur as a result of this sudden release of a huge amount of that energy slowly-accumulated over many years. If the earthquake is large enough, the seismic waves are recorded on seismographs around the world and can cause the ground to quake strongly.

Earthquakes do not occur at random but are distributed according to a well-defined pattern. About 90% of earthquake activity is associated with plate-boundary processes, so the global seismicity patterns reveals a strong correlation between plate boundaries and the presence of intercontinental fault zones, indicating that earthquakes often occur at tectonic plate boundaries. We can say, without committing a gross error, that the alignments of earthquakes indicate the boundaries of tectonic plates (see Figure 6.1).

After the initial fracture, a number of secondary ruptures corresponding to the progressive adjustment of fractured rocks may occur, causing successive lower intensity earthquakes called aftershocks. If these vibrations occur at the sea floor they can produce a long and smooth waving that in shallow water becomes authentic water columns known as tidal waves or tsunamis.

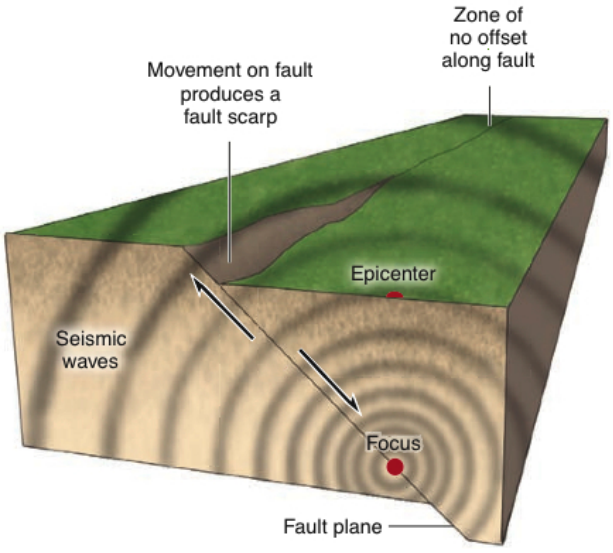


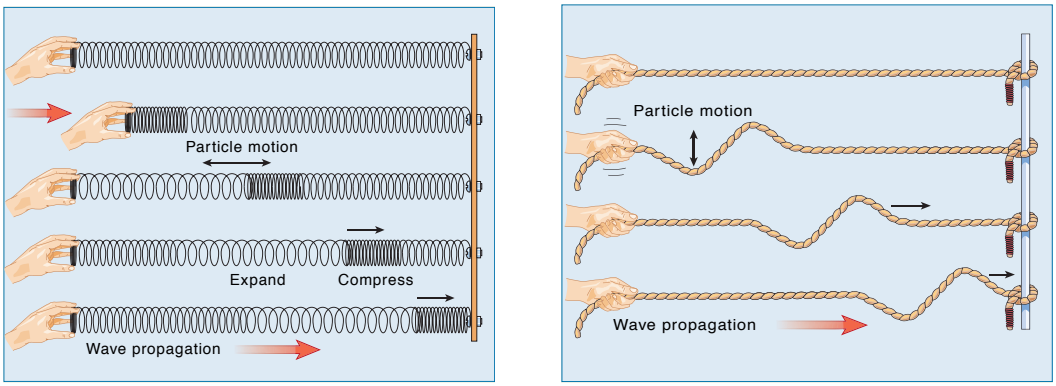
Figure 6.2: Earthquake features. (From The Good Earth: Introduction to Earth Science. McConnell et al. (2007). Courtesy of David McConnell, David Steer, Catherine Knight, Katharine Owens and Lisa Park, with permission of McGraw-Hill Education LLC, Copyright 2008, McGraw-Hill).

The point on the fault where the radiation of seismic waves begins is termed the hypocenter or focus. The epicenter is the projection on the surface of the Earth directly above the hypocenter. The movement begins at the focus and travels outward along the fault surface (see Figure 6.2).

The earthquakes initiate a number of phenomena, termed seismic hazards, such as shaking, tsunami or liquefaction, which can cause significant damage to the built environment and a great loss of life. They not only destroy villages, towns and cities but the aftermath leads to destabilise the economic and social structure of the nation.

Therefore, earthquakes, such as volcanoes, represent the more energetic and rapid manifestations of the planet's internal dynamics.

The fault rupture generates a wave phenomenon similar to the effect of a stone dropped into a pool of water, since the seismic waves radiate out in all directions from the earthquake's hypocenter. There are two basic types of elastic waves that make up the shaking felt and causes damage in an earthquake: the body waves, that travel through the interior of the Earth, and surface waves, travelling only along the Earth's surface.

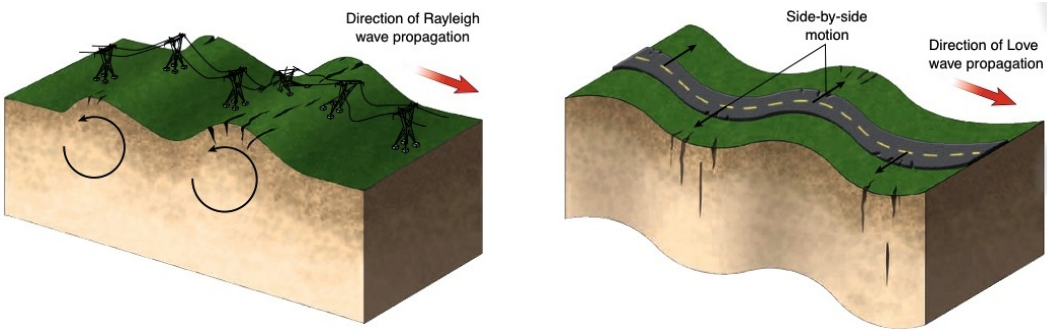


(a) Primary wave (b) Secondary wave

Figure 6.3: P-wave and S-wave motions. (a) P waves are similar to the passage of a vibration through a slinky. The vibration occurs in the same direction that the wave travels. (b) S-wave motion is analogous to a vibration moving along a rope. The vibration occurs perpendicular to the direction in which the wave travels. (From The Good Earth: Introduction to Earth Science. McConnell et al. (2007). Courtesy of David McConnell, David Steer, Catherine Knight, Katharine Owens and Lisa Park, with permission of McGraw-Hill Education LLC, Copyright 2008, McGraw-Hill).

The **body waves**, primary waves (P) and secondary waves (S), transmit foreshocks that have little destructive power whereas **surface waves**, Love waves and Rayleigh waves, produce tremors in all directions causing the most destruction (see Figures 6.3 and 6.4). Surface waves result from the interaction between body waves and the superficial earth materials and usually have the strongest vibrations probably causing most of the damage done by earthquakes.

The faster of the body waves are appropriately called the primary waves (P). They travel in straight lines, alternately compressing and dilating solids and liquids they pass through, so its energy is transmitted via push–pull motion. Similar to sound waves, they are able to travel through all types of material. The slower waves are called the secondary waves (S). As they propagate, they shear the rocks sideways at right angles to the direction of travel so, at the ground surface, S waves can produce both vertical and horizontal motions. Because liquids have no shear resistance, the S waves cannot propagate in the liquid parts of the Earth, travelling only through solids.



(a) Rayleigh wave (b) Love wave

Figure 6.4: Two types of surface waves. Highways and electric transmission lines are examples of man-made structures that can be destroyed by an earthquake. (a) Rayleigh waves produce vertical motions of the land surface. (b) Love waves move sideways, but not vertically. (From The Good Earth: Introduction to Earth Science. McConnell et al. (2007). Courtesy of David McConnell, David Steer, Catherine Knight, Katharine Owens and Lisa Park, with permission of McGraw-Hill Education LLC, Copyright 2008, McGraw-Hill).

The Love waves, horizontally oscillating, moves like S waves that have no vertical displacement, ie, it moves the ground side to side in a horizontal plane parallel to

the Earth's surface, but at right angles to the direction of propagation. The Rayleigh waves, vertically oscillating, spread like rolling ocean waves, ie, move both vertically and horizontally in a vertical plane pointed in the direction in which the waves are travelling, causing fractures perpendicular to their travel by stretching the ground.

The surface waves, appear on the surface after the P and S waves reach the epicenter and travel more slowly than body waves. Having a lower frequency, surface waves have a greater effect on solids, which makes them more destructive. Love waves generally travel faster than Rayleigh waves and the last ones are not recorded by vertical instruments. More details can be found, for example, in McConnell et al. (2007).

6.2.3 Quantification of earthquakes

The scientific analysis of earthquakes requires measurement. The size of an earthquake can be measured in several ways. The early methods used a kind of numerical scale based on a synthesis of observed effects, called the *intensity* scales. Some attempts to relate intensity to the amplitude of ground motion led to a quantity called *magnitude*, based on the records of ground amplitudes normalised for their variation with distance from the earthquake epicenter. However, the known magnitudes present a saturation point which does not allow a correct estimation of the true earthquake size of larger earthquakes, underestimating it. Moreover, it turns out that larger earthquakes, which have larger rupture surfaces, systematically radiate more long-period energy. Then, nowadays, modern seismologists are increasingly turning to describe the physical effects of earthquakes by the estimation of the radiated *energy* or the *seismic moment* of the displaced ground. For more details see e.g. Howell (1990) and Day (2002).

This section defines and discusses each of these measures.

Intensity

Prior to the invention of modern scientific instruments, earthquakes were qualitatively measured by their observed effects, ie, the effects on humans and their structures, in a kind of numerical scale, which differed from place to place. Such measures are called intensity scales and are the oldest useful way to express the “strength” of an earthquake.

The first intensity scale was developed by de Rossi and Forel in the 1880s. It originally had values from I to X and were based on the observation of the effects of seismic activity. A more refined scale was devised in 1902 by the italian volcanologist and seismologist Mercalli expressing the intensity of an earthquake’s effects in a given locality in values ranging from I to XII. The first few levels consist of barely perceptible sensations and the highest levels apply to the destruction of buildings. This scale is widely used to compare levels of damage among different regions and socioeconomic conditions. Mercalli’s scale was modified by Sieberg in 1912 and by H. Wood and F. Neumann in 1931. The Wood-Neumann version is still in use. Earthquake intensity scales, specially the Rossi-Forel and various versions of the Mercalli scale, were used almost universally to measure earthquake size for about 50 years.

Although the intensity scale is relatively easy to use and is helpful in gauging the human impact of an earthquake, it is not widely applied in the scientific analysis of modern earthquakes for several reasons. The scale is based on damage, but the amount of damage depends on how many people live in a particular area and the number of buildings there. A minimal damage in sparsely populated areas with few buildings are usually underestimated while highly populated areas with many buildings might be overestimated. Also, there are significant differences in individual interpretations, one person may define “considerable damage” differently from another person, and different types of buildings constructions lead to a great difference in live losses. Moreover, it depends upon the observer’s location relative to the earthquake’s epicenter, since damage generally decreases moving away from the epicenter, but the decreasing is not the same in every direction. Thus, the amount of damage depends on several factors

that may be unrelated to the earthquake itself.

Since intensity scales are extremely subjective, they are ineffective for the scientific comparison of earthquakes. However, the scale does provide useful local data that can be used to identify which areas are most susceptible to shaking in regions of rare earthquake activity. The Modified Mercalli scale is also useful in evaluating damage from regions having insufficient seismographs and in historical earthquakes that occurred before the widespread use of seismographs.

Magnitude

One of the weaknesses that earthquake intensity scales had is that they could be applied only where there were observers to note the effects or structures to be affected. If sizes of earthquakes are to be compared world-wide, a measure is needed that does not depend on the density of population or type of construction, which can be used to compare the strength of earthquakes apart from their effects.

Although similar seismographs have existed since the 1890s, it was only in the 1930s that Charles F. Richter and Kiyoo Wadati introduced the concept of earthquake *magnitude*, a strictly quantitative scale that can be applied to earthquakes in both inhabited and uninhabited regions. Such a measure define the magnitude of a local earthquake as the logarithm of base 10 of the maximum seismic wave amplitude recorded on a Wood-Anderson seismograph located at a distance of 100 kilometers from the earthquake epicenter. An important feature of this scale is that the levels increase exponentially, ie, the rise of one level on the scale represents 10 times of increase on the ground motion and an approximate 32 times of increase in energy released. The smallest quakes normally felt by people have magnitude 2, from a magnitude of 6 they are commonly considered major and great earthquakes have magnitude of 8 or more. This procedure allows that all stations were able to determine the same earthquake magnitude for a given quake.

The original Richter magnitude is called *local magnitude*, m_l, because it varies from

place to place depending on variations in the local geology. As instruments became more sophisticated, it became clear that it could not accurately measure the largest earthquakes and a variety of magnitude scales have emerged. “Richter scale” has actually fallen out of use by the scientists who study earthquakes.

Once earthquakes excite both body waves and surface waves, two magnitude scales evolved: the *body-wave magnitude*, m_b, and the *surface-wave magnitude*, m_s. These magnitudes measure the size of ground motions at very different periods of vibration: the m_s magnitude scale measures the amplitude of Rayleigh waves in the period range from 18 to 22 seconds while the m_b scale is based in the compressional body P wave amplitudes at a period of about 1 second.

Another commonly used magnitude scale is the *moment magnitude*, m_w, but since it has a different concept behind and it is strongly related with the seismic moment, we only present this magnitude in the next topic.

Seismic Moment

Although theoretically there is no upper bound to the magnitude, it was found that it underestimates the true size of larger earthquakes, since the m_b scale saturates about 6.5 - 6.8, and the saturation point of m_s is about 8.3 - 8.7. However, scientists have found a correlation between fault length and earthquake magnitude: the longer the fault rupture, the bigger the earthquake. It turns out that larger earthquakes, which have larger rupture surfaces, systematically radiate more long-period energy.

The saturation problem can be avoided if seismic moment is used as a measure of earthquake size rather than magnitude. The seismic moment, M, provides more accurate measures of the energy released from an earthquake taking into account the rock properties, such as its rigidity, μ, the area of the fault plane that actually moves, A, and the amount of movement on the fault, D, and combining these three factors in the following form

$$M = \mu A D.$$

Takuo Maruyama, in 1963, appears to have been the first person to give special attention to this property of ground displacement. After that, the seismic moment received increasing attention as a measure of earthquake size and, nowadays, it is preferably adopted for scientific studies. This measure not only avoids the saturation problem, since it does not have an intrinsic upper bound, but also describes the size of an earthquake as a essential combination of physical quantities that really matters at the earthquake source and that determines how strong the seismic motions will be.

Because many people do not really know what means a number with the "size" of seismic moment and since the magnitude scale has been used for a very long time, the need to convert it into some kind of magnitude scale emerged. These factors have led to the definition of a new magnitude scale, the moment magnitude, m_w, based on seismic moment, given by

$$m_w = \frac{2}{3}\left(\log M - 16.1\right), \tag{6.1}$$

where M is in units of dyne-cm.

The moment magnitude scale is the only magnitude scale which does not suffer from the above mentioned saturation problem for great earthquakes. The reason is that it is directly based on the forces that work at the fault rupture to produce the earthquake and not the recorded amplitude of specific types of seismic waves.

Energy

The energy of an earthquake is the fundamental measure of its size. The total amount of energy released is hard to estimate since to determine it one would have to integrate the energy flux over time and space and include the broadest possible spectrum of frequencies generated by an earthquake as it ruptures a fault. During an earthquake, the stored energy is transformed and results in rock deformations, heat and radiated seismic energy. For those who are concerned with the effects of earthquake shaking, what matters is to estimate the amount of energy released as seismic waves during earthquakes. In 1956, Gutenberg and Richter obtained a relation between the radiated

energy E (in ergs) and the surface-wave magnitude m_s given by

$$\log E = 1.5m_s + 11.8. \tag{6.2}$$

Kanamori (1977) proposed the estimation of the radiated energy with the relation

$$E = \frac{M}{2 \times 10^4}. \tag{6.3}$$

He also found that m_w is approximately equal to m_s below the saturation level of about $m_s \leq 8$, which led, jointly with the combination between (6.2) and (6.3), to the moment magnitude m_w defined in (6.1), that do not saturate and provide a scale that quantifies earthquakes on the basis of the radiated energy.

It is relevant to note that, since $10^{1.5} \approx 31.6$, an increase of one magnitude unit is equivalent to approximately 31.6 times more energy release.

6.2.4 Earthquake forecast

Earthquakes are generally considered harmful because of their potential for causing death and destruction. Although the scepticism that accompanied the rise of the scientific study, the earthquake prediction was always seen as a prime goal because is the most useful thing one could do. When the phrase “earthquake prediction” is used, people usually have in mind the accurate forecasting in means of simultaneous prediction of the time of occurrence, location, and approximate size of a specific earthquake to within a matter of days to months, which has long been an unrealised goal of seismology. There are many who maintain that the nature of the earthquake instability makes prediction impossible on such short timescales.

Much of the risk that earthquakes pose, particularly to property, could not be mitigated by short-term prediction, which is an area of seismological research that is afforded a lot more attention by the general public. Even if short term earthquake prediction should someday prove possible and reliable, it would not be possible, for example, to retrofit large engineered structures on a timescale of days to weeks. Then, some developments have been conducted in the field of engineering in an attempt to predict strong

ground motion through data sets relative, for example, to peak ground acceleration (PGA), peak ground velocity (PGV), spectral acceleration (S_a), spectral velocity (S_v), and spectral displacement (S_d).

After many ideas for earthquake prediction have been explored, the sad truth is that reliable prediction of damaging earthquakes is not currently possible on any time scale. Nevertheless, as referred by Kanamori (1974), earthquake prediction is only possible with some statistical uncertainly and, in this sense, methods are being developed to make this uncertainly small enough for practical purposes. Many advances have been made in the probabilistic characterisation, namely the ability to evaluate the probability of an earthquake occurring at some uncertain time during an extensive time period. Other relevant estimations to be done are the damage and loss estimation. In the damage estimation one usually estimates the shaking in terms of PGA or response spectral acceleration at the site, which provides an estimate of the degree of damage for each individual asset.

An important result in the forecast of earthquakes is, for example, the probability in any given year that ground motions of a given intensity will be exceeded. Often, the probability of exceedance during a period of some years is of most interest, for instance when this period represents the lifetime of a building.

The prediction, in a probability sence, of strong ground motion is arguably one of the most important issues that researchers can address and the most important social benefit from earthquake research is the use of that knowledge to reduce the hazard earthquakes pose to mankind.

6.3 Extreme value modelling of earthquake data

In the present section, the EVT using the POT approach is applied to some earthquake data sets in the Harvard Seismic Catalog in order to estimate the parameters quantifying the tails of the distribution of the large earthquakes considered.

We concentrate on the entire route one need to travel from the raising of the data to our ultimate goal of modelling the tail of the distribution of earthquakes seismic moment. In order to do this we begin by explaining the entire procedure which was necessary to make it possible to apply the POT approach to the data chosen for the study. The reliable estimation of the parameters of the earthquake size distribution is only possible if some usual assumptions required by the general results are satisfied. For this purpose, a first analysis in order to verify the validity of these assumptions is performed. Also in this section, an application of the core extreme value analysis to these real earthquake data sets is provided with the estimation of the tail parameters of the seismic moment distribution. The geometric-type and the Hill estimators are considered for the estimation of the tail index and are employed on POT estimator for the quantiles estimation. A comparison between the suggested estimators is carried out and their performance is carefully discussed. All the analysis is supported by graphical tools that show in a clear way the properties of the data that are regarded as relevant to the study that is addressed here.

6.3.1 Description of the earthquake data

We consider the earthquake data obtained from the Harvard Seismic Catalog, available at Global Centroid-Moment-Tensor (CMT) web page (cf. e.g. Dziewonski et al. (1981) and Ekström et al. (2012)). Here, we restrict the territory of our study to earthquakes occurring within the Philippines and Vanuatu Islands, and the analysis was performed in a similar way for the both regions. In particular, we extract and analyse the information about their seismic moments covering the period 01.01.1976 - 31.12.2010. The original data-sets contain 1255 events for Philippines Islands and 1012 events for Vanuatu Islands. However, in order to apply the POT method we selected an adequate and large enough threshold $u = 10^{24}$ dyne-cm, that corresponds to a moment magnitude $m_w \approx 5.27$, the same value considered in related works such as in Pisarenko and Sornette (2003). The observations under this threshold were removed. Since we detect a failure in data acquisition of the Vanuatu Islands until 01-01-1980, we just consider

the Vanuatu Islands data subsequent to this date. So the final data sets, on which the analysis that follows has been based, consider 821 cases for Philippines Islands and 647 cases for Vanuatu Islands. We did not exclude the aftershocks because apart from owning a greatly reduced fraction on the range of seismic moments considered, its removal may introduce a bias in the parameters estimation (cf. e.g. Pisarenko and Sornette (2003)). As the considered region has a lot of deep earthquakes, they also were not excluded. Thus, after the space, time and seismic moment has been selected, no further elimination of events is performed.

6.3.2 Heavy tails detection

Before proceeding it would be useful to discuss if the Pareto-type model provide a plausible fit to the seismic moment distributions of the data under study. We analyse it through quantile-quantile (QQ) plots and hypotheses tests.

The QQ plots constitute a very informative and powerful tool to graphically evaluate how close two distributions are from each other, using for it their quantiles. In most of the time, as in this case, the most convenient comparison is between the empirical quantiles of the sample and the quantiles of the theoretical distribution intended. If the sample data and the reference distribution are derived from populations with a common distribution, the QQ plot should show a strong linear trend.

Since we believe that our data are heavy tailed, we present the Pareto QQ plots of our data sets in Figure 6.5.

Given that $Y \stackrel{D}{=} \log X$, where X and Y are Pareto and Exponential distributed r.v., respectively, then the usual Pareto QQ plots are Exponential QQ plots of the log-transformed data.

In the resultant scatterplot a linear pattern is evident, which is indicative of the good agreement between observed values and the values predicted by the model. We carefully analyse the behaviour of the QQ plot on its upper right part, which represents

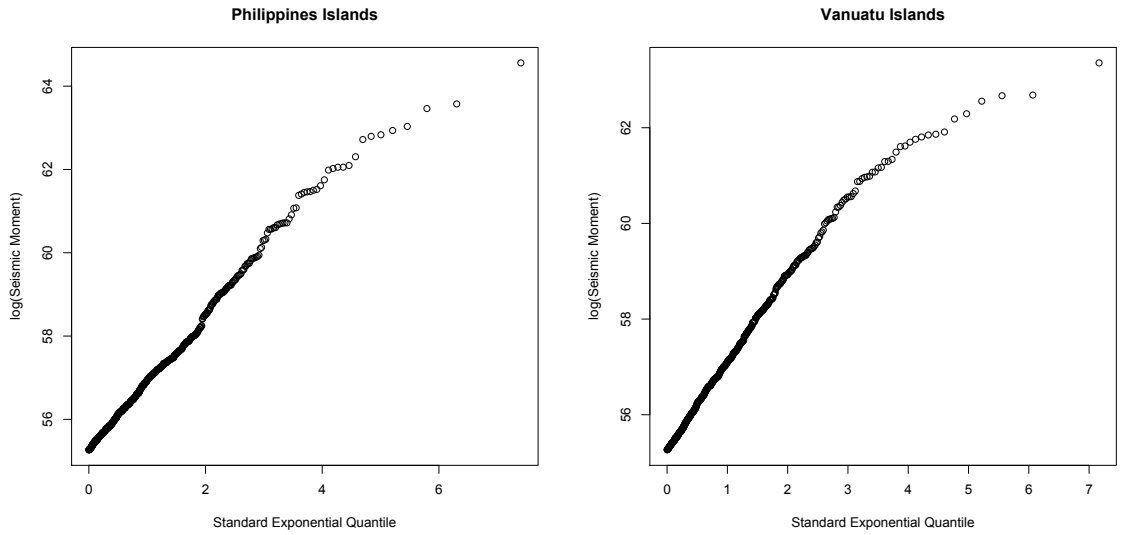


Figure 6.5: Pareto QQ plot for Philippines (left) and Vanuatu (right) Islands seismic moment data.

the most extreme values and, although slightly less than in the remaining part of the plot, a linear tail behaviour is made apparent. The visual impressions based on the Pareto QQ plots suggests that the Vanuatu and Philippines Islands earthquake data sets do seem to follow a Pareto distribution, ie, that we are dealing with a possible heavy-tailed underlying distribution ($\gamma > 0$).

The QQ plots are very useful tool since they provide an indication of the sign of the tail index, thus providing a useful orientation to the choice of the most appropriate estimators.

For the estimation of the quantities of interest we assume that the underlying d.f. F belongs to the domain of attraction of an extreme value distribution, ie, that the extreme value condition (2.1) is satisfied. In order to check the validity of this condition, Dietrich et al. (2002) developed the test statistic

$$E_{k,n} = \int_0^1 \left(\frac{\log X_{(n-kt,n)} - \log X_{(n-k,n)}}{\widehat{\gamma}_+} - \frac{t^{-\widehat{\gamma}_-} - 1}{\widehat{\gamma}_-} \left(1 - \widehat{\gamma}_-\right) \right)^2 t^2 \, dt, \tag{6.4}$$

for $k \ll n$, where $\widehat{\gamma}_+$ is an estimator for $\gamma_+ = \max(0, \gamma)$ and $\widehat{\gamma}_-$ is an estimator for $\gamma_- = \min(0, \gamma)$.

They also refer that in case where only nonnegative values of γ play a role, a simplified version is available

$$T_{k,n} = \int_0^1 \left(\frac{\log X_{(n-kt,n)} - \log X_{(n-k,n)}}{\widehat{\gamma}_+} - \log t \right)^2 t^2 \, dt. \tag{6.5}$$

A table of critical points for various values of γ is available in Hüsler and Li (2006).

In order to test the null hypothesis $H_0 : F \in DA(G_\gamma)_{\gamma>0}$, we consider the E-Test and the T-Test given in (6.4) and (6.5), respectively. In Figures 6.6 and 6.7, we show the behaviour of the E-Test and T-Test statistics, respectively, as a function of k, and considering significance levels $\alpha = 0.01$ and $\alpha = 0.05$.

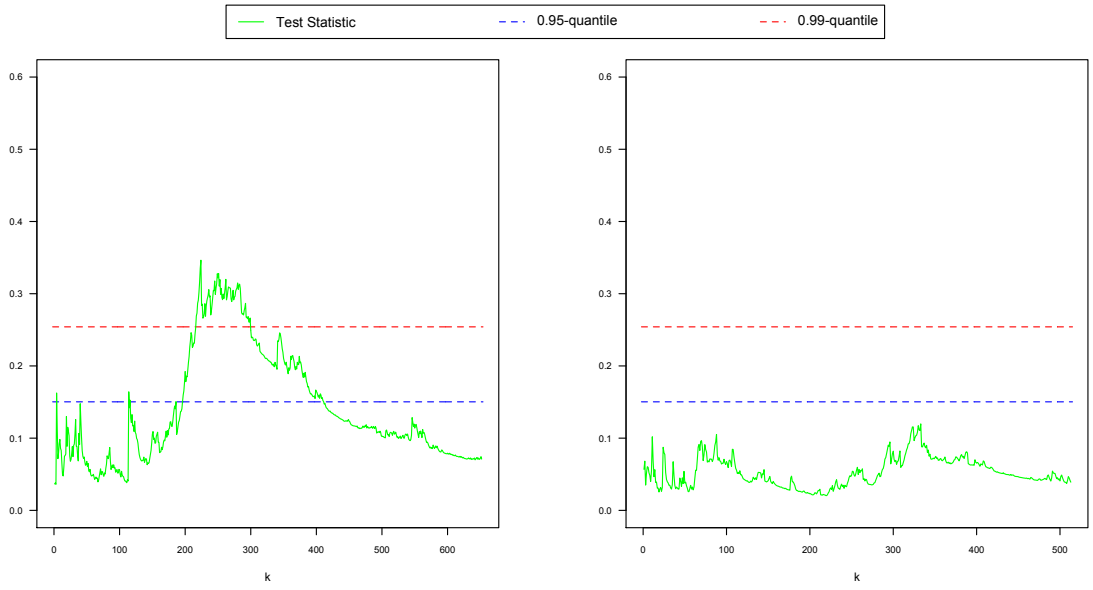


Figure 6.6: Plot of the sample paths for E-Test statistics (Dietrich et al. (2002)), for seismicity of Philippines (left) and Vanuatu (right) Islands, as function of k. The dashed lines represent the asymptotic 0.95-quantile (≈ 0.15) and 0.99-quantile (≈ 0.25).

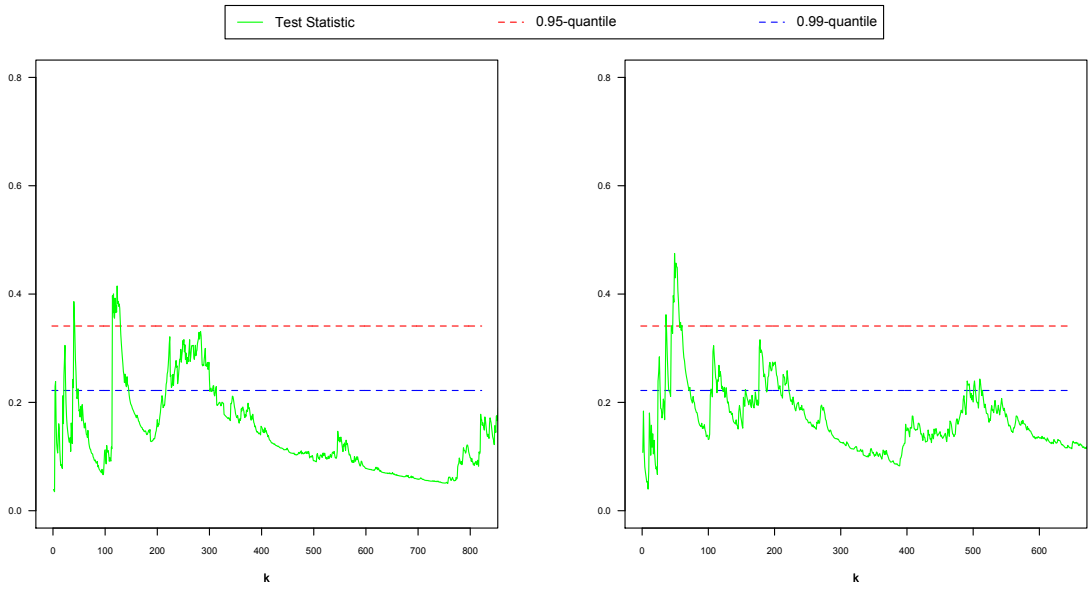


Figure 6.7: Plot of the sample paths for T-Test statistics (Dietrich et al. (2002)), for seismicity of Philippines (left) and Vanuatu (right) Islands, as function of k. The dashed lines represent the asymptotic 0.95-quantile (≈ 0.22) and 0.99-quantile (≈ 0.34).

For Philippines and Vanuatu Islands data sets, we observe that the values for both test statistics are smaller than the corresponding asymptotic quantiles for a large range of k-values. So, since the sample paths of both test statistics are almost always outside the rejection region (i.e. below the critical test value), except for a small range of k, we find no evidence to reject the null hypothesis.

6.3.3 Data stationarity

For the purpose of analysis, the stationarity is a desirable property to have in the data under study, since, in this case, the statistical parameters properties do not change over time. However, strong stationary is never observed in practice but whenever adequate, data is considered approximately stationary. To identify the stationarity we plot the normalised cumulative number of earthquakes versus time.

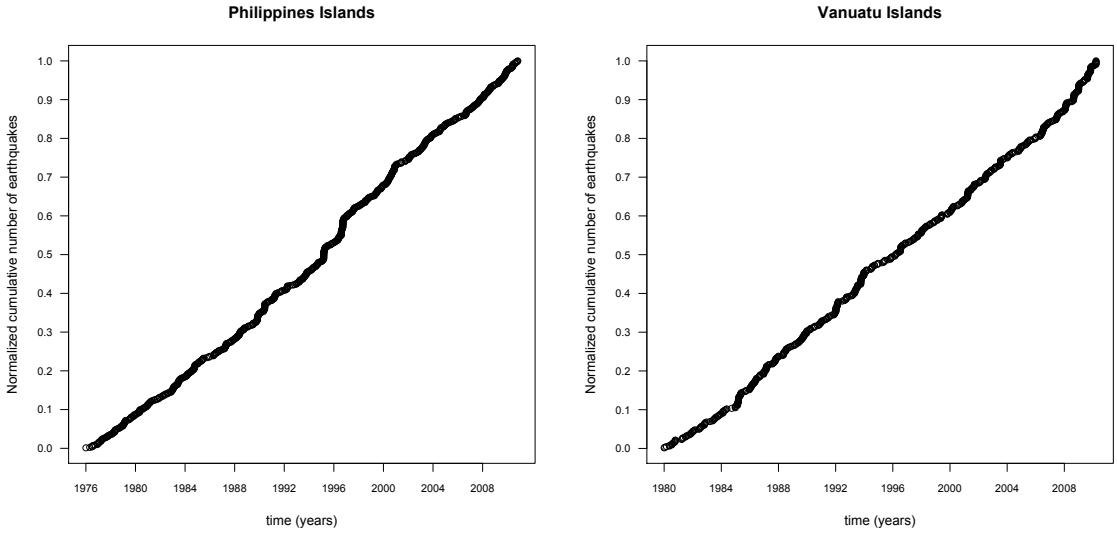


Figure 6.8: Cumulative number of earthquakes normalised by the total number in the period considered as a function of time, for seismicity of Philippines (left) and Vanuatu (right) Islands with $M \geq 10^{24}$.

The linear behaviour that we can observe in Figure 6.8 is an indication of the stationary behaviour of the two data sets over the selected time window, thus the data is approximately homogeneous in time and assumed as stationary.

6.3.4 Investigation of independence

Another relevant property that we are interested to verify before proceeding with the extreme value analysis of the data is the independence, since most of the results in EVT require it as assumption.

In our case, the goal is to analyse the existence of correlations between consecutive seismic moments, ie, verify how the seismic moment of one event, M_{i-1}, influences the seismic moment of the next, M_i.

Here we investigate this statistical dependence through the conditional probability density determined by

$$\frac{P\left(\eta \leq M_i < \eta + \Delta_\eta \mid M_{i-1} \geq M_c'\right)}{\Delta_\eta},$$

where M_c' is the threshold considered on the previous seismic moment when this con-

dition is imposed. Here we denote the initial threshold, u, as M_c, and the condition $M \geq M_c$ is always satisfied (see e.g. Corral (2006)).

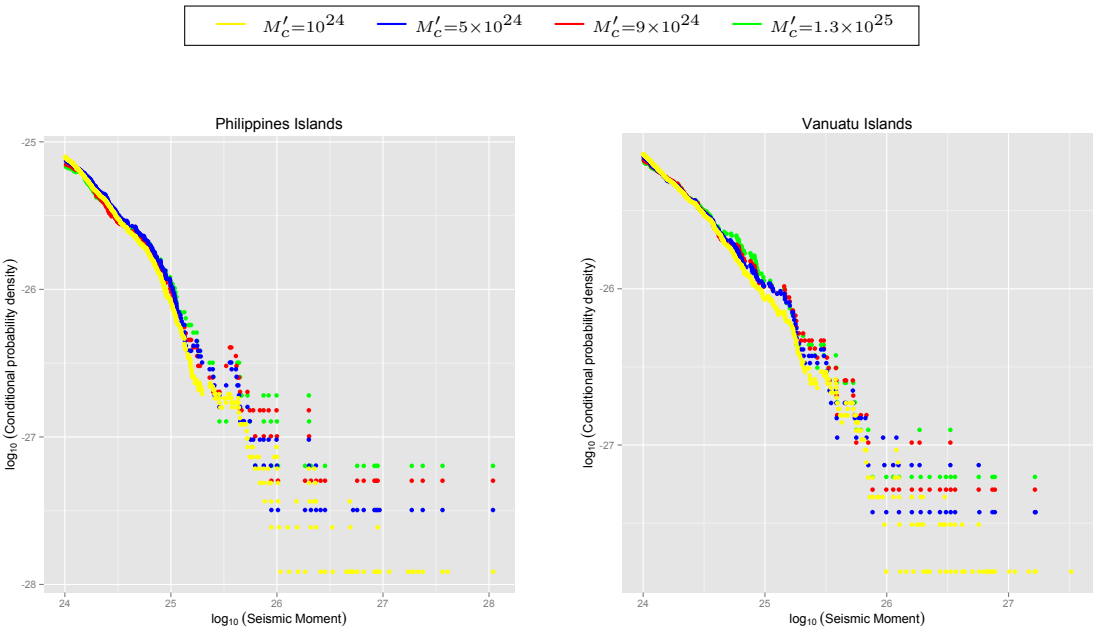


Figure 6.9: Conditional probability densities of earthquake seismic moments, for seismicity of Philippines (left) and Vanuatu (right) Islands, evaluated using different thresholds M'_c and with a constant $M_c = 10^{24}$ ($\Delta_\eta = 10^{25}$).

The conditional probability density of a seismic moment is then defined as the probability of the seismic moments are within a small interval of values, divided by the length of the small interval, Δ_η, tending to zero, considering only the cases in which the seismic moment of the immediately previous event is bigger than a threshold M'_c.

The dependences will be given by the distribution described above. If the conditional distribution of M_i given that $M_{i-1} \geq M'_c$ is identical to the unconditional distribution, then the seismic moment M_i is statistically independent of an event $M_{i-1} \geq M'_c$. Note that the case $M_c = M'_c$ gives the unconditioned distribution.

We observe in Figure 6.9 that, in general, the different densities using different thresholds M'_c share the same properties, which suggest the independence of seismic moments M_i with their history. The small oscillations between the densities may be caused

by the errors associated to the finite sample and the dependence that arises from this is apparently weak enough to lead to major differences in the distributions.

6.3.5 Estimation of tail parameters

In this section we formalise our main objective of investigating the extremal behaviour of the large earthquakes and how the proposed estimators behave with this type of real data.

Then, we discuss the estimation of the tail parameters through the POT approach. The GT and the Hill estimators are considered for the estimation of the tail index and are employed on POT estimator for the quantiles estimation.

Some graphical plots illustrate the tail parameters of large earthquake data, as a function of k.

The bias corrected estimation requires a previous choice of the τ-value more appropriate to be used in the estimation of the second order parameters. As usual, the means whereby we do this choice passes by portraying the sample paths of $\widehat{\rho}_\tau(k)$ in (4.4) for the values $\tau \in \{0, 0.5, 1\}$, as functions of k, and use the following algorithm as a stability criterion for large values of k:

1. Consider $\widehat{\rho}_\tau(k)$, $\tau \in \{0, 0.5, 1\}$, for the integer values $k \in (\lfloor n^{0.995} \rfloor, \lfloor n^{0.999} \rfloor)$ and compute their median, denoted by χ_τ;

2. Choose the *tuning parameter* $\tau^* = \arg\min_\tau \sum_k (\widehat{\rho}_\tau(k) - \chi_\tau)^2$;

3. Compute the ρ estimates $\widehat{\rho}_{\tau^*}(k_{h1})$ and $\widehat{\rho}_{\tau^*}(k_{h2})$, and the β estimates $\widehat{\beta}_{\rho_{\tau^*}(k_{h1})}(k_{h1})$ and $\widehat{\beta}_{\rho_{\tau^*}(k_{h2})}(k_{h2})$, with k_{h1} and k_{h2} given by (5.1).

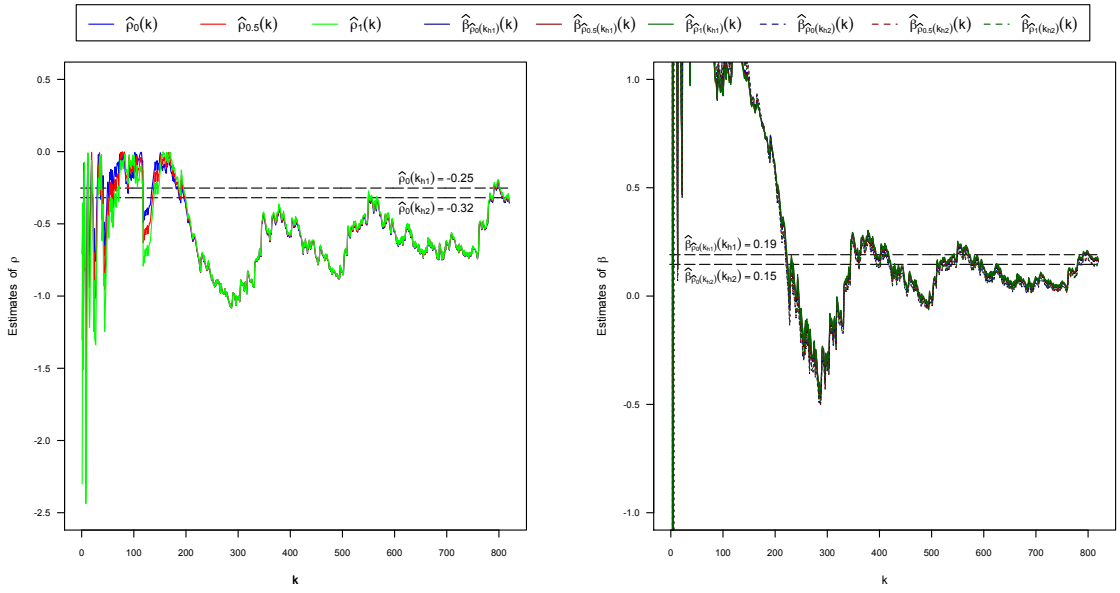


Figure 6.10: Estimates of the second order parameters ρ (left) and β (right) for seismicity of Philippines Islands.

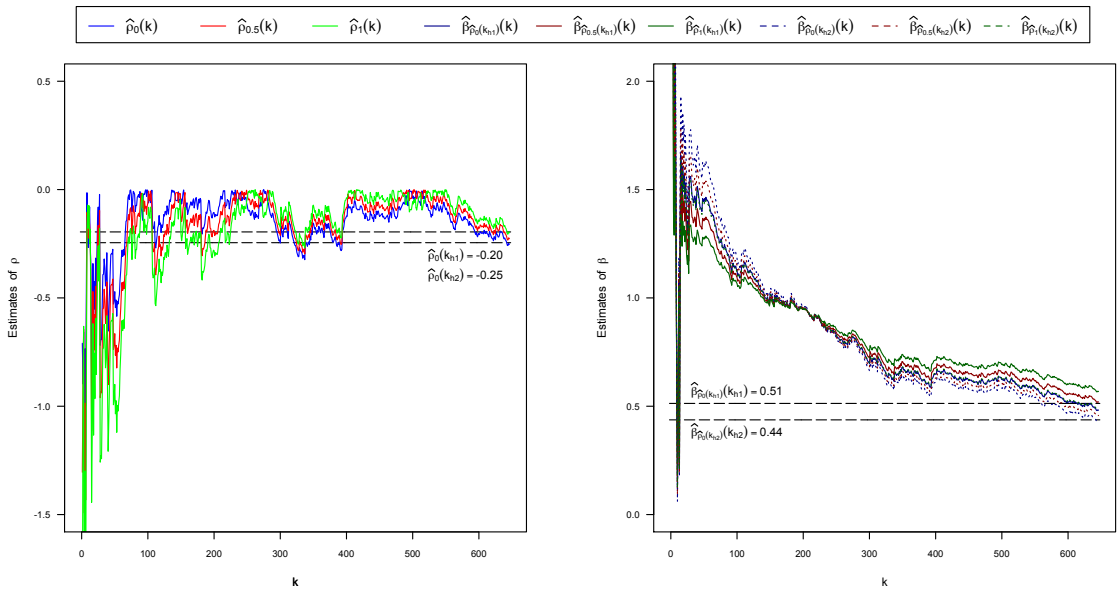


Figure 6.11: Estimates of the second order parameters ρ (left) and β (right) for seismicity of Vanuatu Islands.

The Figures 6.10 and 6.11 show the sample paths of the second order parameter esti-

mators, $\widehat{\rho}$ and $\widehat{\beta}$, based on the Philippines and Vanuatu seismic moment observations, respectively. We might see that the sample paths of $\widehat{\rho}$ for the three different values of τ have a very similar behaviour. It is however apparent that the behaviour of $\widehat{\rho}$ is slightly better when considering $\tau = 0$, specially for data concerning the Vanuatu Islands. As the above described algorithm also points to the choice of $\tau = 0$ in both cases, we chose this value of τ to estimate ρ.

Thus, for Philippines Islands, we have $k_{h1} = \lfloor 821^{0.995} \rfloor = 793$ and $k_{h2} = \lfloor 821^{0.999} \rfloor = 815$, that is, the corresponding estimates of ρ are $\widehat{\rho}_0(793) \approx -0.25$ and $\widehat{\rho}_0(815) \approx -0.32$ and the corresponding estimates of β are $\widehat{\beta}_{\widehat{\rho}_0(793)}(793) \approx 0.19$ and $\widehat{\beta}_{\widehat{\rho}_0(815)}(815) \approx 0.15$, being both represented graphically through straight lines. Doing the same procedure to Vanuatu Islands, we have $k_{h1} = \lfloor 647^{0.995} \rfloor = 626$ and $k_{h2} = \lfloor 647^{0.999} \rfloor = 642$, that is, the corresponding estimates of ρ are $\widehat{\rho}_0(626) \approx -0.20$ and $\widehat{\rho}_0(642) \approx -0.25$ and the corresponding estimates of β are $\widehat{\beta}_{\widehat{\rho}_0(626)}(626) \approx 0.51$ and $\widehat{\beta}_{\widehat{\rho}_0(642)}(642) \approx 0.44$.

Since from the $\widehat{\beta}$ sample paths it is not readily apparent significant differences between the use of k_{h1} or k_{h2} and due to the fact that the tail index estimation is more affected by the ρ fluctuations than the β ones, we use the both levels in the remaining study.

Moreover, here we also present a possible optimal level k_0 of top observations to consider when the geometric-type estimator is used to estimate γ, through the minimisation of the asymptotic mean square error ($AMSE$) of the geometric-type estimator. Considering again the representation of the geometric-type estimator in Theorem 3.2.4, we get what we need to calculate the $AMSE(\widehat{GT})$ and provide for their minimisation

$$\begin{aligned}\frac{\partial}{\partial k}\left[AMSE\left(\widehat{GT}\right)\right] = 0 &\Longleftrightarrow \frac{\partial}{\partial k}\left[V\left(\widehat{GT}\right) + \left(Bias\left(\widehat{GT}\right)\right)^2\right] = 0 \\ &\Longleftrightarrow \frac{\partial}{\partial k}\left[\frac{2\gamma^2}{k} + \left(\frac{\gamma\beta}{(1-\rho)^2}\right)^2 \left(\frac{n}{k}\right)^{2\rho}\right] = 0.\end{aligned}$$

Solving the equation in order to k and denoting the result as $k_0^{\widehat{GT}}$, we obtain

$$k_0^{\widehat{GT}} = \left[\frac{2(1-\rho)^4}{-2\rho\beta^2}\right]^{1/(1-2\rho)} n^{-2\rho/(1-2\rho)}.$$

Although this is not the optimal value for the bias corrected estimators, the value of the tail index and quantiles calculated with the geometric-type estimator at the $k_0^{\widehat{GT}}$ level is represented in some illustrations for comparison.

As a first step we estimate the tail index, γ, using GT estimator and Hill's estimator.

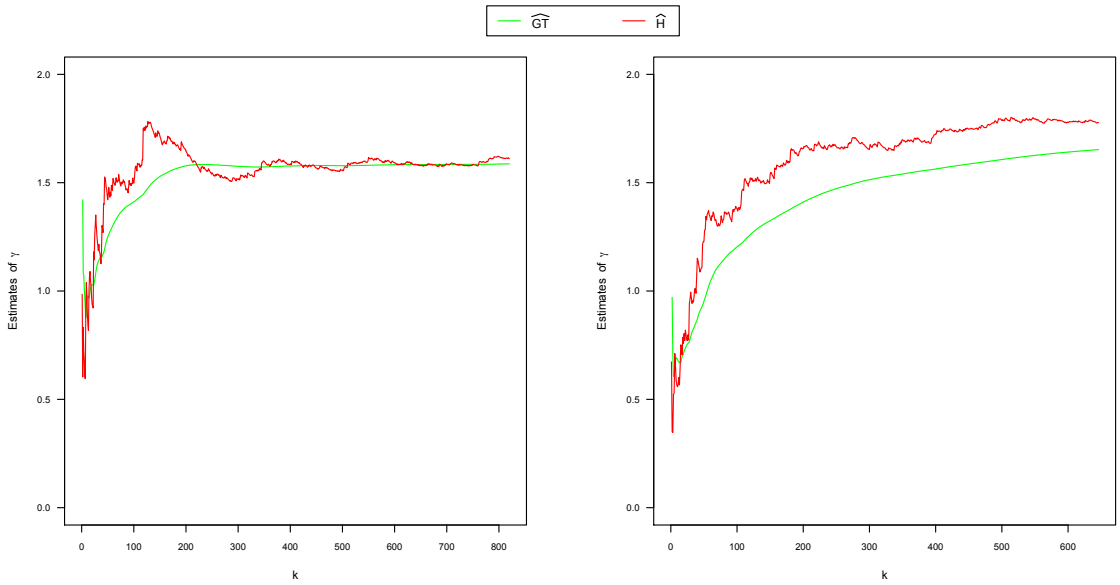


Figure 6.12: Plot for the GT estimator, $\widehat{GT}$, and for the Hill estimator, $\widehat{H}$, of γ, for seismicity of Philippines (left) and Vanuatu (right) Islands.

Concerning the shape parameter γ, the Figure 6.12 displays the estimated values of GT and Hill estimators, as a function of k, for Philippines and Vanuatu Islands data. As one can observe, for Philippines Islands data both estimators give similar results stabilising around the same value of γ, which is 1.6, with basically the same scatter for moderate and high values of k, although it is worth to give emphasis to the smoothness that the geometric-type estimator shows.

For the Vanuatu Islands data, though not so explicit as to the Philippines data, the behaviour of GT tends to stabilise around the value of 1.64 as k increases. The same is true for the Hill estimator around the value of 1.78, although in a slightly more erratic way.

The GT estimator presents the best performance specially for Philippines Islands data, displaying almost a straight line around 1.58 for k-values larger than 300.

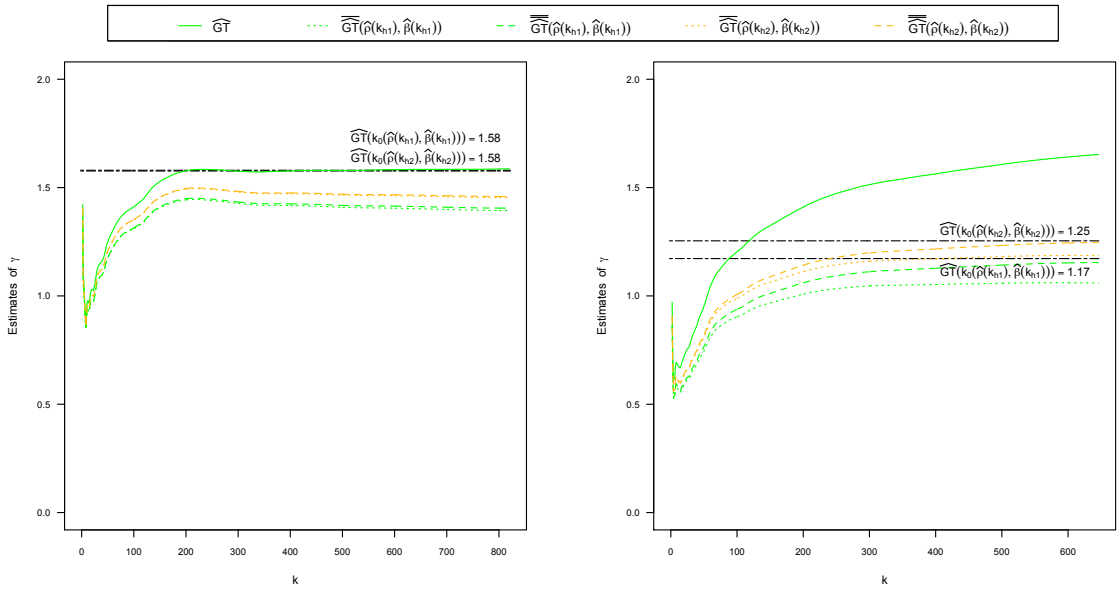


Figure 6.13: Plot for the GT estimator, $\widehat{GT}$, and for the corresponding GT bias corrected estimators, $\widehat{\overline{GT}}$ and $\widehat{\overline{\overline{GT}}}$, of γ, for seismicity of Philippines (left) and Vanuatu (right) Islands.

In Figure 6.13 it is possible to compare the behaviour of the GT estimator with its corrected versions, $\widehat{\overline{GT}}$ and $\widehat{\overline{\overline{GT}}}$. We note that the corrected estimators maintain the good behaviour, having less variation in the initial values of k, and stabilising at slightly lower values than the uncorrected estimator. Depending on the unknown value of the tail index parameter, that we seek, this type of behaviour seems to be indicative of a better performance of the corrected estimators. Particularly for Vanuatu Islands data, this improvement seems to be evident since the corrected estimators begin to stabilise sooner than the non corrected ones, showing a very satisfactory behaviour, right from the initial values of k.

In order to make the comparison between the bias corrected GT estimators and the Hill ones, we draw the sample paths of one against the other.

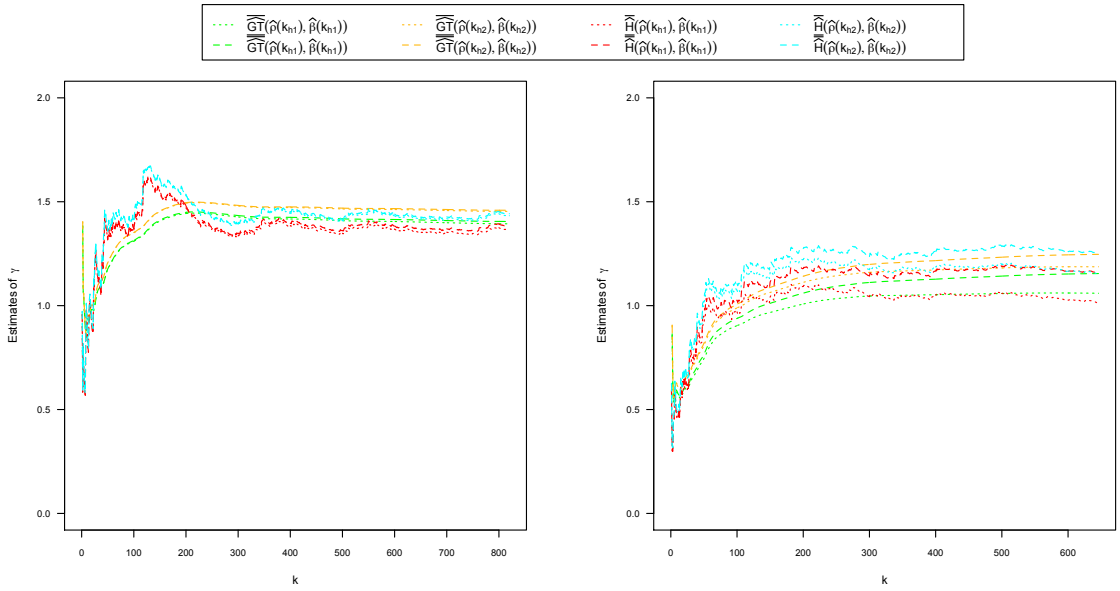


Figure 6.14: Plot for the GT bias corrected estimators, $\overline{\widehat{GT}}$ and $\overline{\overline{\widehat{GT}}}$, and for the Hill ones, $\overline{\widehat{H}}$ and $\overline{\overline{\widehat{H}}}$, of γ, for seismicity of Philippines (left) and Vanuatu (right) Islands.

We might see from Figure 6.14 that the estimates provided by the corrected Hill estimators are around the same values of the estimates given by the corrected GT estimators. However, it is quite clear that the Hill estimators hold a rather irregular behaviour compared to the GT estimators, specially for smaller values of k.

It is suggestive that the value of γ that best describes the seismic moment of the Philippines Islands is a little below 1.5 and of the Vanuatu Islands is slightly above 1.

As in most of the applications, the main interest lays not on the tail index but in the quantiles of the extreme distributions, which are more stable and robust. Now we analyse the sample paths of the quantiles estimators. We estimate the values of POT high quantiles estimator, in (2.2), based on the GT and Hill estimators, as a function of k, for Philippines and Vanuatu Islands data, considering the percentile 99%. Each tail index estimator leads to a different estimation of large quantiles, which is, also, dependent on k. The straight dashed line represents the estimate of the empirical

99% quantile. When more than one straight line are present, the empirical quantile is represented by the inferior one.

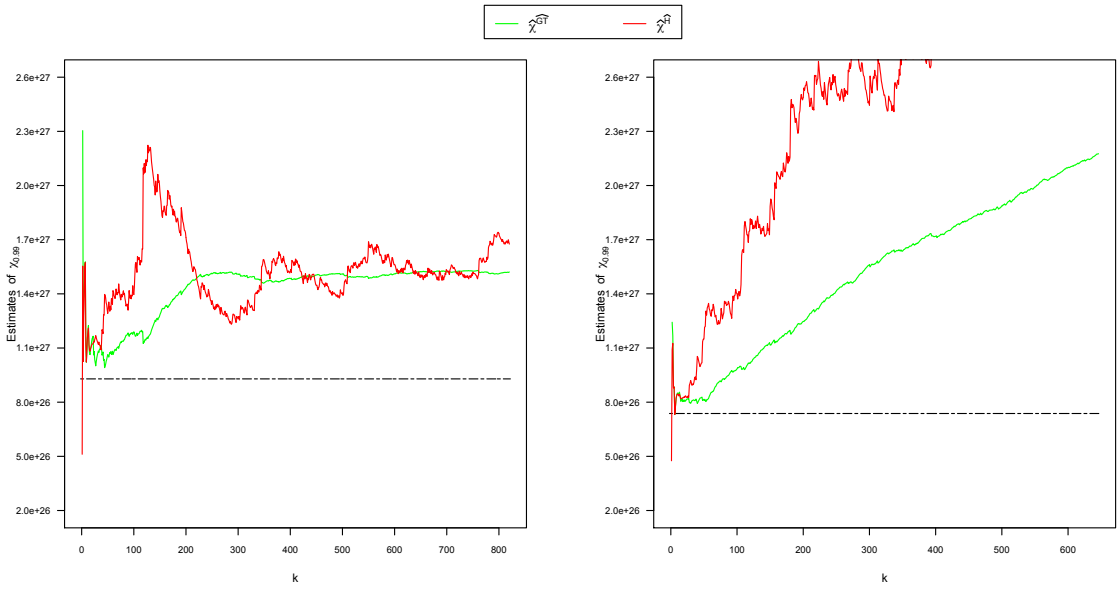


Figure 6.15: Plot for the 99-quantiles estimators based on the GT estimator, $\widehat{\chi}^{\widehat{GT}}$, and on the Hill estimator, $\widehat{\chi}^{\widehat{H}}$, of $\chi_{0.99}$, for seismicity of Philippines (left) and Vanuatu (right) Islands (empirical quantiles $\chi_{0.99} = 9.29 \times 10^{26}$ and $\chi_{0.99} = 7.37 \times 10^{26}$, for Philippines and Vanuatu Islands, respectively).

We might see from Figure 6.15 that, for the Philippines Islands, both estimates do not present values close to the empirical quantile. For values of k larger than 300, the estimates tend to stabilise, being apparent that this stabilisation process is significantly more regular for the GT based quantiles estimator. The uneven performance that the Hill quantile plot shows, make it extremely hard to decide upon a specific value for k. For the Vanuatu Islands the behaviour of both estimators is not the best, but the Hill based quantiles estimator presents a much more irregular behaviour.

Now comparing the GT based quantiles estimator with its corrected versions, we can observe in Figure 6.16 that the improvement caused by the correction is quite remarkable. It is also to be noted that considering the k_{h2} level to estimate the second order parameters, the performance seems to be a little better. Also in Figure 6.16, and for Philippines Islands data, it can be seen that the quantile value calculated using the

geometric-type estimator at its optimal levels $k_0^{\widehat{GT}}$, represented by the superior straight lines, almost coincides with the value of the quantiles estimator based on the geometric-type estimation for k-values larger than 200, which highlights the fairly stable behaviour of this quantiles estimator in this range of values.

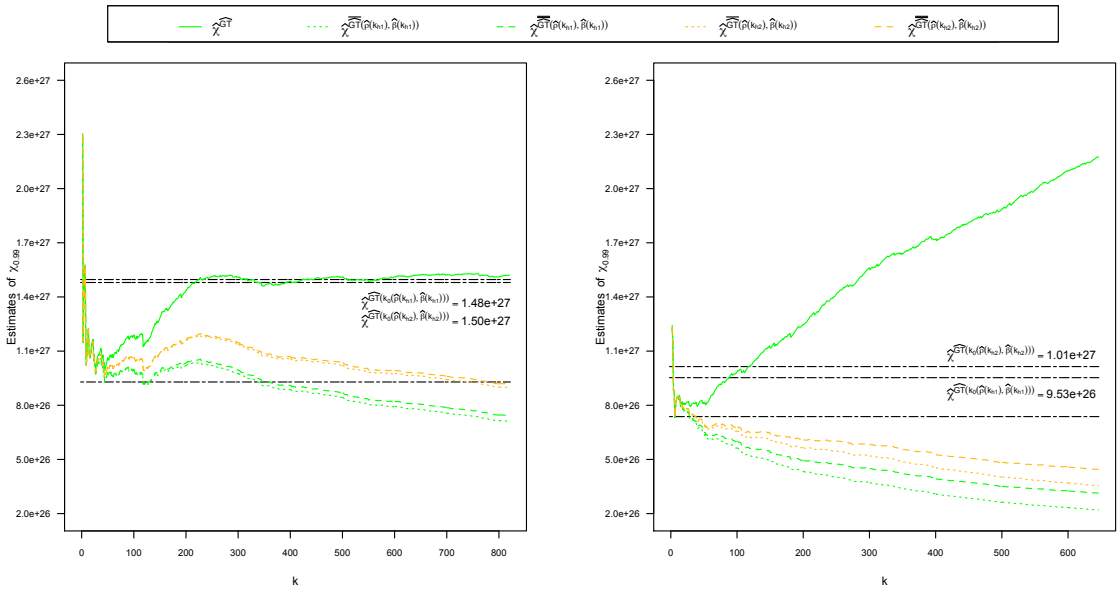


Figure 6.16: Plot for the 99-quantiles estimators based on the GT estimator, $\widehat{\chi}^{\widehat{GT}}$, and on the corresponding geometric-type bias corrected estimators, $\widehat{\chi}^{\overline{\widehat{GT}}}$ and $\widehat{\chi}^{\overline{\overline{\widehat{GT}}}}$, of $\chi_{0.99}$, for seismicity of Philippines (left) and Vanuatu (right) Islands (empirical quantiles $\chi_{0.99} = 9.29 \times 10^{26}$ and $\chi_{0.99} = 7.37 \times 10^{26}$, for Philippines and Vanuatu Islands, respectively).

In Figure 6.17 we can observe that the bias corrected Hill quantiles estimators present estimate values very similar to the ones presented by the bias corrected GT quantiles estimators. Although the corrected Hill quantiles estimators using the k_{h2} level to compute the second order parameters seem to have values more close to the empirical quantile than the corresponding corrected GT quantiles estimators, in case of Philippines Islands only for k-values greater that 300, their erratic and much less stable behaviour may be a factor of considerable disadvantage.

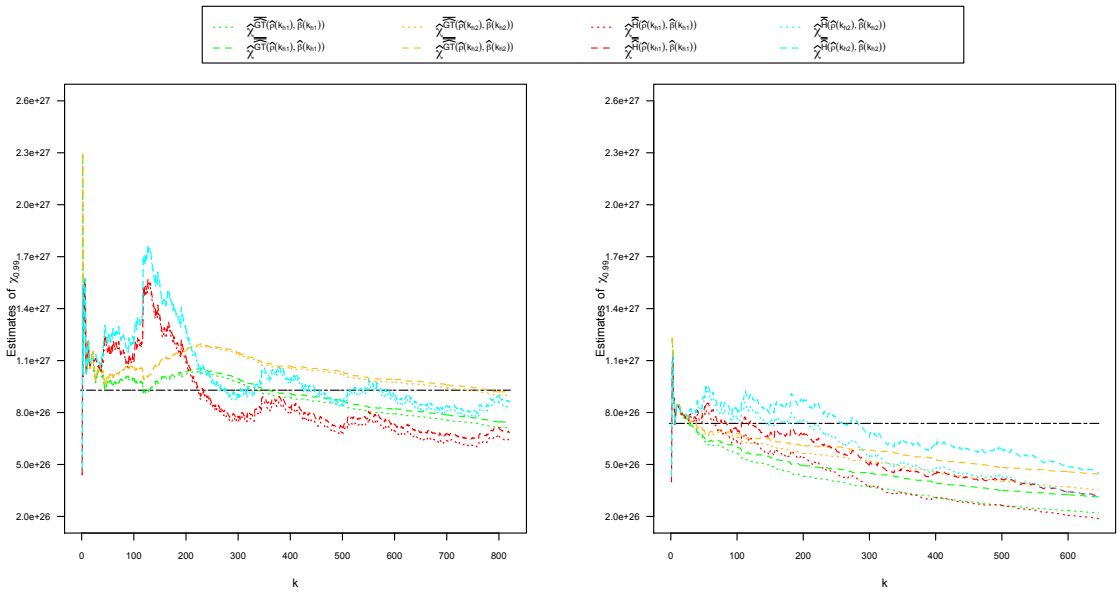


Figure 6.17: Plot for the 99-quantiles estimators based on the geometric-type bias corrected estimators, $\widehat{\chi}^{\overline{GT}}$ and $\widehat{\chi}^{\overline{\overline{GT}}}$, and on the Hill bias corrected estimators, $\widehat{\chi}^{\overline{H}}$ and $\widehat{\chi}^{\overline{\overline{H}}}$, of $\chi_{0.99}$, for seismicity of Philippines (left) and Vanuatu (right) Islands (empirical quantiles $\chi_{0.99} = 9.29 \times 10^{26}$ and $\chi_{0.99} = 7.37 \times 10^{26}$, for Philippines and Vanuatu Islands, respectively).

In this real case study the quantiles estimator using the geometric-type estimator shows a better performance. These results are improved when we apply the geometric-type bias corrected estimators to the quantiles estimator.

In general it is possible to conclude that the smoother behaviour is a common quality both for the estimates obtained for the GT tail index estimators as for GT based quantiles estimators, which show a very small variability, reflecting the more regular behaviour of the GT estimators. Although the Hill estimator is generally more unstable, it also displays an adequate behaviour.

As one knows, the performance of the estimators depends on the distribution of the data and there is not a uniformly best estimator. Nevertheless, from the results of the practical example conducted in this section, one could deduce that for this type of data the GT estimator turns out to be the best choice for tail index estimator and when used in the POT estimator for high quantiles.

Bibliography

Balkema, A.A. and de Haan, L. (1974). Residual life time at great age. *Annals of Probability* 2, 792–804.

Beirlant, J., Caeiro, F. and Gomes, M.I. (2012). An overview and open research topics in statistics of univariate extremes. *Revstat* 10, 1–31.

Beirlant, J., Dierckx, G., Goegebeur Y. and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes* 2, 177–200.

Beirlant, J., Dierckx, G. and Guillou, A. (2005). Estimation of the extreme-value index and generalised quantile plots. *Bernoulli* 11, 949–970.

Beirlant, J., Figueiredo, F., Gomes, M.I. and Vandewalle, B. (2008). Improved reduced-bias tail index and quantiles estimators. *J. Statist. Plann. and Inference*, 138, 1851–1870.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. John Wiley and Sons, Ltd., England.

Ben-Menahem, A. (1995). A concise history of mainstream seismology - Origins, legacy, and perspectives. *Bull. Seism. Soc. Am.*, 85, 1202–1225.

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation*. Cambridge University Press.

Brito, M. and Freitas, A.C.M. (2003). Limiting behaviour of a geometric estimator for tail indices. *Insurance: Mathematics and Economics* 33, 221–226.

Brito, M. and Freitas, A.C.M. (2006). Weak convergence of a bootstrap geometric-type estimator with applications to risk theory. *Insurance: Mathematics and Economics* 38, 571–584.

Caeiro, F. and Gomes, M. I. (2008). Minimum-variance reduced-bias tail index and high quantiles estimation. *Revstat* 6, 1–20.

Caeiro, F., Gomes, M.I. and Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat* 3, 113–136.

Casella G. and Berger R.L. (1990). *Statistical Inference*. Duxbury Press.

Castillo, E. and Hadi, A. (1997). Fitting the generalised Pareto distribution to data. *J. Amer. Statist. Assoc.* 92, 1609–1620.

Chen, W. and Scawthorn, C. (Editors) (2003). *Earthquake engineering handbook*. CRC Press LLC.

Coles, S. (2001). *An Introduction to Statistical Modelling of Extreme Values*. Springer-Verlag, London, Berlin, Heidelberg.

Coles, S. and Dixon, M.J. (1999). Likelihood-based inference for extreme value models. *Extremes* 2, 5–23.

Coles, S. and Powell, E. (1996). Bayesian methods in extreme value modelling: a review and new developments. *Int. Statist. Rev.* 64, 119–136.

Corral, A. (2006). Dependence of earthquake recurrence times and independence of magnitudes on seismicity history. *Tectonophysics* 424, 177–193.

Csörgő, S., Deheuvels, P. and Mason, D.M. (1985). Kernel estimates of the tail index of a distribution. *Ann. Statist.* 13, 1050–1077.

Csörgő, S. and Mason, D.M. (1985). Central limit theorems for sums of extreme events. *Math. Proc. Camb. Phil. Soc.* 98, 547–558.

Csörgő, S. and Viharos, L. (1995). On the asymptotic normality of Hill's estimator. *Math. Proc. Cambridge Philos. Soc.* 118, 375–382.

Csörgő, S. and Viharos, L. (1998). Estimating the tail index. In: *Asymptotic Methods in Probability and Statistics*, (B. Szyszkowicz, ed), North Holland, Amesterdam, 833–881.

Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds (with discussion). *J. R. Statist. Soc.* 52, 237–254.

Day, R.W. (2002). *Geotechnical Earthquake Engineering*. McGraw-Hill Professional, New York.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer, Boston.

de Haan, L. and Rootzén, H. (1993). On the estimation of high quantiles. *J. Statist. Plann. Inference* 35, 1–13.

Deheuvels, P., Haeusler, E. and Mason, D.M. (1988). Almost sure convergence of the Hill estimator. *Math. Proc. Camb. Phil. Soc.* 104, 371–381.

Dekkers, A.L.M. and de Haan, L. (1989). On the estimation of the extreme-value index and large quantiles estimation. *Ann. Stat.* 17, 1795–1832.

Dekkers, A.L.M. and de Haan, L. (1993). Optimal choice of sample fraction in extreme-value estimation. *J. Multivar. Anal.* 47, 173–195.

Dekkers, A.L.M., Einmahl, J.H.J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* 17, 1833–1855.

Diebolt, J., Guillou, A., Naveau, P. and Ribereau, P. (2008). Improving probability-weighted moment methods for the generalized extreme value distribution. *REVSTAT* 6, 33–50.

Dietrich, D., de Haan, L. and Hüsler, J. (2002). Testing extreme value conditions. *Extremes* 5, 71–85.

Drees, H. (1996). Refined Pickands estimators with bias correction. *Comm. Statist. Theory and Meth.* 25, 837–851.

Dziewonski, A.M., Chou, T.-A. and Woodhouse, J.H. (1981). Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *J. Geophys. Res.* 86, 2825–2852.

Ekström, G., Nettles, M. and Dziewonski, A.M. (2012). The global CMT project 2004-2010: Centroid-moment tensors for 13,017 earthquakes. *Phys. Earth Planet. Inter.* 200–201, 1-9.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.

Falk, M., Hüsler, J. and Reiss, R.-D. (2010). *Laws of Small Numbers: Extremes and Rare Events*. 3rd edition, Springer-Basel.

Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.* 27, 760–781.

Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.* 24, 180–190.

Fraga Alves, M.I., Gomes, M.I. and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Math.* 60, 193–213.

Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Math. Polon.* 6, 93–116.

Geluk, J. and de Haan, L. (1987). *Regular Variation, Extensions and Tauberian Theorems*. CWI Tract 40, Center of Mathematics and Computer Science, Amsterdam, Netherlands.

Global CMT Catalogue. Available from <http://www.globalcmt.org/>. [last accessed August 2013].

Gnedenko, B.V. (1943). Sur la distribuition limite du terme maximum d'une série aléatoire. *Ann. Math.* 44, 423–453.

Gomes, M. I., Caeiro, F. and Figueiredo, F. (2004). Bias reduction of a tail index estimator trough an external estimation of the second order parameter. *Statistics* 38, 497–510.

Gomes, M.I. and Martins, M.J. (2002). "Asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes* 5, 5–31.

Gomes, M.I., Martins, M.J. and Neves, M. (2000). Alternatives to a semi-parametric estimator of parameters of rare events – the jackknife methodology. *Extremes* 3, 207–229.

Gomes, M.I, Martins, M.J. and Neves, M. (2007). Improving second order reduced bias extreme value index estimation. *Revstat* 5, 177–207.

Gomes, M.I. and Pestana, D. (2007). A simple second order reduced bias tail index estimator. *J. Statist. Comput. and Simulation* 77, 487–504.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form, *Wat. Resour. Res.* 15, 1049–1054.

Gumbel, E.J. (1941). The return period of flood flows. *Annals of Mathematical Statistics* 12, 163–190.

Gutenberg, B. and Richter, C.F. (1956). Magnitude and energy of earthquakes. *Annali di geofisica* 9, 1–15.

Haeusler, E. and Teugels, J.L. (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation. *Ann. Statist.* 13, 743–756.

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. Royal Statist. Soc.* B44, 37–42.

Hill, B.M. (1975). A simple approach to inference about the tail of a distribution. *Ann. Stat.* 3, 1163–1174.

Hosking, J. and Wallis, J. (1987). Parameter and quantiles estimation for the Generalised Pareto Distribution. *Technometrics* 29, 339–349.

Hosking, J.R.M., Wallis, J.R. and Wood, E.F. (1985). Estimation of the generalised extreme-value distribution by the method of probability-weighted moments. *Technometrics* 27, 251–261.

Howell, B.F., Jr (1990). *An Introduction to Seismological Research*. Cambridge University Press.

Hüsler, J. and Li, D. (2006). On testing extreme value conditions. *Extremes* 9, 69–86.

Jenkinson, A.F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quart. J. Roy. Meteo. Soc.* 81, 158–171.

Kanamori, H. (1974). Earthquake Prediction. *Engineering and Science* 38, 18–21.

Katz, R., Parlange, M. and Naveau, P. (2002). Statistics of extremes in hydrology. *Adv. Water Resour.* 25, 1287–1304.

Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions – Theory and Applications*. World Scientific Publishing.

Kratz, M. and Resnick, S. (1996). The qq-estimator of the index of regular variation. *Commun. Statist. Stochastic Models* 12, 699–724.

Landwehr, J. M., Matalas, N. C. and Wallis, J. R. (1979). Probability weighted moments com- pared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resour. Res.* 15, 1055–1064.

Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer Series in Statistics. Springer-Verlag, New York.

Martins, E. and Stedinger, J. (2000). Generalised maximum-likelihood generalised extreme-value quantiles estimators for hydrologic data, *Water Res. Research* 36, 737–744.

Mason, D.M. (1982). Laws of large numbers for sums of extreme values. *Ann. Prob.* 10, 754–764.

Matthys, G. and Beirlant, J. (2003). Estimating the extreme value index and high quantiles with exponential regression models. *Statistica Sinica* 13, 853–880.

McConnell, D.A., Steer, D.N., Owens, K., Knight, C., and Park, L.E. (2007). *The Good Earth: An Introduction to Earth Science*, McGraw-Hill Education, New York.

McNeil, A.J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management – Concepts, Techniques and Tools*. Princeton University Press.

McNeil, A.J. and Saladin, T. (1997). The peaks over thresholds method for estimating high quantiles of loss distributions. *Proceedings of 28th International ASTIN Colloquium*, 23–43.

Peng, L. (1998). Asymptotically unbiased estimator for the extreme-value-index. *Statistics and Probability Letters* 38, 107–115.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Stat.* 3, 119–131.

Pisarenko, V.F. and Sornette, D. (2003). Characterization of the frequency of extreme events by Generalised Pareto Distribution. *Pure and Applied Geophysics* 160, 2343–2364.

R Development Core Team (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reiss, R.-D. and Thomas, M. (2007). *Statistical Analysis of Extreme Values, with Application to Insurance, Finance, Hydrology and Other Fields*. 3rd edition, Birkhäuser Verlag.

Schultze, J. and Steinebach, J. (1996). On least squares estimates of an exponential tail coefficient. *Statist. Decisions* 14, 353–372.

Shearer, P.M. (2009). *Introduction to seismology*. 2nd edition, Cambridge University Press, UK.

Smirnov, N.V. (1967). Some remarks on limit laws for order statistics. *Theory Probab. Appl.* 12, 337–339.

Tiago de Oliveira, J. (1990). Perspectiva da Estatística de Extremos: Resultados Básicos e Alguns Problemas em Aberto. *Revista Real Academia de Ciências Exactas, Físicas y Naturales*, Madrid, t. XXXIV, cuad. III, 453–481.

von Mises, R. (1936). La distribuition de la plus grande de n valeurs. *Rev. Math. Union Interbalcanique* 1, 141–160.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist.* 73, 812–815.