A review of more than one hundred Pareto-tail index estimators

Fedotenkov, Igor

Russian Presidential Academy of National Economy and Public Administration

15 November 2018

A review of more than one hundred Pareto-tail index estimators

Igor Fedotenkov[1]

Abstract

This paper reviews more than one hundred Pareto (and equivalent) tail index estimators. It focuses on univariate estimators for nontruncated data. We discuss basic ideas of these estimators and provide their analytical expressions. As samples from heavy-tailed distributions are analysed by researchers from various fields of science, the paper provides nontechnical explanations of the methods, which could be understood by researchers with intermediate skills in statistics. We also discuss strengths and weaknesses of the estimators, if they are known. The paper can be viewed as a catalog or a reference book on Pareto-tail index estimators.

JEL Classification: C13, C14.
AMS Subject Classification: 62-02, 62F99, 62G05.

Keywords: Heavy tails, Pareto distribution, tail index, review.

[1]Russian Presidential Academy of National Economy and Public Administration

E-mail: i.fedotenkov@gmail.com

1 Introduction

Since the seminal work of Pareto (1897), the classical application of heavy-tailed distributions in economics is allocations of income and wealth (Gastwirth 1972; Cowell and Flachaire 2007; Ogwang 2011; Benhabib, Bisin, and Zhu 2011; Toda 2012; Benhabib, Bisin, and Luo 2017). Right-hand heavy-tails are also common when analysing consumption data (Nordhaus 2012; Toda and Walsh 2015), price of land (Kaizoji 2003), CEO compensations (Gabaix and Landier 2008), firm sizes (Simon and Bonini 1958; Axtell 2001) and productivities (Chaney 2008). In international economics, revealed comparative advantage and changes in exchange rates also follow a Pareto distribution (Hinloopen and Van Marrewijk 2012). In finance, fluctuations of stock and commodity prices follow power laws as well (Mandelbrot 1963; Gabaix, Gopikrishnan, Plerou, and Stanley 2003; Zhong and Zhao 2012; Das and Halder 2016). Most applications of heavy tails in economics and finance were discussed in detail by Gabaix (2009). Apart from applications in economics, heavy tailed distributions are used to describe the upper tails of the sizes of cities (Rosen and Resnick 1980; Soo 2005), lakes (Seekell and Pace 2011) and sets of mineral deposits (Agterberg 1995). They are also common in biology (Ferriere and Cazelles 1999; Seekell and Pace 2011), telecommunications (Huebner, Liu, and Fernandez 1998), seismology (Pisarenko and Sornette 2003; Nordhaus 2012) and many other fields (Newman 2005).

Most of statistical and econometric methods are based on laws of large numbers and central limit theorems. The Kolmogorov's law of large numbers requires an existence of the first finite moment. Lyapunov's version of the central limit theorem assumes an existence of the finite moment of an order higher than two. If data comes from a heavy-tailed distribution, these assumptions may not necessarily be satisfied. The existence of specific finite moments is closely linked to the concept of a tail index, and estimation of the tail index is one of key problems in statistics. Many methods were proposed and numerous modifications of existent methods were made. The goal of this paper is to review existing tail-index estimators.

There are numerous tail-index estimators. They are based on various assumptions, have diverse asymptotic and finite-sample properties. Unfortunately, the literature in this field is unstructured. Data analysts and policy-makers have a tough time choosing the best techniques for their particular cases. Even statisticians, who work with heavy-tailed distributions, often face difficulties searching the literature and they are not always aware of many developments in this field. As a consequence, a few estimators were derived independently by different authors. The goal of our paper is to fill this gap. The paper reviews more than one hundred tail-index estimators, discusses their assumptions and provides closed-form expressions. We also correct a number of typos present in the original works and these corrections were okayed by the authors.[2] The paper also aims to provide nontechnical explanations of methods, which could be understood by researchers with intermediate skills in statistics. The paper

[2]The authors who replied to our queries are listed in the acknowledgments section.

can be considered as a reference book on tail index estimators for researchers from various fields of science.

An excellent review of the advantages in the extreme value theory was made by Gomes and Guillou (2015). They discussed a large number of underlying theories and reviewed a number of tail-index estimators, providing equations for thirteen of them. We cover a much wider range of estimators, and provide analytical expressions for more than one hundred various tail index estimators. We avoid repeating information provided by Gomes and Guillou; however, the most famous estimators were also briefly discussed in order to have a possibility to compare them with other methods.

A number of works are devoted to comparisons of various estimators. For example, De Haan and Peng (1998) compared asymptotic minimal mean squared errors of four tail-index estimators and Caeiro and Gomes compared asymptotic properties of several reduced-bias tail-index estimators. A few works performed Monte-Carlo simulations in order to compare finite-sample properties of the estimators (Gomes and Oliveira 2003; Brzezinski 2016; Paulauskas and Vaičiulis 2017b; Kang and Song 2017). However, typically the number of compared estimators is not large (lower than ten). Our work facilitates more complete Monte-Carlo comparisons of the existing methods. Computer codes written in R for most of the estimators, reviewed in this paper, are readily available on the author's webpage.

The focus of this review is univariate Pareto-type tail index estimators for i.i.d. nontruncated data. We focus on the right tails only. To the best of our knowledge, this is the most complete review of tail-index estimators.

1.1 Notation

Before we proceed, it is useful to introduce notation, which is used throughout the paper.

$$U_j = j(\log X_{(n-j+1)} - \log X_{(n-j)}). \tag{1}$$

$$V_{i,k} := X_{(n-i+1)}/X_{(n-k)}. \tag{2}$$

$$M_{k,n}^{(l)} = \frac{1}{k}\sum_{i=0}^{k-1}\big(\log X_{(n-i)} - \log X_{(n-k)}\big)^l. \tag{3}$$

1.2 Assumptions

A common assumption is that observations are i.i.d. However, a number of estimators, reviewed in this paper, also allow for dependent observations. A large number of estimators are based on an assumption that the right tail of a distribution has a Pareto form.

- A1: $1 - F(x) \sim L(x)x^{-\alpha}$ as $x \to \infty$, where α is a tail index. It is assumed that $\alpha > 0$. $L(x)$ is a slowly varying function, i.e. $\lim_{t\to\infty} L(t)/L(tx) \to 1$ as $t \to \infty$, for every $x > 0$.

It is also convenient to denote an extreme value index (EVI) $\gamma = \alpha^{-1}$. Very often estimators are designed for γ (instead of α) estimation. Sometimes γ is also allowed to take on negative values. Note that even if $\hat{\gamma}$ is an unbiased estimator of γ, $\hat{\gamma}^{-1}$ is a biased estimator of α.

A more general functional form of the tail was introduced by Hall (1982).

- A2: $1 - F(x) = 1 - F(x) = Cx^{-\alpha}[1 + D_1 x^{-\alpha} + ... + D_m x^{-m\alpha} + o(x^{-m\alpha})]$ as $x \to \infty$.

Another functional form with second-order parameters $D > 0$, $\rho < 0$ was studied in detail by Hall and Welsh (1985).

- A3: $1 - F(x) = Cx^{-\alpha}[1 + Dx^{\rho} + o(x^{\rho})]$ as $x \to \infty$.

A more general assumption of a functional form, which allows for a second-order parameter ρ is

- A4: $\lim_{t\to\infty} \frac{\log V(tx) - \log V(t) - \gamma \log x}{A(t)} = \begin{cases} \frac{x^{\rho}-1}{\rho} & \text{if } \rho < 0 \\ \log x & \text{if } \rho = 0 \end{cases}$,

where $A(\cdot)$ is a suitably chosen function of constant sign near infinity. Estimators, which take the second-order parameters into account often perform badly with $\rho = 0$; therefore, this case in assumption A4 is often omitted.

Sometimes the second-order parameter ρ is equalised to -1, resulting in the following simplified version of assumption $A3$:

- A5: $1 - F(x) = c_1 x^{-\alpha} + c_2 x^{-\alpha-1} + o(x^{-\alpha-1})$ as $x \to \infty$.

Assumptions A1-A5 are based on Pareto distributions, and they allow for positive α (and γ) only. A more general case, which allows for $\gamma \in \mathbb{R}$ results from assumptions based on the generalised Pareto distribution (GPD), or Extreme value distribution. Negative values of γ correspond to distributions with a finite right endpoint. The GPD is defined as follows:

- A6: $1 - F(x) = \begin{cases} \left(1 + \frac{\gamma(x-\mu)}{\sigma}\right)^{-1/\gamma} & \text{if } \gamma \neq 0,\ x \geq \mu \\ e^{-\frac{x-\mu}{\sigma}} & \text{if } \gamma = 0,\ \mu \leq x \leq \mu - \sigma/\gamma \end{cases}$,

where σ, $\sigma > 0$ is a scale parameter, and μ, $\mu \in (-\infty, \infty)$ is a location parameter. There are many parametric methods, which assume that an entire sample is drawn from a GPD. But, GPD can also be applied for a tail only. It is often assumed that the difference between values exceeding a certain (high) threshold and the threshold itself has a GPD distribution.

- A7: The sample is drawn from a distribution which belongs to the maximum domain of attraction of the generalised extreme value distribution. I.e. for some sequences $a_n > 0$, b_n, $n = 1, 2, ...$ and $\gamma \in \mathbb{R}$, $\lim_{n\to\infty} P\left\{\frac{max(X_1, X_2, ..., X_n) - b_n}{a_n} \leq x\right\} = \exp\left(-(1+\gamma x)^{-1/\gamma}\right)$, $1+\gamma x > 0$.

Another popular assumption is

- A8: The sample is drawn from a distribution, which belongs to the domain of attraction of a stable distribution with $0 < \alpha < 2$.

The stability parameter α in assumption A8 corresponds to the tail index if $0 < \alpha < 2$. The weakness of this assumption is that it imposes strict limitations on α, but it can be extended, by raising the data to a power.

Despite assumptions A1-A8 sound differently, the parameter α (γ) always determines the heaviness of the tail. Many classical estimators are based on assumption A1. If asymptotic properties of an estimator are derived using an alternative assumption, we specify this fact in the text. More details on theories underlying these assumptions are provided by Gomes and Guillou (2015).

Most of the estimators are based on the k largest observations, with the following assumptions about k:

- B1: $k(n) \to \infty$ as $n \to \infty$.
- B2: $k = k(n) = o(n)$ as $n \to \infty$.

If the methods use different assumptions, we specify them in the text. Otherwise, we suppose that B1-B2 hold. It is also necessary to mention that nowadays attempts to get rid of these assumptions are being made (Müller and Wang 2017).

2 Hill estimator and other estimators for $\gamma > 0$.

2.1 Hill estimator

The most popular estimator of tail indexes is the Hill estimator (Hill 1975). Hill suggested to approximate the k-th largest observations with a Pareto distribution as in assumption A1 and used a maximum likelihood estimator (MLE) for γ estimation. Suppose that $X_{(1)}, ..., X_{(n)}$ are the order statistics of the sample. Then

$$\hat{\gamma}_n^H(k) = \frac{1}{k} \sum_{i=0}^{k-1} \log\left(\frac{X_{(n-i)}}{X_{(n-k)}}\right). \quad (4)$$

Despite Hill estimator is MLE, it is classified as a semi-parametric method because the Pareto distribution is only assumed about the limiting behavior of the tail. If k is chosen to be too high, the variance of the estimator increases. If k is too low, usually the bias of the estimator increases. The problem of the choice of k is also relevant to many other estimators.

Mason (1982) showed week consistency of the Hill estimator under assumptions A1, B1 and B2. For strong consistency an additional technical assumptions is also required (Deheuvels et al. 1998). Its asymptotic normality was analysed in detail by Hall (1982), Haeusler and Teugels (1985), Csörgoő and Mason (1985), Beirlant and Teugels (1989), Haan and Resnick (De Haan and Resnick 1998) and the others.

2.2 Kernel estimator

Csörgő et al. (1985) extended the Hill estimator with kernels. Its form is the following:

$$\hat{\gamma}_n^K(k) = \frac{\sum_{j=1}^{n-1} \frac{j}{k} K\left(\frac{j}{k}\right)\left(\log X_{(n-j+1)} - \log X_{(n-j)}\right)}{\frac{1}{k}\sum_{j=1}^{n-1} K\left(\frac{j}{k}\right)}, \tag{5}$$

$K(\cdot)$ is a kernel function with the following properties: 1) $K(u) > 0$, $0 < u < \infty$, 2) it is nonincreasing and right continuous on $(0, \infty)$, 3) $\int_0^\infty K(u)du = 1$; 4) $\int_0^\infty u^{-1/2}K(u)du < \infty$. The Hill estimator is obtained when $K(u) = \mathbb{1}(0 < u < 1)$. If $K(u) = u^{-1}\mathbb{1}(0 < u < 1)$ the kernel estimator reduces to the De Haan and Resnick estimator discussed below (eq. 92). ($\mathbb{1}$ is a unit indicator function).

2.3 Fraga Alves estimator

Fraga Alves (1995) developed a simple estimator for positive γ. It uses fewer observations than the Hill estimator and; therefore, its computation is faster.

$$\hat{\gamma} := \frac{1}{\log c} \log \frac{X_{(n-k+1)}}{X_{(n-ck+1)}}, \tag{6}$$

were c, $c > 1$ is an integer and $ck < n+1$. The estimator is based on assumption A7.

2.4 Aban& Meerschaert shifted Hill's estimator

As Hill estimator is not shift-invariant, Aban and Meerschaert (2001) changed assumption A1 to $1 - F(x) \sim C(x-s)^{-\alpha}$, where s is a shift parameter, and C is a constant $C > 0$. The MLE is

$$\hat{\alpha} = \left[\frac{1}{k}\sum_{i=0}^{k-1} \log\left(\frac{X_{(n-i)} - \hat{s}}{X_{(n-k)} - \hat{s}}\right)\right]^{-1}. \tag{7}$$

and the optimal shift satisfies

$$\hat{\alpha}(X_{(n-k)} - \hat{s})^{-1} = \left(\frac{\hat{\alpha}+1}{k}\right)\sum_{i=0}^{k-1}(X_{(n-i)} - \hat{s})^{-1}.$$

The optimal α and s can be solved from these two equations.

2.5 Danielsson, Jansen, De Vries estimator

Danielsson et al. (1996) suggested the following moment ratio estimator based on the assumption A3:

$$\hat{\gamma}^l_{MR} = \frac{M^{(l+1)}_{k,n}}{(l+1)M^{(l)}_{k,n}}, \tag{8}$$

where $M^{(l)}_{k,n}$ is defined in equation (3). l is a tuning parameter. Often l is equalised to 1.

2.6 Jurečková and Picek estimator

Jurečková (2000) and Jurečková and Picek (2001) suggested a semiparametric test for testing a hypothesis that the right tail of a distribution is heavier than that of a Pareto distribution with the tail index α_0. Jurečková and Picek (2004) reversed the question, and suggested to use the underlying idea of these tests for tail-index estimation. Split the sample into T nonintersecting subsamples of size m and denote the maximal element of sabsample j as $X^j_{(m)}$. Denote $\hat{F}^*(a) = T^{-1}\sum_{j=1}^{T} \mathbb{1}[X^j_{(m)} \leq a]$ - the empirical distribution of the subsamples' maxima, and $a_{T,s} = (mT^{1-\delta})^{\frac{1}{s}}$, $0 < \delta < 1/2$. The estimator is defined as

$$\hat{\alpha}_T = \frac{1}{2}(\hat{\alpha}^+_T + \hat{\alpha}^-_T), \tag{9}$$

where

$$\hat{\alpha}^+_T = sup\{s : \big(1 - F^*(a_{T,s})\big) < T^{-(1-\delta)}\},$$
$$\hat{\alpha}^-_T = inf\{s : \big(1 - F^*(a_{T,s})\big) > T^{-(1-\delta)}\}.$$

2.7 Davydov, Paulauskas, Račkauskas (DPR) estimator

LePage et al. (1981) studied asymptotic properties of order statistics in the case of the domain of attraction of a non-Gausian stable law. Based on their findings, Davydov, Paulauskas and Račkauskas (2000) proposed a (DPR) estimator, which was further studied by Paulauskas (2003). Its idea is the following: Divide the sample into T groups $V_1, ..., V_T$ of size m. It is assumed that $m \to \infty, T \to \infty$ as $n \to \infty$ unless the sample is not entirely from the Pareto distribution. In the former case it is better to take $m = 2$. Denote the largest observation in group V_j as $M^{(1)}_j$ and $M^{(2)}_j$ - the second largest observation in the same group. Also denote

$$Z_T = \frac{1}{T}\sum_{j=1}^{T}\frac{M^{(2)}_j}{M^{(1)}_j}.$$

They showed that

$$\hat{\alpha} = \frac{Z_T}{1 - Z_T} \tag{10}$$

is an asymptotically unbiased tail-index estimator. The proper choice of m is rather tricky, but it is possible to make a plot $(m, \hat{\alpha}_m)$, $2 \leq m \leq n/2$, similar to Hill plots (Paulauskas 2003). Also, to improve the finite sample properties of this estimator it is wise to apply it for different permutations of the sample, and take an arithmetic mean or median of the estimates.

2.8 Qi estimator

Qi (2010) used the idea of the DPR estimator to modify the Hill estimator. Observations are divided into T groups, of size m_i, $i = 1, ..., T$. Then, Hill estimator for every group is applied, and the arithmetic mean of them is taken:

$$\hat{\gamma} = \frac{1}{\sum_{i=1}^{T} k_i} \sum_{i=1}^{T} \sum_{j=1}^{k_i} (\log X_{(m_i-j+1)}^{(i)} - \log X_{(m_i-k_i)}^{(i)}), \tag{11}$$

where $X_{(m_i-j+1)}^{(i)}$ denotes the $m_i - j + 1$-th order statistics in group i, and k_i is the group-specific number of observations treated as the tail.

2.9 Fialova et al. estimator

Fialova et al. (2004) suggested an estimator, which takes all observations into account; however, it requires some preliminary knowledge about the tail index and uses an additional parameter. The sample size n is randomly partitioned into N non-overlapping sub-samples of size r: $(X_1^{(1)}, .., X_r^{(1)})$, ..., $(X_1^{(N)}, .., X_r^{(N)})$. Denote $(\bar{X}^{(1)}, ..., \bar{X}^{(N)})$ as a vector of arithmetic means of the sub-samples. Denote $\hat{F}_{\bar{X}_n}^{(N)}(x) = N^{-1} \sum_{i=1}^{N} \mathbb{1}(\bar{X}^{(i)} \leq x)$ the distribution function of the sub-sample means. Also, suppose that there is a preliminary information that tail index α is lower than a certain value α_0. I.e. the tail is heavier than that of a Pareto distribution with the tail index α_0. Then, choose $x_N = N^{\frac{1-\delta}{\alpha_0}}$, for a fixed δ, $0 < \delta < 1$ and calculate

$$\tilde{\alpha}(x_N) = -\frac{\log\left(1 - \hat{F}_{\bar{X}_n}^{(N)}(x_N)\right)}{\log x_N},$$

$$\hat{\alpha}(x_N) = \tilde{\alpha}(x_N)\mathbb{1}\left[0 < F_{\bar{X}_n}^{(N)}(x_N) < 1\right] + \tilde{\alpha}_0 \mathbb{1}\left[F_{\bar{X}_n}^{(N)}(x_N) = 0 \quad or \quad 1\right]. \tag{12}$$

$\hat{\alpha}(x_N)$ is the tail index estimator.

2.10 Nuyts estimator

Nuyts (2010) suggested to use the Simpson's rule for an improvement of Hill estimator. The tail index α is estimated numerically as a solution of the following equation:

$$\frac{1}{k} \sum_{i=0}^{k-1} \log\left(X_{(n-i)}\right) = \frac{1}{\alpha} + \frac{\log X_{(k+1)} X_{(k+1)}^{-\alpha} - \log X_{(n)} X_{(n)}^{-\alpha}}{X_{(k+1)}^{-\alpha} - X_{(n)}^{-\alpha}}. \tag{13}$$

2.11 Hall class of estimators

Hall (1982) developed a tail index estimator under A2 assumption.

Define a_j and b_j as constants satisfying $a_{m+1} > a_m > ... > a_1 > 0$ (m is defined in A2) and

$$\sum_{j=1}^{m+1} b_j a_j^s = \begin{cases} 1 & \text{if } s = 0 \\ 0 & \text{if } 1 \le s \le m \end{cases}$$

The class of estimators is defined as

$$\hat{\alpha} = \left(\sum_{j=1}^{m+1} b_j \hat{\gamma}_n^H(\lfloor a_j k \rfloor) \right)^{-1}, \tag{14}$$

where $\hat{\gamma}_n^H(k)$ denotes estimates received with the Hill estimator and $\lfloor \cdot \rfloor$ - the integer part of $a_j k$.

2.12 Vaičiulis (2009) estimator

Vaičiuls (2009) proposed an estimator based on an increment ratio statistics:

$$IR_{n,m} := \frac{1}{n-2m+1} \sum_{i=0}^{n-2m} \frac{|\sum_{t=i+1}^{i+m} X_t^2 - \sum_{t=i+m+1}^{i+2m} X_t^2|}{\sum_{t=i+1}^{i+m} X_t^2 + \sum_{t=i+m+1}^{i+2m} X_t^2},$$

where m is a bandwidth. Using Monte-Carlo simulations, he suggested the following expression for $IR_{n,m}$ conversion to $\hat{\alpha}$

$$\hat{\alpha} \approx \big(2.0024 - 0.4527 IR_{n,m} + 0.4246 IR_{n,m}^2 - 0.1386 IR_{n,m}^3\big) cos\left(\frac{\pi IR_{n,m}}{2}\right). \tag{15}$$

3 Quantile plots

First, in this section, quantile plots for $\gamma > 0$ are discussed. Next, the case of $\gamma \in \mathbb{R}$ is considered.

3.1 Zipf plots

Quantile plots for tail index estimation were introduced by Zipf (1941, 1949). Kratz and Resnick (1996), Schultze and Steinebach (1996) and Beirlant et al. (1996b) studied this method from various perspectives. Examine a scatter-plot with coordinates $\big(-\log\big(j/(n+1)\big), \log X_{(n-j+1)}\big)$, $j = 1, ..., n$. If the right side of this plot is 'almost' linear (suppose, this is so for the last k observations) its slope corresponds to γ, and it can be estimated by applying the OLS with intercept:

$$\hat{\gamma}_1 = \frac{\sum_{j=1}^k \log\big((k+1)/j\big) \log X_{(n-j+1)} - k^{-1} \sum_{j=1}^k \log\big((k+1)/j\big) \sum_{j=1}^k \log X_{(n-j+1)}}{\sum_{j=1}^k \log^2\big((k+1)/j\big) - k^{-1}\big(\sum_{j=1}^k \log\big((k+1)/j\big)\big)^2}. \tag{16}$$

It is also possible to apply a weighted OLS regression (Csörgő and Viharos 1998; Viharos 1999).

3.2 Schultze and Steinebach estimators

In addition, Schultze and Steinebach (1996) introduced two other estimators. One is based on the regression line with no intercept:

$$\hat{\gamma}_2 = \frac{\sum_{j=1}^{k} \log(n/j) \log X_{(n-j+1)}}{\sum_{j=1}^{k} \log^2(n/j)}. \tag{17}$$

The second one is similar to the estimator (16), but with reversed dependent and explanatory variables:

$$\hat{\gamma}_3 = \frac{\sum_{j=1}^{k} \log^2\big(X_{(n-j+1)}\big) - k^{-1}\big(\sum_{j=1}^{k} \log X_{(n-j+1)}\big)^2}{\sum_{j=1}^{k} \log\big((k+1)/j\big) \log X_{(n-j+1)} - k^{-1} \sum_{j=1}^{k} \log\big((k+1)/j\big) \sum_{j=1}^{k} \log X_{(n-j+1)}}. \tag{18}$$

3.3 Brito and Freitas estimator

Brito and Freitas (2003) suggested using a geometric mean of the estimators $\hat{\gamma}_1$ and $\hat{\gamma}_3$: $\hat{\gamma} = \sqrt{\hat{\gamma}_1 \hat{\gamma}_3}$, introduced by Schultze and Steinebach. They showed that $\hat{\gamma}_1 \leq \hat{\gamma} \leq \hat{\gamma}_3$.

3.4 Hüsler et al. estimator

Hüsler et al. (2006) introduced a weighted least squares estimator of the following form:

$$\hat{\gamma} = \frac{\sum_{i=0}^{k-1} g\big(\frac{i+1}{k+1}\big) \log\big(\frac{i+1}{k+1}\big) \log\big(\frac{X_{(n-i)}}{X_{(n-k)}}\big)}{\sum_{i=0}^{k-1} g\big(\frac{i+1}{k+1}\big) \log\big(\frac{i+1}{k+1}\big)^2}, \tag{19}$$

where $g(\cdot) \geq 0$ on the interval (0,1).

3.5 Beirlant et al. estimators

Beirlant et al. (1996b) suggested the regression line to force going through the anchor point $\big(-\log\big((k+1)(n+1)^{-1}\big), \log X_{(n-k)}\big)$. They suggested using the weighted OLS estimators of the regression line. If the weight for observation j is taken as $w_j = 1/\log[(k+1)/j]$, this method coincides with the Hill estimator. The other weights make the method similar to the Kernel estimator (Csörgő, Deheuvels, and Mason 1985). The advantage of this method is that k can be chosen so that it minimizes the weighted mean squared error.

3.6 Aban and Meerschaert estimators

Aban and Meerschaert (2004) studied the best linear unbiased estimator of the quantile regression, which takes into account the mean and covariance structure of the largest order statistics. The resulting estimator is equivalent to

$$\hat{\gamma} = k\hat{\gamma}_k^H/(k-1). \tag{20}$$

Another estimator, which has slightly higher variance, suggested by Aban and Meerschaert, is

$$\hat{\gamma} = \sum_{i=1}^{k} s_i \log X_{(n-i+1)}, \tag{21}$$

were

$$s_i = \frac{\bar{a}_k(a_{n-i+1} - \bar{a}_k)}{\sum_{i=1}^{k}(a_{n-i+1} - \bar{a}_k)^2},$$

$a_r = \sum_{j=1}^{r}(n-j+1)^{-1}$ and $\bar{a}_k$ is the arithmetic mean of a_{n-i+1}, $i = 1, ..., k$.

In a remark, Aban and Meerschaert also suggested to reduce the bias by using a regression of the following form:

$$\log X_{(n-j+1)} = const - \gamma \log\left(\frac{j-1/2}{n}\right), \quad j = 1, ..., k. \tag{22}$$

Formally, its consistency was shown by Gabaix and Ibragimov (2011).

3.7 Gabaix and Ibragimov estimator

Apart from showing consistency of the estimator (22), Gabaix and Ibragimov (2011) also considered a similar estimator with reversed dependent and explanatory variables, and showed that $-1/2$ is an optimal shift:

$$\log\left(\frac{j-1/2}{n}\right) = const - \alpha \log X_{(n-j+1)}, \quad j = 1, ..., k. \tag{23}$$

Furthermore, they considered two harmonic estimators. Define $H(j) = \sum_{i=1}^{j} j^{-1}$.

$$H(j-1) = const - \gamma \log X_{(n-j+1)}, \quad j = 1, ..., k, \tag{24}$$

$$\log X_{(n-j+1)} = const - \alpha H(j-1), \quad j = 1, ..., k. \tag{25}$$

3.8 Beirlant et al. (1999) bias-reduced quantile plots

Beirlant et al. (1999) suggested a modification of the quantile plot estimation method, which reduces a bias of the estimate. They noticed that the slowly varying function $L(x)$ in assumption A1 may cause a serious bias, and suggested to take it into account when analyzing quantile plots. Again, the Pareto quantile plot to the right of the point $-\log\big((k+1)/(n-1), \log X_{(n-k)}\big)$ was analyzed. The method relies on the maximum likelihood estimation of the parameters γ, ρ and $b_{n,k}$ in the following model:

$$U_{j,k} = \left(\gamma + b_{n,k}\left(\frac{j}{k+1}\right)^{-\rho}\right) f_{j,k}, \quad 1 \le j \le k, \tag{26}$$

where $U_{j,k}$ is defined in equation (1) and $f_{j,k}$ is an i.i.d exponential random variable. If the term $b_{n,k}\big(j/(k+1)\big)^{-\rho}$ is ignored, the MLE of the equation (26) reduces to the Hill estimator. The asymptotic properties of this estimator were shown by Beirlant et al. (2002). There is also a modification of this estimator for censored data (Beirlant and Guillou 2001).

3.9 Generalized quantile plots

Beirlant et al. (1996a) suggested to use quantile plots for γ which is not limited by positive values. The method results in fitting a linear regression through the k points which correspond to k largest X.

$$\big(-\log(j/n), \log UH_{j,n}\big), \quad j = 1, ..., n, \tag{27}$$

where

$$UH_{j,n} = X_{(n-j)}\hat{\gamma}_n^H(k),$$

and $\hat{\gamma}_n^H(k)$ is the Hill estimator.

Beirlant et al. (2005) proposed a simplification of the OLS estimator for $\gamma \in \mathbb{R}$:

$$\hat{\gamma} = \frac{1}{k}\sum_{j=1}^{k}\left(\log\frac{k+1}{j} - \frac{1}{k}\sum_{i=1}^{k}\log\frac{k+1}{i}\right)\log UH_{j,n}. \tag{28}$$

They also generalized the Hill estimator:

$$\hat{\gamma} = \frac{1}{k}\sum_{i=1}^{k}\log\left(\frac{UH_{j,n}}{UH_{k+1,n}}\right). \tag{29}$$

However, it works for $\gamma > 0$ only.

4 Minimal distance estimators

Methods based on quantile plots fit a regression line to a linear part of the Q-Q plot. Usually, this is made with an OLS or a similar method. There are also several methods which were designed as minimal distance estimators with no relations to Q-Q plots.

4.1 Vanderwale et al. estimator

Vandewalle et al. (2007) suggested to use a minimization of the least square errors for tail index estimations. Namely,

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \int \big(f(x|\theta) - f(x)\big)^2 dx,$$

were $f(\cdot)$ denotes a density function of normalized tail observations: $X_{(n-j+1)}/X_{(n-k)}$. Denote $\theta = \{\gamma, \delta, \rho\}$. Then, having assumed the following density function

$$f_\theta(x) = (1-\delta)\left[\frac{1}{\gamma}x^{-(1+1/\gamma)}\right] + \delta\left[\left(\frac{1}{\gamma} - \rho\right)x^{-(1+1/\gamma)-\rho}\right],$$

the estimator reduces to that in equation (30).

$$\varrho(\theta) = \frac{(1+\delta)^2}{(\gamma+2)\gamma} + 2\frac{\delta(1-\delta)(1-\gamma\rho)}{(\gamma+2-\gamma\rho)\gamma} + \frac{\delta^2(1-\gamma\rho)^2}{(\gamma+2-2\gamma\rho)\gamma},$$

$$(\hat{\theta}, \hat{w}) = \underset{\theta, w}{\operatorname{argmin}} \left[w^2 \varrho(\theta) - \frac{2w}{k} \sum_{i=1}^{k} f_\theta \left(\frac{X_{(n-j+1)}}{X_{(n-k)}} \right) \right]. \tag{30}$$

$\hat{\gamma}$ is one of the $\hat{\theta}$ elements. The authors claim that the method is more robust compared with those based on ML.

4.2 Tripathi et al. estimators

Tripathi et al. (2014) improved the Hill estimator taking specific parametric forms, which generalize the Hill estimator, and minimizing the loss function $\left(\hat{\alpha}/\alpha - 1\right)^2$. Denote $S_k = \sum_{i=0}^{k-1} \log\left(X_{(n-i)}/X_{(n-k)}\right)$. The best tail-index estimator in the class of s/S_k is found to be

$$\hat{\alpha}_1 = \frac{k-3}{S_k}. \tag{31}$$

In the class of estimators of the form $s/(S_k + \max(0, \log X_k))$, the optimal 'supremum' estimator is

$$\hat{\alpha}_2 = \frac{k-2}{S_k + \max(0, \log X_k)}. \tag{32}$$

The 'infimum' estimator is:

$$\hat{\alpha}_3 = \frac{k-3}{S_k + \max(0, \log X_k)}. \tag{33}$$

The authors showed that these estimators dominate the Hill estimator performance in terms of the quadratic loss function; however, the optimization was made on the $0 < \alpha < 1$ interval. This assumption restricts the set of possible applications.

5 Bias-reduced estimators

Often bias-reduced estimators take the second-order parameter ρ into account, and are based on assumptions A2-A4.

5.1 Feuerverger and Hall estimators

Feuerverger and Hall (1999) tried to reduce the bias of the Hill and OLS tail index estimators under assumption A3: Denote $v_i = logU_i$, where U_i is defined in equation (1) The MLE estimator of α is given by

$$\hat{\alpha} = \left[k^{-1} \sum_{i=1}^{k} U_i \exp\left\{ -\hat{D}(i/n)^{-\hat{\rho}} \right\} \right]^{-1}. \tag{34}$$

and $\hat{\rho}$ and $\hat{D}$ are obtained from minimization of

$$L(D,\rho) = Dk^{-1}\sum_{i=1}^{k}(i/n)^{-\rho} + \log\Big[k^{-1}\sum_{i=1}^{k}U_i \exp\big\{-D(i/n)^{-\rho}\big\}\Big].$$

OLS estimate of α is received from

$$S(\mu, D, \rho) = \sum_{i=1}^{k}\big\{v_i - \mu - D(i/n)^{-\rho}\big\}^2$$

minimization with respect to μ, D and ρ, and

$$\hat{\alpha} = exp(\Gamma'(1) - \hat{\mu}). \tag{35}$$

where $\Gamma'(1) \approx -0.5772157$ is a derivative of gamma function at point 1. Alternatively, one can plug OLS estimates $\hat{\rho}$ and $\hat{D}$ into estimator (34). The weakness of these methods is that convergence problems often exist (Gomes and Oliveira 2003).

Gomes and Martins (2004) simplified Feuerverger and Hall estimators by assuming $\rho = -1$. Furthermore, they approximated $\exp\big\{-Di/n\big\}$ as $1 - Di/n$, leading o the following approximation of equation (34):

$$\hat{\gamma}^{ML}(k) = \hat{\gamma}_n^H(k) - \Big(\frac{1}{k}\sum_{i=1}^{k} iU_i\Big)\frac{\sum_{i=1}^{k}(2i-k-1)U_i}{\sum_{i=1}^{k} i(2i-k-1)U_i}, \tag{36}$$

where $\hat{\gamma}_n^H(k)$ is the Hill estimator. The OLS estimator (35) is rewritten as

$$\hat{\gamma}^{LS}(k) = exp\Big\{\frac{2(2k+1)}{k(k-1)}\sum_{i=1}^{k} logU_i - \Gamma'(1) - \frac{6}{k(k-1)}\sum_{i=1}^{k} ilogU_i\Big\}, \tag{37}$$

Gomes et al. (2007) suggested estimating the second-order parameters separately with a larger k.

5.2 Peng estimators

Using assumption A4, Peng (1998) developed an asymptotically unbiased estimator of the following form ($\gamma > 0$):

$$\hat{\rho} = (\log 2)^{-1}\log\frac{M_n(n/(2\log n)) - 2(\hat{\gamma}_n^H(n/(2\log n)))^2}{M_n(n/\log n) - 2(\gamma_n^H(n/\log n))^2}$$

$$\gamma_n(k) = \hat{\gamma}_n^H - \frac{M_n(k) - 2(\hat{\gamma}_n^H(k))^2}{2\hat{\gamma}_n^H(k)\hat{\rho}_n}(1 - \hat{\rho}_n), \tag{38}$$

where $M_n(k) = M_{k,n}^2$ defined in equation (3), and $\hat{\gamma}_n^H$ denotes the Hill estimator (eq. 4), ρ is the second-order parameter. The author claims that the estimator remains unbiased even for large k; however, its asymptotic properties were still shown using assumptions B1 and B2. For a more general γ, Peng modified the Pickands estimator:

$$\hat{\rho} = (\log 2)^{-1} \log \frac{\hat{\gamma}_n^P(n/(2\log n)) - \hat{\gamma}_n^P(n/(4\log n))}{\hat{\gamma}_n^P(n/\log n) - \hat{\gamma}_n^P(n/(2\log n))},$$
$$\hat{\gamma}_n(k) = \hat{\gamma}_n^P(k) - \frac{\hat{\gamma}_n^P(k) - \hat{\gamma}_n^P(k/4)}{1 - 4^{\hat{\rho}}}, \quad (39)$$

where $\hat{\gamma}_n^P(k)$ is the Pickands estimator defined in equation (70). Despite of good theoretical asymptotic properties of the Peng estimators, our simulations show that the expressions under the logarithms in $\hat{\rho}$ estimation obtain negative values sometimes, leading to a collapse in calculations.

5.3 Huisman et al. estimator

Huisman et al. (2001) noted, that if assumption A2 is satisfied with $m = 1$, the bias of the Hill estimator is almost linear in k. As small values of k result in a lower bias (at the cost of higher variance), they suggested to estimate $\gamma(r)$ with the Hill estimator $r = 1, ..., k$, and then run a regression

$$\gamma(r) = \beta_0 + \beta_1 r + \epsilon(r), \quad r = 1, ..., k. \quad (40)$$

$\hat{\beta}_0$ is a bias-free estimate of the tail index.

5.4 Gomes et al. (2000,2002) Jackknife estimators

Gomes et al. (2000) studied Jackknife estimators of the following form:

$$\hat{\gamma} = \frac{\hat{\gamma}_1 - q\hat{\gamma}_2}{1 - q},$$

where $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are some consistent estimators of γ and $q = d_1(n)/d_2(n)$ - is the ratio of biases of these estimators. They considered the following special cases:

$$\hat{\gamma}_1 = 2\hat{\gamma}^{(2)}(k) - \hat{\gamma}^{(1)}(k) \quad (41)$$
$$\hat{\gamma}_2 = 4\hat{\gamma}^{(3)}(k) - 3\gamma^{(1)}(k), \quad (42)$$
$$\hat{\gamma}_3 = 3\hat{\gamma}^{(2)}(k) - 2\gamma^{(3)}(k), \quad (43)$$
$$\hat{\gamma}_4 = 2\hat{\gamma}^{(1)}(k/2) - \gamma^{(1)}(k), \quad (44)$$

where $\hat{\gamma}^{(1)}(k)$ is the Hill estimator (eq. 4), $\hat{\gamma}^{(2)}(k)$: Danielsson et al. estimator (eq. 8) with $l = 1$, and $\hat{\gamma}^{(3)}(k)$ is a Gomes and Martins estimator with $l = 2$ (eq. 83):

The estimator (42) is a simplified version of the Peng's estimator (38) with $\rho = -1$. More complete versions of estimators (43) and (44) were studied by Gomes and Martins (2002) and Gomes et al. (2002) under assumption A4:

$$\hat{\gamma}_3 = \frac{-(2 - \hat{\rho})\hat{\gamma}^{(2)}(k) + 2\hat{\gamma}^{(3)}(k)}{\hat{\rho}}, \quad (45)$$

$$\hat{\gamma}_4 = \frac{\hat{\gamma}^{(1)}(k) - 2^{-\hat{\rho}}\hat{\gamma}^{(1)}(k/2)}{1 - 2^{-\hat{\rho}}}. \tag{46}$$

$\hat{\rho}$ can be estimated as

$$\hat{\rho}_\tau(k) = -\left|\frac{3(T^{(\tau)}(k) - 1)}{(T^{(\tau)}(k) - 3)}\right|$$

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

$M_{k,n}^{(l)}$ is defined in equation (3) and τ is a tuning parameter. This estimator of ρ was introduced by Fraga Alves et al. (2003).

In Gomes et al. (2002) also several other generalized Jackknife estimators were considered. One is based on the work of Quenouille (1956):

$$\hat{\gamma}_{n,\hat{\rho}}^G(k) = \frac{\hat{\gamma}_n - \left(n/(n-1)\right)^{\hat{\rho}}\bar{\hat{\gamma}}_n}{1 - \left(n/(n-1)\right)^{\hat{\rho}}}, \quad \bar{\hat{\gamma}}_n = \frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_{n-1,i}(k), \tag{47}$$

where $\hat{\gamma}_n(k)$ is a consistent estimator of γ based on $n-1$ observations obtained from the original sample, after exclusion of the i-th element, and $\hat{\rho}$ is a suitable estimator of ρ. They also studied a simplified version of this estimator with $\rho = -1$.

The second estimator is a modification of estimator (46) for the case of Fréchet model:

$$\hat{\gamma}^{G_2} = \frac{\hat{\gamma}_n(k) - \frac{log(1-k/n)}{log(1-k/(2n))}\hat{\gamma}_n(k/2)}{1 - \frac{log(1-k/n)}{log(1-k/(2n))}} \tag{48}$$

The third estimator is designed for generalized Pareto and the Burr models:

$$\hat{\gamma}^{G_1} = \frac{(2 + k/n)\hat{\gamma}_n(k/2) - \hat{\gamma}_n(k)}{1 + k/n}. \tag{49}$$

5.5 Gomes et al. (2005) Jackknife estimator

Another Jackknife estimator was introduced by Gomes et al. (2005). First, they introduced another generalised class of estimators:

$$\hat{\gamma}_n^{(s)}(k) = \frac{s^2}{k^s}\sum_{i=1}^{k} i^{s-1} log\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right), \quad s \geq 1. \tag{50}$$

If $s = 1$, this estimator simplifies to the Hill estimator. Combining the Hill estimator with an estimator with $s \neq 1$, they received a Jackknife estimator:

$$\hat{\gamma}_n^G = -\frac{s(1-\hat{\rho})}{\hat{\rho}(s-1)}\left[\hat{\gamma}_n^{(1)}(k) - \frac{s-\hat{\rho}}{s(1-\hat{\rho})}\hat{\gamma}^{(s)}(k)\right]. \tag{51}$$

5.6 Gomes et al. (2007) Jackknife estimator

Gomes et al. (2007) considered generalised estimators

$$\hat{\gamma}_1^{(s)}(k) = \frac{s}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{s-1} U_i, \quad s \geq 1 \tag{52}$$

$$\hat{\gamma}_2^{(s)}(k) = -\frac{s^2}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{s-1} \log\left(\frac{i}{k}\right) U_i, \quad s \geq 1 \tag{53}$$

They considered a number of Jackknife estimators similar to the previous papers. An optimal (variance reducing) combination of two estimators (52) (with different s) is

$$\hat{\gamma}^{GJ_1}(\hat{\rho}) = \frac{(1-\hat{\rho})^2\hat{\gamma}_1^{(1)}(k) - (1-2\hat{\rho})\hat{\gamma}_1^{(1-\hat{\rho})}(k)}{\hat{\rho}^2}. \tag{54}$$

where $\hat{\rho}$ is an estimate of the second-order parameter. Note that $\hat{\gamma}_1^{(1)}(k)$ is the Hill estimator. An optimal combination of estimators (52) and (53) is

$$\hat{\gamma}^{GJ_2}(\hat{\rho}) = \frac{1}{\hat{\rho}}(\hat{s}\hat{\gamma}_1^{(\hat{s})}(k) - (\hat{s}-\hat{\rho})\gamma_2^{(\hat{s})}(k)). \tag{55}$$

where $\hat{s}$ is such that

$$3\hat{s}^3 - 5\hat{s}^2 + \hat{s}(\hat{\rho}^2 - \hat{\rho} + 3) - (2\hat{\rho}^2 - 2\hat{\rho} + 1) = 0.$$

In the above-mentioned paper, also an optimal Jackknife estimator based on (53) with two different s, was studied, but the authors do not provide an optimal combination. The authors also study the case of $\hat{\rho}$ equalized to -1. In this case, the Jackknife estimator with the lowest variance based on (52) is $\hat{\gamma}^{GJ_1}(-1) = 4\hat{\gamma}_1^{(1)}(k) - 3\hat{\gamma}_1^{(2)}(k)$.

5.7 Caeiro et al. 2005 bias-reduced Hill's estimator

Caeiro et al. (2005) introduced two direct ways for reducing the bias of the Hill estimator:

$$\hat{\gamma}_{\hat{\beta},\hat{\rho}}(k) = \hat{\gamma}_n^H(k)\left(1 - \frac{\hat{\beta}}{1-\hat{\rho}}\left(\frac{n}{k}\right)^{\hat{\rho}}\right), \tag{56}$$

$$\hat{\gamma}_{\hat{\beta},\hat{\rho}}(k) = \hat{\gamma}_n^H(k)\exp\left(-\frac{\hat{\beta}}{1-\hat{\rho}}\left(\frac{n}{k}\right)^{\hat{\rho}}\right), \tag{57}$$

where $\hat{\gamma}_n^H(k)$ is the Hill estimator. ρ and β are second-order parameters. Consistent estimators for $\hat{\rho}$ can be found in subsections (5.2) or (5.4). β can be estimated as

$$\hat{\beta}(k) = \left(\frac{k}{n}\right)^{\hat{\rho}} \frac{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\hat{\rho}}\right)\left(\frac{1}{k}\sum_{i=1}^{k}U_i\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\hat{\rho}}U_i\right)}{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\hat{\rho}}\right)\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\hat{\rho}}U_i\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-2\hat{\rho}}U_i\right)}.$$

The authors argued that it could be wise to use a higher k than that in the tail index estimation for estimation of the second-order parameters. Alternatively, one may use estimators of the second-order parameters discussed in detail by Caeiro and Gomes (2006).

5.8 Gomes, Figueiredo and Mendonça best linear unbiased estimators (BLUE)

Gomes et al. (2005) look for a BLUE in the presence of second-order regular variation condition. Define $\boldsymbol{T} = \hat{\gamma}_i$; $i = k-m+1, ..., k$; $1 \le m \le k$, where $\hat{\gamma}_i$ is an estimator of γ. They assume that the covariance matrix of $\boldsymbol{T}$ can be approximated as $\gamma^2\Sigma$, and its mathematical expectation can be asymptotically approximated as $\gamma\boldsymbol{s} + \phi(n,k)\boldsymbol{b}$, where $\boldsymbol{s}$ is a vector of unities of length m, and the second term in the sum accounts for a bias. Then, they look for a vector $\boldsymbol{a}' = (a_1, a_2, ..., a_m)$ such that $a'\Sigma a$ is minimal, with the restrictions $\boldsymbol{a}'\boldsymbol{s} = 1$ and $\boldsymbol{a}'\boldsymbol{b} = 0$. The result of this optimization problem is:

$$\boldsymbol{a} = \Sigma^{-1}\boldsymbol{P}(\boldsymbol{P}'\Sigma^{-1}\boldsymbol{P})^{-1}\left[\begin{array}{c}1\\0\end{array}\right],$$

where $\boldsymbol{P} = [\boldsymbol{s}\ \boldsymbol{b}]$. The BLUE estimator is given by

$$\hat{\gamma}^{BLUE} = \boldsymbol{a}'\boldsymbol{T}.$$

In case of Hill estimator, and having assumed a model similar to assumption A3, $m = k$, and $\rho = -1$, this estimator simplifies to

$$\hat{\gamma} = \frac{6}{k^2-1}\sum_{i=1}^{k-1} i\hat{\gamma}_n^H(i) - \frac{2k-1}{k+1}\hat{\gamma}_n^H(k), \tag{58}$$

where $\gamma_n^H(i)$ is the Hill estimator (4). The authors also provide an explicit expression for the case of a more general ρ:

$$\hat{\gamma} = \sum_{i=1}^{k+1} a_i(\hat{\rho}) \log X_{(n-i+1)}, \tag{59}$$

with

$$a_i(\hat{\rho}) = \frac{1}{k}\left(\frac{1-\hat{\rho}}{\hat{\rho}}\right)^2 \left\{1 - \frac{k(1-2\hat{\rho})}{1-\hat{\rho}}\left[\left(\frac{i}{k}\right)^{1-\hat{\rho}} - \left(\frac{i-1}{k}\right)^{1-\hat{\rho}}\right]\right\}, \quad i = 1, ..., k,$$

$$a_{k+1}(\hat{\rho}) = -\left(\frac{1-\hat{\rho}}{\hat{\rho}}\right),$$

where $\hat{\rho}$ is a consistent estimate of the second-order parameter.

5.9 Beirlant et al. 2008 estimator

Beirlant et al. (2008) proposed a bias-reduced tail-index estimator with asymptotic variance being equal to the Hill estimator:

$$\hat{\gamma} = \hat{\gamma}_n^H - \hat{\beta}_{\hat{\rho},k}\left(\frac{n}{k}\right)^{\hat{\rho}} P_{\hat{\rho},k}, \tag{60}$$

where $\hat{\gamma}_n^H$ is the Hill estimator, $\hat{\rho}$ is a consistent estimator of the second-order parameter and

$$\hat{\beta}_{\hat{\rho},k} = \left(\frac{k+1}{n+1}\right)^{\hat{\rho}} \frac{p_{\hat{\rho},k}P_{0,k} - P_{\hat{\rho},k}}{p_{\hat{\rho},k}P_{\hat{\rho},k} - P_{2\hat{\rho},k}},$$
$$P_{\hat{\rho},k} = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k+1}\right)^{-\hat{\rho}} U_i.$$

$p_{\hat{\rho},k}$ in the definition of $\hat{\beta}_{\hat{\rho},k}$ can be calculated as

$$p_{\hat{\rho},k} = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k+1}\right)^{-\hat{\rho}}.$$

They also found that the asymptotic optimal level for k used for $P_{\hat{\rho},k}$ estimation is $k_0 = \left((1-2\rho)n^{-2\rho}/(-2\rho\beta^2)\right)^{1/(1-2\rho)}$, where $\beta = D\rho/\gamma$, and D is defined in assumption A3. At the same time, it is reasonable to estimate parameters $\hat{\rho}$ and $\hat{\beta}$ using k_1 observations, which is higher than k, such as $k_1 = n^{1-\epsilon}$, with a (relatively) small ϵ, such as 0.05.

5.10 Baek and Pipiras estimators

Baek and Pipiras (2010) considered a distribution with second-order parameter in the form of assumption A5, which resulted in the following OLS estimator:

$$argmin_{\beta_0,\alpha,\beta_1}\sum_{i=1}^{k}\left(\log\left(\frac{i}{k}\right) - \beta_0 + \alpha\log\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right) - \beta_1\left(\frac{X_{(n-k)}}{n-i+1}\right)\right)^2. \tag{61}$$

They also suggested changing $\log(i/k)$ with $\log(i-0.5)/k$. Alternatively, it is possible to minimize

$$argmin_{\beta_0,\alpha,\beta_1}\sum_{i=1}^{k}\left(\log\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right) - \beta_0 + \frac{1}{\alpha}\log\left(\frac{i}{k}\right) - \beta_1\left(\frac{X_{(n-k)}}{n-i+1}\right)\right)^2. \tag{62}$$

Apart from these two estimators, Baek and Pipiras adopted a number of other estimators for assumption A5. For example, using methodology of Aban and Meerschaert (2004) for searching for the BLUE they received

$$\hat{\gamma} = \hat{\gamma}_H - \frac{\beta_1}{k}\sum_{i=1}^{k}(X^{-1}_{(n-i+1)} - X^{-1}_{(n-k+1)}), \tag{63}$$

where β_1 could be estimated from equation (62). Having adopted the methodologies of Gomes et al. (2000) and Peng (1998) they received the following estimators:

$$\hat{\alpha}_G = \frac{2(2+\gamma_H)}{M_{k,n}^{(2)}} - \frac{2}{\hat{\gamma}_H}\sqrt{\frac{2}{M_{k,n}^{(2)}}}. \tag{64}$$

$$\hat{\alpha}_P = -(\gamma_H)^{(-2)} + 2(\gamma_H+1)/M_{k,n}^{(2)}, \tag{65}$$

where $M_{k,n}^{(2)}$ is defined in equation (3). Furthermore, they provided expressions for a modified MLE introduced by Feuerverger and Hall (1999) and conditional MLE.

5.11 Brito et al. (2016) estimators

Brito et al. (2016) suggested the following estimator:

$$\hat{\gamma}_B(k) = \frac{M_{k,n}^{(2)} - [M_{k,n}^{(1)}]^2}{k^{-1}\sum_{i=1}^{k}\log^2(n/i) - \Big(k^{-1}\sum_{i=1}^{k}\log(n/i)\Big)^2}. \tag{66}$$

Furthermore, they analysed its two biased-reduced modifications, which are similar, but not absolutely equivalent, to those introduced by Caeiro et al. (2005) (eq. 56 and 57):

$$\hat{\gamma}_{\hat{\beta},\hat{\rho}}(k) = \hat{\gamma}_B(k)\bigg(1 - \frac{\hat{\beta}}{(1-\hat{\rho})^2}\Big(\frac{n}{k}\Big)^{\hat{\rho}}\bigg), \tag{67}$$

$$\hat{\gamma}_{\hat{\beta},\hat{\rho}}(k) = \hat{\gamma}_B(k)\exp\bigg(-\frac{\hat{\beta}}{(1-\hat{\rho})^2}\Big(\frac{n}{k}\Big)^{\hat{\rho}}\bigg). \tag{68}$$

5.12 Vaičiulis' bias reduced DPR

Intending to reduce the bias of the DPR estimator, Vaičiulis (2012) introduced its following modification:

$$S_{t,r} = \frac{1}{t}\sum_{j=1}^{t} f_r\bigg(\frac{M_j^{(2)}}{M_j^{(1)}}\bigg), \quad f_r(b) = \log^r \frac{1}{b}.$$

$$\hat{\alpha}_l = \frac{\sum_{r=1}^{l}(-1)^{r+1}\big(\Gamma(r+1)\big)^{-1}S_{t,r}}{\sum_{r=2}^{l+1}(-1)^{r}\big(\Gamma(r+1)\big)^{-1}S_{t,r}}, \quad l \in \boldsymbol{N}. \tag{69}$$

If l is even, $\alpha \neq 1$.

6 Estimators which allow for negative γ

The advantage of tail index estimators based on the extreme value distribution over those based on the Pareto is that the values of the EVI are not limited to positive values. Often estimators allow for $\gamma \in R$, negative values corresponding to distributions with a finite right endpoint. However, usually they have a higher asymptotic variance than estimators for positive γ. The consistency of these estimators is usually shown under assumptions A6 or A7.

6.1 Pickands estimator

Pickands (1975) proposed a simple method. Its idea is to consider the quartiles of the k largest observations. Under assumption that the tail of the distribution satisfies assumption A6, he found analytical expressions for the 3/4 and 1/2 quartiles and, having substituted theoretical quartiles with their empirical counterparts, received the following tail index estimator:

$$\hat{\gamma}_n^P(k) = \frac{1}{\log 2} \log \left(\frac{X_{(n-\lfloor k/4 \rfloor)} - X_{(n-\lfloor k/2 \rfloor)}}{X_{(n-\lfloor k/2 \rfloor)} - X_{(n-k)}} \right), \tag{70}$$

where $\lfloor u \rfloor$ denotes the integer part of u. A good property of the Pickands estimator is that it is location invariant, i.e. the estimate does not change if the sample is shifted by a constant.

6.2 Falk estimator

Falk (1994) suggested an improvement of Pickands estimator by taking a linear combination of two different numbers of observations treated as the tail.

$$\hat{\gamma}(k, p) = p\hat{\gamma}_n^P([k/2]) + (1-p)\hat{\gamma}_n^P(k), \quad p \in [0, 1] \tag{71}$$

The optimal level of p depends on the parameter γ itself: $p_{opt} = \big((2^{-2\gamma} + 2) + 2^{1-\gamma}\big)/\big(3(2^{-2\gamma} + 2) + 2^{2-\gamma}\big)$, therefore, γ needs to be preestimated. Falk argued that, in order to calculate p_{opt}, it may be wise to take $\gamma = 0$, which is a turning point between the finite and infinite right endpoint cases. In this case, $p_{opt} = 5/13$. The author also claimed that the resulting estimator outperforms the Pickands estimator in most cases apart from the case of normal distribution.

6.3 Drees improvements of the Pickands estimator

Drees (1995) extended the Falk's refinement of the Pickands estimator to

$$\hat{\gamma} = \sum_{i=1}^{m_n} c_{n,i} \hat{\gamma}_n^P(4i), \tag{72}$$

where m_n is an intermediate sequence and $c_{n,i}$, $i = 1, ..., m$ are weights $0 \le c_{n,i} \le 1$, which sum to unity. The author sets some restrictions on the choice of m_n and $c_{n,i}$ for showing asymptotic normality, and also discusses an optimal choice for weights; however, this choice depends on the parameter γ and it is rather tricky (see pages 2064-2065 in the above-mentioned paper).

Drees (1996) proposed another refinement of the Pickands estimator. He noticed that its bias can be estimated as $\hat{b} = \big(\hat{\gamma}_n^P(k) - \hat{\gamma}_n^P([k/2])\big)^2 / \big(\hat{\gamma}_n^P(k) - 2\hat{\gamma}_n^P(k/2) + \hat{\gamma}_n^P(k/4)\big)$. Therefore, Pickands estimator can be corrected as $\hat{\gamma}_n^P(k) - \hat{b}$, where for an estimation of $\hat{b}$ a preestimated value of $\hat{\gamma}$ is used. Similarly, estimator (72) was refined as

$$\hat{\gamma} = \int \hat{\gamma}_n^P(\lfloor kt \rfloor)\nu(dt) - \frac{\big(\int \hat{\gamma}_n^P(\lfloor kt \rfloor) - \hat{\gamma}_n^P(\lfloor kt \rfloor)\nu(dt)\big)^2}{\int \hat{\gamma}_n^P(\lfloor kt \rfloor) - 2\hat{\gamma}_n^P(\lfloor kt/2 \rfloor) + \hat{\gamma}_n^P(\lfloor kt/4 \rfloor)\nu(dt)}, \tag{73}$$

where $\nu(\cdot)$ denotes a probability measure on the Borel-σ-field $\mathbb{B}[0,1]$.

Drees (1998b, 1998a) proposed a generalized form of the Pickands estimator: $T(Q_n)$, where $T(\cdot)$ is a smooth functional and Q_n is a quantile function $Q_n(t) = F_n^{-1}\big(1 - tk_n/n\big) = X_{(n-\lfloor k_n t \rfloor)}$, and F_n^{-1} is an empirical quantile function. Despite this generalized form is of limited value to practitioners (because the choice of the functional $T(\cdot)$ is rather complicated), it may simplify an introduction of new estimators, which can be expressed in a similar form.

6.4 Yun estimator

Yun (2002) generalized the Pickands estimator in the following way:

$$\hat{\gamma}_{n,m}(u,v) = \frac{1}{\log v} \log \frac{X_{(m)} - X_{(\lfloor um \rfloor)}}{X_{(\lfloor vm \rfloor)} - X_{(\lfloor uvm \rfloor)}}, u, v > 0, \quad u, v \neq 1, \tag{74}$$

where $m \geq 1$, $\lfloor um \rfloor, \lfloor vm \rfloor, \lfloor uvm \rfloor \leq n$. Pickands estimator corresponds to the $\hat{\gamma}_{n,m}(1/2, 1/2)$. With the optimal values of u and v it is possible to reduce the asymptotic variance of the estimator (see the paper for a numeric algorithm). We shall also mention that earlier, Fraga Alves (1995) analysed an estimator of the $\hat{\gamma}_{n,m}(c,c)$ form, and Yun (2000) introduced a less general estimator $\hat{\gamma}_{n,m}(c, 4/c)$ $(1/4 < c < 1)$.

6.5 Segers 2005 generalized Pickands estimator

Segers (2005) proposed the following generalization of the Pickands estimator:

$$\hat{\gamma}(c,\lambda) = \sum_{j=1}^{k} \left(\lambda\Big(\frac{j}{k}\Big) - \lambda\Big(\frac{j-1}{k}\Big)\right) \log(X_{(n-\lfloor cj \rfloor)} - X_{(n-j)}), \tag{75}$$

where $\lambda(\cdot)$ is a signed Borel measure on the $(0,1]$ interval. The paper discusses an adaptive procedure for optimal $\lambda(\cdot)$ selection.

6.6 Smith estimator

Smith (1987) suggested the following tail index estimator, based on the GPD:

$$\hat{\gamma} = \frac{1}{m} \sum_{i=1}^{m} \log(1 + Y_i/u), \tag{76}$$

where u is a relatively high threshold, $Y_i = X_j - u$, where j is the index of the ith exceedance and m is the number of observations higher than u.

6.7 Moment estimator

Dekkers et al. (1989) proposed a moment estimator. It has the following form:

$$\hat{\gamma}_n^M(k) = M_{k,n}^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{\left(M_{k,n}^{(1)}\right)^2}{M_{k,n}^{(2)}}\right)^{-1}, \tag{77}$$

where $M_{k,n}^{(i)}$, $i = 1, 2$, is defined in equation (3). To show the consistency of the moment estimator, the assumption B1 about the behavior of k was changed to $k(n)/\log n^{\delta} \to \infty$ as $n \to \infty$ for some $\delta > 0$. The moment estimator was also adopted for censored data (Beirlant et al. 2007).

6.8 Gronenboom et al. kernel estimator

Following Dekkers et al. generalization of the Hill estimator (77), Gronenboom et al. (2003) generalised kernel estimator (5) for $\gamma \in \mathbb{R}$. It has the following form: $\hat{\gamma} = \hat{\gamma}_n^K(k) - 1 + q_1/q_2$, where $\hat{\gamma}_n^K(k)$ is defined in equation (5) and

$$q_1 = \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^t K_h\left(\frac{i}{n}\right)(\log X_{n-i+1} - \log X_{n-i}). \tag{78}$$

$$q_2 = \sum_{i=1}^{n-1} \frac{d}{du}[u^{t+1}K_h(u)]_{t=i/n}(\log X_{n-i+1} - \log X_{n-i}).$$

$K_h(u) = K(u/h)/h$, t is a tuning parameter $t > 0.5$, and h is a bandwidth ($h > 0$). Apart from the assumptions used by Csörgő et al., there are also assumptions that K, K' and K'' are bounded.

6.9 Müller and Rufibach smooth estimators

Müller and Rufibach (2009) noticed that order statistics $X_{(i)}$ can be rewritten as $F_n^{-1}(i/n)$, $i = 1, ..., n$, where $F_n(\cdot)$ is an empirical distribution function of X. They suggested to change F_n with a smooth estimate of the empirical distribution function $\tilde{F}_n$. The methods for such a smoothing are well known and date back to the seminal work of Nadaraya (1964). They proposed such estimators:

$$\hat{\gamma}_1(H) = \frac{1}{\log 2} \log \left(\frac{H^{-1}\big((n-r(H)+1)/n\big) - H^{-1}\big((n-2r(H)+1)/n\big)}{H^{-1}\big((n-2r(H)+1)/n\big) - H^{-1}\big((n-4r(H)+1)/n\big)} \right), \quad (79)$$

$$\hat{\gamma}_2(H) = \frac{1}{k-1} \sum_{j=2}^{k} \log \left(\frac{X_{(n)} - H^{-1}\big((n-j+1)/n\big)}{X_{(n)} - H^{-1}\big((n-k)/n\big)} \right), \quad (80)$$

where

$$r(H) = \begin{cases} \lfloor k/4 \rfloor & \text{if } H = F_n \\ k/4 & \text{if } H = \tilde{F}_n \end{cases}$$

$\hat{\gamma}_1$ is valid for $k = 4, ..., n$ and $\hat{\gamma}_2$ for $k = 3, ..., n-1$. If $H = F_n$, $\hat{\gamma}_1$ and $\hat{\gamma}_2$ boil down to Pickands' (eq. 70) and Falk's (eq. 71) estimators.

6.10 Fraga Alves et al. 2009 mixed moment estimator

Fraga Alves et al. (2009) considered conditions that a distribution function shall follow in order to be in the domain of attraction of the generalized extreme value distribution function. These conditions were introduced by De Haan (1970). Having substituted these conditions with their empirical counterparts, they received the following estimator, which is valid for $\gamma \in \mathbb{R}$:

$$\hat{\phi}_n(k) = \big(M_{k,n}^{(1)} - L_{n,k}\big)/\big(L_{n,k}\big)^2,$$

where $M_{k,n}^{(1)}$ is defined in equation (3), and $L_{n,k}$:

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

$$\hat{\gamma}^{MM} = \frac{\hat{\phi}_n(k) - 1}{1 + 2min(\hat{\phi}_n(k) - 1, 0)}. \quad (81)$$

They also considered a location-invariant peaks over random threshold (PORT) version of this estimator, when the original sample X_i is replaced by $X_i^* = X_i - X_{([np]+1)}$, with a tuning parameter p, $0 < p < 1$.

7 Generalised classes of estimators

7.1 Gomes and Martins generalizations of the Hill estimator

Gomes and Martins (2001) introduced two estimators, which generalize the Hill estimator:

$$\hat{\gamma}_1^{(l)}(k) = \frac{M_{k,n}^{(l)}}{\Gamma(l+1)[\hat{\gamma}^{(1)}(k)]^{l-1}}. \quad (82)$$

$$\hat{\gamma}_2^{(l)}(k) = \left(\frac{M_{k,n}^{(l)}}{\Gamma(l+1)}.\right)^{\frac{1}{l}}. \tag{83}$$

where $l > 0$ and $M_{k,n}^{(l)}$ is defined in (3). If $l = 1$, they reduce to the Hill estimator. Estimator (83) is a special case of a class of estimators introduced by Segers (2001). Moreover, both estimators are special cases of a very general class of estimators introduced by Paulauskas and Vaičiulis (2017a).

7.2 Caeiro and Gomes generalized class of estimators

Caeiro and Gomes (2002) proposed another generalized class of estimators:

$$\gamma_n^{(l)}(k) = \frac{\Gamma(\alpha)}{M_{k,n}^{(l-1)}(k)}\left(\frac{M_{k,n}^{(2l)}(k)}{\Gamma(2l+1)}\right)^{1/2}, \quad \alpha \geq 1. \tag{84}$$

and $M_{k,n}^{(0)}(k) = 1$. l is a tuning parameter. When $l = 1$, Caeiro and Gomes estimator (84) and Gomes and Martins estimator (83) (with the same l) coincide. An optimal $\hat{l}$ is given by

$$\hat{l} = -\frac{log[1-\hat{\rho}-\sqrt{(1-\hat{\rho})^2-1}]}{log(1-\hat{\rho})},$$

where $\hat{\rho}$ is a consistent estimate of the second-order parameter. Gomes et al. (2004) showed that estimator (84) may achieve a high efficiency in comparison to the Hill estimator, if a number of top-order statistics is larger than the one usually used for the estimation through the Hill estimator.

7.3 Mean-of-order-p class of estimators

Mean-of-order-p class of estimators was independently introduced by Brilhante et al. (2013) and Beran et al. (2014). The authors noticed that the Hill estimator can be expressed as the natural logarithm of geometric mean of $V_{i,k}$, $1 \leq i \leq k$. $V_{i,k}$ is defined in equation (2). They suggested to generalize it to the following mean-of-order-p form:

$$\hat{\gamma}(p) = \begin{cases} \frac{1}{p}\left(1-\left(\frac{1}{k}\sum_{i=1}^{k} V_{i,k}^{p}\right)\right) & \text{if } p \leq 1/\gamma,\ p \neq 0, \\ \log\left(\prod_{i=1}^{k} V_{i,k}\right)^{1/k} & \text{if } p = 0. \end{cases} \tag{85}$$

Under assumption A4, Brilhante et al. (2014) found an optimal p to be equal to $p^* = \phi_\rho/\gamma$, where $\phi_\rho = 1 - 0.5\rho - 0.5\sqrt{\rho^2 - 4\rho + 2}$, $\phi_\rho \in (0, 1-\sqrt{2}/2]$, and ρ is the second-order parameter.

Caeiro et al. (2016) noticed that under assumption A3 this estimator is biased, and suggested to reduce the bias in the following way:

$$\hat{\gamma}(p,\beta,\rho,\phi) = \hat{\gamma}(p)\left(1 - \frac{\beta(1-\phi)}{1-\rho-\phi}\left(\frac{n}{k}\right)^{\rho}\right) \tag{86}$$

β is a second-order parameter $D = \gamma\beta$, and D is defined in assumption A3. Parameters β and ρ can be estimated (see the above-mentioned paper), $\phi = p\hat{\gamma}(p)$. It is suggested to take p as in Brilhante et al. (2014).

Gomes et al. (2016) suggested replacing X_i with $X_i^* = X_i - X_{([nq]+1)}$, with a tuning parameter q, $0 < q < 1$ (PORT methodology) and contributed on the optimal choice of k: $\hat{k}^* = \left((1-\hat{\rho})n^{-\hat{\rho}}/(\hat{\beta}\sqrt{-2\hat{\rho}})\right)^{2/(1-2\hat{\rho})}$.

7.4 Segers (2001) generalized class of estimators

Segers (2001) showed that the statistics $R_{n,k}$,

$$R_{n,k} = \frac{1}{k}\sum_{j=1}^{k} f(X_{(n-j+1)}/X_{(n-k)}),$$

converges in probability to $\mathbb{E}[f(Y^\gamma)]$, if $f : [1:\infty) \to \mathbb{R}$ is an almost everywhere (in terms of Lebesgue measure) continuous function and $|f(x)| \leq Ax^{(1-\delta)/\gamma}$, for some $A > 0$, $\delta \in (0,1)$ and $x \geq 1$. This finding gave raise to a number of methods of tail-index estimators. For example, under some additional technical assumptions if $f(y) = (\log y)^\beta$, with $\beta > 0$, tail index estimator reduces to one of the estimators studied by Gomes and Martins (2001) (eq. 83). $f(x) = x^{-p}$ with $p > -1/\gamma$ results in the following estimator:

$$\hat{\gamma} = \frac{1}{p}\left[\left(\frac{1}{k}\sum_{j=1}^{k}\left(\frac{X_{(n-k)}}{X_{(n-j+1)}}\right)^p\right)^{-1} - 1\right] \tag{87}$$

The authors also showed that in this class of estimators the Hill estimator has the smallest asymptotic variance.

7.5 Ciuperca and Mercadier generalized class of estimators

Ciuperca and Mercadier (2010) analysed in detail the following class of estimators:

$$\hat{\gamma}(g,l) = \frac{\frac{1}{k}\sum_{i=1}^{k} g\left(\frac{i}{k+1}\right)\left[\log\frac{X_{(n-i+1)}}{X_{(n-k)}}\right]^l}{\int_0^1 g(x)(-\log(x))^l dx}, \tag{88}$$

where $l > 0$ and $g(x)$ is a positive, non-increasing and integrable weight function defined on $(0,1)$. It is assumed that there exists $\delta > 0.5$ satisfying $\int_0^1 g(x)x^{-\delta}dx < \infty$ and $0 < \int_0^1 g(x)(1-x)^{-\delta}dx < \infty$. $\hat{\gamma}(1,1)$ is the Hill estimator (eq. 4), $\hat{\gamma}(1,l)$ - Gomes and Martins (2001) (eq. 83), $\hat{\gamma}(g,1)$ - corresponds to the weighted least squares estimator of Hüsler et al. (2006) (eq. 19).

7.6 Generalised DPR and Hill estimators

Paulauskas and Vaičiulis (2011) suggested to apply a Box-Cox transformation for the DPR estimator (eq. 10):

$$f_r(x) = \frac{x^r - 1}{r}, \; -\alpha < r < \infty, r \neq 0; \quad f_0(x) = \log x.$$

$$S_t = -\frac{1}{T}\sum_{j=1}^{T} f_r\left(\frac{M_j^{(2)}}{M_j^{(1)}}\right).$$

$$\hat{\alpha}_r = \frac{1 - rS_T}{S_T}. \tag{89}$$

If $1 - F(x) = C_1 x^{-\alpha} + C_2 x^{-\beta} + o(x^{-\beta})$, with $0 < \alpha < \beta \leq \infty$, the optimal r^* and m^* have the following expressions:

$$r^* = -\frac{1}{2}\left(\alpha + \beta - \sqrt{(\alpha+\beta)^2 - 2\alpha^2}\right).$$

$$m^* = \left(\frac{2n\zeta(\alpha + 2r^*)}{\alpha}\left(\frac{C_2\beta\zeta\Gamma(\beta/\alpha)}{C_1^{\beta/\alpha(\beta+r^*)}}\right)^2\right)^{\frac{1}{1+2\zeta}}$$

where $\zeta = (\beta - \alpha)/\alpha$. Nevertheless, Hill estimator with a properly selected k^* has a lower variance than the optimal Generalized DPR estimator. See also Paulauskas and Vaičiulis (2012) for applications of this estimator for max-aggregated data.

Paulauskas and Vaičiulis (2013) applied the idea of the generalized DPR estimator for a similar generalization of the Hill estimator:

$$H_n^l(k, r) = \frac{1}{k}\sum_{i=0}^{k-1} f_r^l\left(\frac{X_{n-i}}{X_{n-k}}\right), \quad l = 1, 2.$$

$$\hat{\lambda}_n^1(k, r) = H_n^{(1)}(k, r),$$

$$\hat{\lambda}_n^2(k, r) = H_n^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{H_n^{(1)}(k, r)H_n^{(1)}(k, 2r)}{H_n^{(2)}(k, r)}\right)^{-1},$$

$$\hat{\lambda}_n^3(k, r) = \frac{H_n^{(2)}(k, r)}{2H_n^{(1)}(k, 2r)}.$$

The tail index estimators are received from the relation:

$$\gamma^{(l)}(k, r) = \frac{\lambda^{(l)}(k, r)}{1 + r\lambda^{(l)}(k, r)}. \tag{90}$$

The estimator $\gamma_n^1(k, r)$ generalizes the Hill estimator (eq. 4), $\gamma_n^2(k, r)$ - moment estimator (eq. 77), and $\gamma_n^3(k, r)$ - Danielsson et al. estimator (eq. 8). The estimator $\gamma_n^1(k, r)$ was also independently introduced by Brilhante et al. (2013)

and Beran et al. (2014). With the appropriate choice of r, this estimator has a lower asymptotic variance compared to the classical Hill estimator.

In the subsequent work, Paulauskas and Vaičiulis (2017a) introduced a new general class of estimators.

$$g_{r,u}(x) = x^r \log^u(x),$$

$$G_n(k, r, u) = \frac{1}{k}\sum_{i=0}^{k-1} g_{r,u}\left(\frac{X_{(n-i)}}{X_{(n-k)}}\right)$$

They analysed a large number of estimators expressed in terms of statistics $G_n(k, r, u)$, and also introduced a couple of new estimators:

$$\hat{\gamma}_4 = \frac{2G_n(k, r, 1)}{2rG_n(k, r, 1) + 1 + \sqrt{4rG_n(k, r, 1) + 1}} \tag{91}$$

$$\hat{\gamma}_5 = \begin{cases} (rG_n(k, r, 1) - G_n(k, r, 0) + 1)(r^2 G_n(k, r, 1))^{-1} & \text{if } r \neq 0, \\ \hat{\gamma}^{MR}(k) & \text{if } r = 0, \end{cases}$$

were $\hat{\gamma}^{MR}(k)$ corresponds to the Danielsson et al. estimator (eq. 8) with $l = 1$.

8 Stable distribution tail index estimators

These tail-index estimators are developed under an assumption A8. For a review of classical methods for parameter estimation of specific stable distributions we refer to the paper of Mittnik and Rachev (1993). Below, we provide a couple of more general methods. Before we proceed, it is important to note that in case of the stable distribution tail indexes $0 < \alpha < 2$ correspond to heavy-tailed distributions with corresponding indexes; however, the case $\alpha = 2$ corresponds to the normal distribution.

8.1 De Haan and Resnick estimator

De Haan and Resnick (1980) introduced a simple estimator of the following form:

$$\hat{\gamma}_n^{HR}(k) = \frac{\log X_{(n)} - \log X_{(n-k+1)}}{\log k}. \tag{92}$$

The weakness of the estimator is that it converges to its limit in a very slow rate ($\log^{-1} n$).

8.2 Bacro and Brito

Bacro and Brito (1995) modified the De Haan and Resnick estimator by excluding a (small) number of the highest-order statistics:

$$\hat{\gamma}_n^{BB}(k) = -\frac{\log X_{(n-\lfloor \nu k \rfloor+1)} - \log X_{(n-k+1)}}{log\nu}, \quad 0 < \nu < 1. \tag{93}$$

ν is a fixed constant.

8.3 De Haan and Pereira estimator

De Haan and Pereira (1999) suggested another estimator based on the assumption A8 of the following form:

$$\hat{\beta}_n = \frac{kX^2_{(n-k)}}{\sum_{i=1}^{n-k} X^2_{(i)}} \tag{94}$$

Its interesting property is that it uses lower order statistics than $n-k$.

8.4 Fan estimator

Fan (2004) suggested an estimator based on permutations. Fist define $h(x_1,...,x_m) = (\log m)^{-1} log(\sum_{i=1}^m x_i)$.

$$\hat{\gamma} = \binom{n}{m}^{-1} \sum_{1\le i_1<...<i_m\le n} h(X_{i_1},...,X_{i_m}). \tag{95}$$

Summation is made over all combinations of observations. But, if permutations consume too much time, they can be changed by resampling. The interpretation of m differs from k in other methods. Its meaning is closer to a window of a kernel function. Nevertheless, the assumptions are similar $m \to \infty$ and $m = o(n^{1/2})$ as $n \to \infty$.

8.5 Meerschaert and Scheffer estimator

The properties of the growth rate of the logged sample second moment led to a simple estimator introduced by Meerschaert and Scheffler (1998). It has the following form:

$$\hat{\gamma} = \frac{\log_+ \sum_{i=1}^n (X_i - \bar{X}_n)^2}{2\log n}, \tag{96}$$

where $\log_+(x) = \max(\log x, 0)$. The estimator is consistent, asymptotically unbiased and asymptotically log stable if the data is in the domain of attraction of a stable law; however, the estimator is not scale-invariant. It is consistent for $\alpha \in (0,2]$.

8.6 Politis estimator

A similar method based on the divergence speed of logged sample second moment was introduced by Politis (2002). Define $S_j = j^{-1}\sum_{i=1}^k X_i^2$ and $Y_j = \log S_j$. Politis noticed that the tail index can be derived from the slope of Y_j regression on $\log j$. The estimator of α is designed as follows:

$$\hat{\mu} = \frac{\sum_{j=1}^n (Y_j - \bar{Y})(\log k - \overline{\log n})}{\sum_{j=1}^n (\log j - \overline{\log n})^2}.$$

$$\hat{\alpha} = \frac{2}{\hat{\mu} + 1}, \tag{97}$$

were $\bar{Y} = n^{-1} \sum_{j=1}^{k} Y_j$ and $\overline{\log n} = n^{-1} \sum_{j=1}^{n} \log j$. The estimator is consistent for $\alpha \in (0, 2]$. Similarly to the Meerschaert and Scheffer estimator, Politis estimator is not scale invariant. Also different permutations of the data may lead to different results. To deal with the last problem, the author suggests to apply the estimator for a number of permutations and to take the median value of the estimates. The method can be applied to time series.

8.7 McElroy and Politis estimators

McElroy and Politis (2007) modified the Meerschaert and Scheffer estimator (eq. 96) in the following way:

$$\hat{\gamma}_r = \frac{\log \sum_{i=1}^{n} X_i^{2r}}{2r \log n}, \tag{98}$$

where r is large enough, such that the $2r$-th moment does not exist. The bias of this estimator reduces slowly. Apart from this estimator, McElroy and Politis studied reduced bias estimators:

$$\hat{\gamma}_n^{CEN} = \frac{\log S_n(X^2) - \log S_{\sqrt{n}}(X^2)}{2 \log n}.$$

where $S_n(X^2) = \sum_{i=1}^{n} X_i^2$.

They also suggested to split a sample into M non-overlapping groups of size b^2 (b is supposed to be relatively small, such as $n^{1/3}$), and compute

$$\hat{\gamma}^{SCEN} = \frac{1}{M} \sum_{m=1}^{M} \hat{\gamma}_{b^2}^{CEN(m)}, \tag{99}$$

where $\hat{\gamma}_{b^2}^{CEN(m)}$ is defined as $\hat{\gamma}_{b^2}^{CEN}$ estimated on data points $\{(m-1)b^2 + 1, ..., mb^2\}$.

Alternatively, denote

$$S_d^{(j)}(X^2) = \sum_{i=(j-1)d+1}^{jd} X_i^2.$$

$$\hat{\gamma}_{b^2}^{RCEN} = \frac{1}{b} \sum_{j=1}^{b} \frac{\log S_{b^2}(X^2) - \log S_b^{(j)}(X^2)}{2 \log b}.$$

Also denote $\hat{\gamma}_{b^2}^{RCEN(m)}$ as $\hat{\gamma}_{b^2}^{RCEN}$ evaluated on data points $\{(m-1)b^2+1, ..., mb^2\}$. Then, another estimator is

$$\hat{\gamma}^{SRCEN} = \frac{1}{M} \sum_{m=1}^{M} \hat{\gamma}_{b^2}^{RCEN(m)}. \tag{100}$$

All enumerated estimators allow for time-series dependence in the data.

9 Small-sample and robust estimators

9.1 Knight estimator

Knight (2007) suggested a robust estimator of the following form. First, choose c, which represents the level of robustness. Next, solve for $\phi(c)$ from equation

$$\phi(c) + exp(-\{c + \phi(c)\}) = 1.$$

The estimator is defined as a solution for $\tilde{\alpha}$

$$\sum_{j=1}^{k_n} \psi_c(U_j, \tilde{\alpha}_n(c)) = 0, \tag{101}$$

where U_j is defined in equation (1) and

$$\psi_c(x, \alpha) = \begin{cases} x - \phi_c/\alpha & \text{if } x \leq (c + \phi(c))/\alpha \\ c/\alpha & \text{otherwise.} \end{cases}$$

If $c = \infty$, the proposed estimator simplifies to the MLE.

9.2 Beran and Schell M-estimator

Beran and Schell (2012) suggested a small-sample M-estimator:

$$\psi_v(x, \alpha) = max(\alpha \log(x) - 1, v) - (v + exp(-(v + 1)))$$

$$\sum_{i=1}^{n} \psi_v(X_i, \alpha) = 0, \tag{102}$$

the estimator is the value of α, which solves equation (102). v is a constant $v > -1$. Higher values of v lead to a larger degree of robustness and a larger bias, which does not vanish asymptotically. An application of a similar M-estimator to Pareto-tail index estimation was also discussed by Victoria-Feser and Ronchetti (1994); however, their algorithm is very sensible to the choice of starting values.

9.3 Dupuis and Victoria-Feser weighted MLE

Dupuis and Victoria-Feser (2006) adjusted a more general weighted MLE of Dupuis and Morgenthaler (2002) to tail index estimation. Its main idea is to analyse a quantile plot first. Next, MLE is applied giving lower weights to observations, which result in residuals of the qunatile plot regressions exceeding a threshold.

$$\sum_{i=1}^{k} w(X_{(n-i+1)}; \theta) \frac{\partial}{\partial \theta} \log f_\theta(X_i) = 0, \tag{103}$$

where $f_\theta(X_i)$ is the density function of the right tail of the distribution with parameters θ ($\alpha \in \theta$) and $w(X_{(n-i+1)};\theta)$ is a weight function:

$$w(X_{(n-i+1)};\theta) = \begin{cases} 1 & \text{if } |\hat{\epsilon}_i| < c \\ c/|\hat{\epsilon}_i| & \text{if } |\hat{\epsilon}_i| \geq c. \end{cases}$$

c, is a parameter which controls for robustness (lower c results in a higher level of robustness). $\hat{\epsilon}_i$ is a standardized residual of a quantile plot, i.e.: $\hat{\epsilon}_i = (Y_i - \hat{Y}_i)/\sigma_i$, $\sigma_i^2 = \sum_{j=1}^{i} 1/[\hat{\alpha}^2(k-i+j)^2]$, $Y_i = log(X_{(n-i+1)}/X_{(k)})$, and $\hat{Y}_i = -1/\hat{\alpha}\log[(k+1-i)/(k+1)]$. The resulting estimator is biased, but the bias can be estimated and removed. For more detail, we refer to lemma 5 in the above-mentioned paper.

10 Peaks Over Random Threshold (PORT) estimators

The notion of PORT estimators was introduced by Santos et al. (2006). The need for this class of estimators arises from the fact that most classical estimators are non-shift invariant. Its general idea is that instead of the original (ordered) sample of size n, $\boldsymbol{X} = \{X_{(1)}, ..., X_{(n)}\}$, a modified sample $\boldsymbol{X^{PORT}} = \{X_{(n-m+1)} - X_{(m)}, ..., X_{(n)} - X_{(m)}\}$ of $n-m$ largest observations is analyzed. The reduction of $X_{(m)}$ from the m largest observations removes the shift existing in the data, and $X_{(m)}$ serves as a random threshold. If a classical estimator for such a sample is applied, it becomes shift-invariant. Fraga Alves (2001) applied this idea to the Hill estimator (eq. 4). She suggested using such m that $(n-m(n)) = o(n)$, $(n-m(n)) \to \infty$ as $n \to \infty$, and $k = o(n-m(n))$, $k \to \infty$ as $n \to \infty$. Santos et al. (2006) suggested using $m = nq+1$, $0 < q < 1$. If the distribution function underlying the initial sample $\boldsymbol{X}$ has a finite left endpoint, also $q = 0$ can be applied.[3] Apart from the Hill estimator, they also applied this idea to the moment estimator (77). Similarly, Fraga Alves et al. (2009) introduced a PORT modification of their estimator (81) and Gomes et al. (2016) developed PORT methodology for mean-of-order-p class of estimators (85). One of the estimators studied by Gomes and Henriques-Rodrigues (2016) was a PORT version of the Caeiro et al. (2005) estimator (56). Li et al. (2008) developed PORT methodology for Caeiro end Gomes estimator (84).

11 Parametric methods

The peculiarity of the parametric methods is that they use an assumption that the entire sample is drawn from a distribution function with a specific functional form, while semi-parametric methods discussed above are less restrictive because they assume some regularity in the tail behaviour only.

[3] However, Gomes (2008) expressed a caution for using $q = 0$.

11.1 Weiss estimator

Weiss (1971) considered a class of probability densities of the form

$$f(x) = c(x-\theta)^{\alpha+1}[1 + r(x-\theta)],\ x \geq \theta; \quad f(x) = 0,\ x < \theta.$$

$C > 0$, $\alpha > 0$, and $|r(y)| \leq Ky^{\nu}$, for all y in some interval $[0, \Delta]$, where K, ν and Δ are positive but unknown constants. The estimator is:

$$\hat{\alpha} = \log 2 \left[\log \frac{X_{(k(n))} - X_{(1)}}{X_{(k(n)/2)} - X_{(1)}} \right]^{-1}. \tag{104}$$

It is assumed that $k(n)$ is an even integer. Assumption B2 is changed into a more restrictive one: $k(n)/n^{\rho} \to 0$ for all $\rho > 0$.

11.2 Brazauskas and Serfling estimator

Brazauskas and Serfling (2000) assumed an exact Pareto distribution as in assumption A1, with $L(x) = C$, where C is a known constant, $x \geq C$. They suggested using MLE for k arbitrary observations: $\hat{\gamma}(x_1, ..., x_k) = k^{-1} \sum_{j=1}^{k} \log x_j - \log \sigma$, where σ is a known scale parameter. Having calculated $\hat{\gamma}(x_1, ..., x_k)$ for all possible combinations of $X_{i_1}, ..., X_{i_k}$, they get a vector H of $n!/(k!(n-k)!)$ size of γ estimates. As a final estimator, they suggested using

$$\hat{\gamma} = median(H) \tag{105}$$

11.3 Finkelstein et al. estimator

Under the same assumptions as Brazauskas and Serfling, Finkelstein et al. (2006) proposed estimating parameter α by solving an equation

$$n^{-1} \sum_{i=1}^{n} \left(\frac{C}{X_i} \right)^{\hat{\alpha} t} = \frac{1}{t+1}, \quad t > 0 \tag{106}$$

for $\hat{\alpha}$. t is a tuning parameter.

The intuition of the method is as follows: $(C/X_j)^{\alpha}$ has a uniform distribution; therefore, if $\hat{\alpha}$ is relatively close to α, the left side of equation (106) is distributed like the arithmetic mean of uniform random variables raised to power t. Having substituted this mean by the mathematical expectation, which equals to $(1+t)^{-1}$, equation (106) is received. Brzezinski (2016) argued that this estimator has nice properties in a sense of compromise between ease of use and robustness against outliers in the small-sample setting.

11.4 McElroy parametric estimators

McElroy (2007) used the properties of $Var(\log |X|)$ for derivation of parametric tail-index estimators for a number of specific heavy-tailed distribution functions.

It was assumed that the data is mean zero, or its location parameter is zero. Denote $\hat{V} = \widehat{Var}(\log|X|)$. $\hat{V}$ is an empirical estimate of logged $|X|$ variance. Taking derivatives of $Var(\log|X|)$ around zero, he showed that, such a function $g(\cdot)$ exists that $g(\alpha) = Var(\log|X|)$. Hence, if $g^{-1}(\cdot)$ exist, $\hat{\alpha} = g^{-1}(\hat{V})$.

For stable distribution the tail index estimator can be expressed as

$$\hat{\alpha} = \frac{2}{\sqrt{1 + \frac{4\hat{V} - \Psi_2(1/2)}{\Psi_2(1)}}}, \tag{107}$$

were $\Psi_2(\cdot)$ denotes the second derivative of the log-gamma function. For Student's t-distribution the expression for $g(\cdot)$ is

$$g(\alpha) = \frac{1}{2}\big(\Psi_2(\alpha/2) + \Psi_2(1/2)\big). \tag{108}$$

There is no analytical expression for g^{-1} in this case, but α can be solved numerically. If the data is drawn from a log-gamma distribution,

$$\hat{\alpha} = \frac{\sqrt{6}}{\sqrt{\hat{V}\widehat{Kur}\log|X|}}, \tag{109}$$

were $\widehat{Kur}\log|X|$ is an empirical estimate of the kurtosis of the logged data.

If the data comes from a Pareto-like distribution, the resulting estimator is

$$\hat{\alpha} = \frac{1}{\sqrt{\hat{V} - \Psi_2(1/2)/4}}. \tag{110}$$

Estimator (110) allows for serially correlated data.

11.5 Hosking and Wallis estimators of the generalized Pareto distribution parameters

Hosking and Wallis (1987) analysed methods for parameters of the generalized Pareto distribution estimation. They proposed three methods: MLE, method of moments, and method of probability weighted moments. Their methods are based on the assumption that the entire sample is drawn from the GPD. We do not present the MLE because it may provide parameters resulting in arbitrary large values, and there is a need to search for a local maximum. In the above-mentioned paper, the Newton-Raphson algorithm failed to converge in 91 samples out of 100 random starting values. The method of moments results in the following γ estimator:

$$\gamma = 0.5(1 - (\bar{X})^2/s^2), \tag{111}$$

where $\bar{X}$ is the sample's mean and s^2 is its variance. The method works if $\gamma < 1/2$, i.e. the second moment is finite, and the normality is shown under the assumption of the fourth finite moment existence. The third estimator presented by Hosking and Wallis is the method of probability weighted moments. It has the following form:

$$\hat{\gamma}_{PWM} = \bar{X}/(2s - \bar{X}), \tag{112}$$

where $s = n^{-1}\sum_{i=1}^{n}(n-i)X_i/(n-1)$.

Dupuis and Tsao (1998) made a modification of this estimator for negative γ, to exclude cases, when the right endpoint of the distribution is estimated to be lower than the largest observation.

11.6 Zhang estimators

Zhang (2007) created a mix of MLE and moment estimator for the GPD. First, equation for b is solved:

$$n^{-1}\sum_{i=1}^{n}(1-bX_i)^p - (1-r)^{-1} = 0, \quad b < X_{(n)}^{-1},$$

where $p = rn/\sum_{i=1}\log(1-bX_i)$ and r is a tuning parameter $r < 1$. It is also assumed that $r\gamma < 1$. When $r = -\gamma$, the method becomes a pure MLE. The estimator of $\hat{\gamma}$ is given by

$$\hat{\gamma} = n^{-1}\sum_{i=1}^{n}\log(1-\hat{b}X_i). \tag{113}$$

In the subsequent papers, Zhang and Stephens (2009) presented a Bayesian modification of this method, and in 2010 improved the Bayesian method for very heavy-tailed ($\hat{\gamma} > 1$) distributions.

11.7 Wang and Chen estimator

Wang and Chen (2016) introduced another hybrid method for the GPD parameter estimation. They suggested minimizing $G(b)$

$$G(b) = -n^{-1}\sum_{i=1}^{n}\Big\{(2i-1)\log[g_i(b)] + (2n+1-2i)\log[1-g_i(b)]\Big\}. \tag{114}$$

$$g_i(b) = 1 - (1-bX_{(i)})^{-n/\sum_{j=1}^{n}\log(1-\theta X_j)}, \quad i = 1, ..., n.$$

with respect to b. Next, $\hat{b}$ it is plugged into equation (113).

11.8 Van Zyl estimator

Van Zyl (2015) suggested normalizing observations before estimating the tail index. Namely, $Z_i = \mu\big[(\gamma/\sigma)X_i + (1-\gamma\mu/\sigma)\big]$, where μ, σ and γ are the location, scale and shape parameters of the GPD. The preliminary values of these parameters can be received using MLE or other methods. For example, $\hat{\gamma}$

can be estimated with the Hill estimator (eq. 4), $\hat{\mu} = X_{(1)}$. $\hat{\sigma}$ can be estimated numerically (see the above-mentioned paper, pages 173-174).

$$\hat{\gamma}_Z = \frac{1}{n} \sum_{i=0}^{n-1} \log \left[(\hat{\gamma}/\hat{\sigma}) X_{n-i} + 1 - \hat{\gamma}\hat{\mu}/\hat{\sigma} \right]. \tag{115}$$

The method works under assumptions that $(\hat{\gamma}/\hat{\sigma}) X_{n-i} + 1 - \hat{\gamma}\hat{\mu}/\hat{\sigma} > 0$ and $0 < \gamma < 1$.

Acknowledgments

I would like to thank Remigijus Leipus and Anton Skrobotov for their help in finding the literature. Also I am grateful to Tucker Sprague McElroy, Constantinos Petropoulos, Martin Van Zyl and Dimitris N. Politis for consulting me about their methods. Their advices helped me to avoid typos, which are present in the original papers.

References

Aban, I. B. and M. M. Meerschaert (2001). Shifted Hill's estimator for heavy tails. *Communications in Statistics-Simulation and Computation 30*(4), 949–962.

Aban, I. B. and M. M. Meerschaert (2004). Generalized least-squares estimators for the thickness of heavy tails. *Journal of Statistical Planning and Inference 119*(2), 341–352.

Agterberg, F. (1995). Multifractal modeling of the sizes and grades of giant and supergiant deposits. *International Geology Review 37*(1), 1–8.

Axtell, R. L. (2001). Zipf distribution of US firm sizes. *Science 293*(5536), 1818–1820.

Bacro, J. N. and M. Brito (1995). Weak limiting behaviour of a simple tail Pareto-index estimator. *Journal of Statistical Planning and inference 45*(1-2), 7–19.

Baek, C. and V. Pipiras (2010). Estimation of parameters in heavy-tailed distribution when its second order tail parameter is known. *Journal of Statistical Planning and Inference 140*(7), 1957–1967.

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

Beirlant, J., G. Dierckx, and A. Guillou (2005). Estimation of the extreme-value index and generalized quantile plots. *Bernoulli 11*(6), 949–970.

Beirlant, J., G. Dierckx, A. Guillou, and C. Staăricaă (2002). On exponential representations of log-spacings of extreme order statistics. *Extremes 5*(2), 157–180.

Beirlant, J., F. Figueiredo, M. I. Gomes, and B. Vandewalle (2008). Improved reduced-bias tail index and quantile estimators. *Journal of Statistical Planning and Inference 138*(6), 1851–1870.

Beirlant, J. and A. Guillou (2001). Pareto index estimation under moderate right censoring. *Scandinavian Actuarial Journal 2001*(2), 111–125.

Beirlant, J., A. Guillou, G. Dierckx, and A. Fils-Villetard (2007). Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes 10*(3), 151–174.

Beirlant, J. and J. L. Teugels (1989). Asymptotic normality of Hill's estimator. In *Extreme value theory*, pp. 148–155. Springer.

Beirlant, J., P. Vynckier, and J. L. Teugels (1996a). Excess functions and estimation of the extreme-value index. *Bernoulli 2*(4), 293–318.

Beirlant, J., P. Vynckier, and J. L. Teugels (1996b). Tail index estimation, Pareto quantile plots regression diagnostics. *Journal of the American statistical Association 91*(436), 1659–1667.

Benhabib, J., A. Bisin, and M. Luo (2017). Earnings inequality and other determinants of wealth inequality. *American Economic Review 107*(5), 593–97.

Benhabib, J., A. Bisin, and S. Zhu (2011). The distribution of wealth and fiscal policy in economies with finitely lived agents. *Econometrica 79*(1), 123–157.

Beran, J. and D. Schell (2012). On robust tail index estimation. *Computational Statistics & Data Analysis 56*(11), 3430–3443.

Beran, J., D. Schell, and M. Stehlík (2014). The harmonic moment tail index estimator: asymptotic distribution and robustness. *Annals of the Institute of Statistical Mathematics 66*(1), 193–220.

Brazauskas, V. and R. Serfling (2000). Robust and efficient estimation of the tail index of a single-parameter Pareto distribution. *North American Actuarial Journal 4*(4), 12–27.

Brilhante, M.F., G.-M. P. D. (2014). The mean-of-order p extreme value index estimator revisited. In *New Advances in Statistical Modeling and Application*, pp. 163–175. Berlin.

Brilhante, M. F., M. I. Gomes, and D. Pestana (2013). A simple generalisation of the Hill estimator. *Computational Statistics & Data Analysis 57*(1), 518–535.

Brito, M., L. Cavalcante, and A. C. M. Freitas (2016). Bias-corrected geometric-type estimators of the tail index. *Journal of Physics A: Mathematical and Theoretical 49*(21), 214003.

Brito, M. and A. C. M. Freitas (2003). Limiting behaviour of a geometric-type estimator for tail indices. *Insurance: Mathematics and Economics 33*(2), 211–226.

Brzezinski, M. (2016). Robust estimation of the Pareto tail index: a Monte Carlo analysis. *Empirical Economics 51*(1), 1–30.

Caeiro, F. and M. I. Gomes (2002). A class of asymptotically unbiased semi-parametric estimators of the tail index. *Test 11*(2), 345–364.

Caeiro, F. and M. I. Gomes (2006). A new class of estimators of a scale second order parameter. *Extremes 9*(3-4), 193–211.

Caeiro, F., M. I. Gomes, J. Beirlant, and T. de Wet (2016). Mean-of-order p reduced-bias extreme value index estimation under a third-order framework. *Extremes 19*(4), 561–589.

Caeiro, F., M. I. Gomes, and D. Pestana (2005). Direct reduction of bias of the classical Hill estimator. *Revstat 3*(2), 113–136.

Chaney, T. (2008). Distorted gravity: the intensive and extensive margins of international trade. *American Economic Review 98*(4), 1707–21.

Ciuperca, G. and C. Mercadier (2010). Semi-parametric estimation for heavy tailed distributions. *Extremes 13*(1), 55–87.

Cowell, F. A. and E. Flachaire (2007). Income distribution and inequality measurement: The problem of extreme values. *Journal of Econometrics 141*(2), 1044–1072.

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

Csörgő, S. and L. Viharos (1998). Estimating the tail index. In *Asymptotic Methods in Probability and Statistics*, pp. 833–881. Elsevier.

Csörgoő, S. and D. M. Mason (1985). Central limit theorems for sums of extreme values. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 98, pp. 547–558. Cambridge University Press.

Danielsson, J., D. W. Jansen, and C. G. De Vries (1996). The method of moments ratio estimator for the tail shape parameter. *Communications in Statistics-Theory and Methods 25*(4), 711–720.

Das, K. P. and S. C. Halder (2016). Understanding extreme stock trading volume by generalized Pareto distribution. *The North Carolina Journal of Mathematics and Statistics 2*, 45–60.

Davydov, Y., V. Paulauskas, and A. Račkauskas (2000). More on p-stable convex sets in Banach spaces. *Journal of Theoretical Probability 13*(1), 39–64.

De Haan, L. and T. T. Pereira (1999). Estimating the index of a stable distribution. *Statistics & Probability Letters 41*(1), 39–55.

De Haan, L. and S. Resnick (1998). On asymptotic normality of the Hill estimator. *Stochastic Models 14*(4), 849–866.

De Haan, L. and S. I. Resnick (1980). A simple asymptotic estimate for the index of a stable distribution. *Journal of the Royal Statistical Society. Series B 42*(1), 83–87.

De Haan, L. d. and L. Peng (1998). Comparison of tail index estimators. *Statistica Neerlandica 52*(1), 60–70.

De Haan, L. F. M. (1970). *On regular variation and its application to the weak convergence of sample extremes*. Mathematisch Centrum, Amsterdam.

Deheuvels, P., E. Haeusler, and D. M. Mason (1988). Almost sure convergence of the Hill estimator. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 104, pp. 371–381.

Dekkers, A. L., J. H. Einmahl, and L. De Haan (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics 17*(4), 1833–1855.

Drees, H. (1995). Refined Pickands estimators of the extreme value index. *The Annals of Statistics 23*(6), 2059–2080.

Drees, H. (1996). Refined Pickands estimators with bias correction. *Communications in Statistics-Theory and Methods 25*(4), 837–851.

Drees, H. (1998a). A general class of estimators of the extreme value index. *Journal of Statistical Planning and Inference 66*(1), 95–112.

Drees, H. (1998b). On smooth statistical tail functionals. *Scandinavian Journal of Statistics 25*(1), 187–210.

Dupuis, D. and M. Tsao (1998). A hybrid estimator for generalized Pareto and extreme-value distributions. *Communications in Statistics-Theory and Methods 27*(4), 925–941.

Dupuis, D. J. and S. Morgenthaler (2002). Robust weighted likelihood estimators with an application to bivariate extreme value problems. *Canadian Journal of Statistics 30*(1), 17–36.

Dupuis, D. J. and M.-P. Victoria-Feser (2006). A robust prediction error criterion for Pareto modelling of upper tails. *Canadian Journal of Statistics 34*(4), 639–658.

Falk, M. (1994). Efficiency of convex combinations of Pickands estimator of the extreme value index. *Journal of Nonparametric Statistics 4*(2), 133–147.

Fan, Z. (2004). Estimation problems for distributions with heavy tails. *Journal of Statistical Planning and Inference 123*(1), 13–40.

Ferriere, R. and B. Cazelles (1999). Universal power laws govern intermittent rarity in communities of interacting species. *Ecology 80*(5), 1505–1521.

Feuerverger, A. and P. Hall (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *The Annals of Statistics 27*(2), 760–781.

Fialová, A., J. Jurecková, and J. Picek (2004). Estimating Pareto tail index based on sample means. *REVSTAT–Statistical Journal 2*(1), 75–100.

Finkelstein, M., H. G. Tucker, and J. Alan Veeh (2006). Pareto tail index estimation revisited. *North American Actuarial Journal 10*(1), 1–10.

Fraga Alves, M. (1995). Estimation of the tail parameter in the domain of attraction of an extremal distribution. *Journal of Statistical Planning and Inference 45*(1-2), 143–173.

Fraga Alves, M. (2001). A location invariant Hill-type estimator. *Extremes 4*(3), 199–217.

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

Fraga Alves, M. I., M. I. Gomes, L. de Haan, and C. Neves (2009). Mixed moment estimator and location invariant alternatives. *Extremes 12*(2), 149–185.

Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics 1*(1), 255–294.

Gabaix, X., P. Gopikrishnan, V. Plerou, and H. E. Stanley (2003). A theory of power-law distributions in financial market fluctuations. *Nature 423*(6937), 267.

Gabaix, X. and R. Ibragimov (2011). Rank- 1/2: a simple way to improve the OLS estimation of tail exponents. *Journal of Business & Economic Statistics 29*(1), 24–39.

Gabaix, X. and A. Landier (2008). Why has CEO pay increased so much? *The Quarterly Journal of Economics 123*(1), 49–100.

Gastwirth, J. L. (1972). The estimation of the Lorenz curve and Gini index. *The Review of Economics and Statistics 54*, 306–316.

Gomes, M. I., M. I. F. Alves, and P. A. Santos (2008). PORT Hill and moment estimators for heavy-tailed models. *Communications in Statistics - Simulation and Computation 37*(7), 1281–1306.

Gomes, M. I., F. Caeiro, and F. Figueiredo (2004). Bias reduction of a tail index estimator through an external estimation of the second-order parameter. *Statistics 38*(6), 497–510.

Gomes, M. I., F. Figueiredo, and S. Mendonça (2005). Asymptotically best linear unbiased tail estimators under a second-order regular variation condition. *Journal of Statistical Planning and Inference 134*(2), 409–433.

Gomes, M. I. and A. Guillou (2015). Extreme value theory and statistics of univariate extremes: A review. *International Statistical Review 83*(2), 263–292.

Gomes, M. I. and L. Henriques-Rodrigues (2016). Competitive estimation of the extreme value index. *Statistics & Probability Letters 117*, 128–135.

Gomes, M. I., L. Henriques-Rodrigues, and B. Manjunath (2016). Mean-of-order-p location-invariant extreme value index estimation. *Revstat 14*(3), 273–296.

Gomes, M. I. and M. J. Martins (2001). Generalizations of the Hill estimator–asymptotic versus finite sample behaviour. *Journal of Statistical Planning and Inference 93*(1-2), 161–180.

Gomes, M. I. and M. J. Martins (2002). Asymptotically unbiased estimators of the tail index based on external estimation of the second order parameter. *Extremes 5*(1), 5–31.

Gomes, M. I. and M. J. Martins (2004). Bias reduction and explicit semi-parametric estimation of the tail index. *Journal of Statistical Planning and Inference 124*(2), 361–378.

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

Gomes, M. I., M. J. Martins, and M. Neves (2002). Generalized Jackknife semi-parametric estimators of the tail index. *Portugaliae Mathematica 59*(4), 393–408.

Gomes, M. I., M. J. Martins, and M. Neves (2007). Improving second order reduced bias extreme value index estimation. *Revstat 5*(2), 177–207.

Gomes, M. I., C. Miranda, and C. Viseu (2007). Reduced-bias tail index estimation and the Jackknife methodology. *Statistica Neerlandica 61*(2), 243–270.

Gomes, M. I. and O. Oliveira (2003). Maximum likelihood revisited under a semi-parametric context-estimation of the tail index. *Journal of Statistical Computation and Simulation 73*(4), 285–301.

Gomes, M. I., H. Pereira, and M. C. Miranda (2005). Revisiting the role of the Jackknife methodology in the estimation of a positive tail index. *Communications in Statistics-Theory and Methods 34*(2), 319–335.

Groeneboom, P., H. Lopuhaä, and P. De Wolf (2003). Kernel-type estimators for the extreme value index. *The Annals of Statistics 31*(6), 1956–1995.

Haeusler, E. and J. L. Teugels (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation. *The Annals of Statistics 13*(1), 743–756.

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society. Series B (Methodological) 44*(1), 37–42.

Hall, P. and A. Welsh (1985). Adaptive estimates of parameters of regular variation. *The Annals of Statistics 13*, 331–341.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics 3*(5), 1163–1174.

Hinloopen, J. and C. Van Marrewijk (2012). Power laws and comparative advantage. *Applied Economics 44*(12), 1483–1507.

Hosking, J. R. M. and J. R. Wallis (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics 29*(3), 339–349.

Huebner, F., D. Liu, and J. Fernandez (1998). Queueing performance comparison of traffic models for internet traffic. In *Global Telecommunications Conference, 1998. GLOBECOM 1998. The Bridge to Global Integration. IEEE*, Volume 1, pp. 471–476. IEEE.

Huisman, R., K. G. Koedijk, C. J. M. Kool, and F. Palm (2001). Tail-index estimates in small samples. *Journal of Business & Economic Statistics 19*(2), 208–216.

Hüsler, J., D. Li, and S. Müller (2006). Weighted least squares estimation of the extreme value index. *Statistics & Probability Letters 76*(9), 920–930.

Jurečková, J. (2000). Test of tails based on extreme regression quantiles. *Statistics & Probability Letters 49*(1), 53–61.

Jurečková, J. and J. Picek (2001). A class of tests on the tail index. *Extremes 4*(2), 165–183.

Jurečková, J. and J. Picek (2004). Estimates of the tail index based on nonparametric tests. In *Theory and Applications of Recent Robust Methods*, pp. 141–152. Springer.

Kaizoji, T. (2003). Scaling behavior in land markets. *Physica A: Statistical Mechanics and its Applications 326*(1-2), 256–264.

Kang, S. and J. Song (2017). Parameter and quantile estimation for the generalized Pareto distribution in peaks over threshold framework. *Journal of the Korean Statistical Society 46*(4), 487–501.

Knight, K. (2007). A simple modification of the Hill estimator with applications to robustness and bias reduction. Unpublished paper: Statistics Department, University of Toronto.

Kratz, M. and S. I. Resnick (1996). The qq-estimator and heavy tails. *Stochastic Models 12*(4), 699–724.

LePage, R., M. Woodroofe, and J. Zinn (1981). Convergence to a stable distribution via order statistics. *The Annals of Probability 9*(4), 624–632.

Li, J., Z. Peng, and S. Nadarajah (2008). A class of unbiased location invariant Hill-type estimators for heavy tailed distributions. *Electronic Journal of Statistics 2*, 829–847.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business 36*, 394–419.

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

McElroy, T. (2007). Tail index estimation for parametric families using log moments. Research Report Series (Statistics Nr. 2007-2).

McElroy, T. and D. N. Politis (2007). Moment-based tail index estimation. *Journal of Statistical Planning and Inference 137*(4), 1389–1406.

Meerschaert, M. M. and H.-P. Scheffler (1998). A simple robust estimation method for the thickness of heavy tails. *Journal of Statistical Planning and Inference 71*(1), 19–34.

Mittnik, S. and S. T. Rachev (1993). Modeling asset returns with alternative stable distributions. *Econometric Reviews 12*(3), 261–330.

Müller, S. and K. Rufibach (2009). Smooth tail-index estimation. *Journal of Statistical Computation and Simulation 79*(9), 1155–1167.

Müller, U. K. and Y. Wang (2017). Fixed-k asymptotic inference about tail properties. *Journal of the American Statistical Association 112*(519), 1334–1343.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications 9*(1), 141–142.

Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary physics 46*(5), 323–351.

Nordhaus, W. D. (2012). Economic policy in the face of severe tail events. *Journal of Public Economic Theory 14*(2), 197–219.

Nuyts, J. (2010). Inference about the tail of a distribution: Improvement on the Hill estimator. *International Journal of Mathematics and Mathematical Sciences 2010*, 1–16.

Ogwang, T. (2011). Power laws in top wealth distributions: Evidence from Canada. *Empirical Economics 41*(2), 473–486.

Pareto, V. (1897). Cours de'conomie politique. Lausanne, Rouge.

Paulauskas, V. (2003). A new estimator for a tail index. *Acta Applicandae Mathematica 79*, 55–67.

Paulauskas, V. and M. Vaičiulis (2011). Several modifications of DPR estimator of the tail index. *Lithuanian Mathematical Journal 51*(1), 36–50.

Paulauskas, V. and M. Vaičiulis (2012). Estimation of the tail index in the max-aggregation scheme. *Lithuanian Mathematical Journal 52*(3), 297–315.

Paulauskas, V. and M. Vaičiulis (2013). On an improvement of Hill and some other estimators. *Lithuanian Mathematical Journal 53*(3), 336–355.

Paulauskas, V. and M. Vaičiulis (2017a). A class of new tail index estimators. *Annals of the Institute of Statistical Mathematics 69*(2), 461–487.

Paulauskas, V. and M. Vaičiulis (2017b). Comparison of the several parameterized estimators for the positive extreme value index. *Journal of Statistical Computation and Simulation 87*(7), 1342–1362.

Peng, L. (1998). Asymptotically unbiased estimators for the extreme-value index. *Statistics & Probability Letters 38*(2), 107–115.

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

Pisarenko, V. and D. Sornette (2003). Characterization of the frequency of extreme earthquake events by the generalized Pareto distribution. *Pure and Applied Geophysics 160*(12), 2343–2364.

Politis, D. N. (2002). A new approach on estimation of the tail index. *Comptes Rendus Mathematique 335*(3), 279–282.

Qi, Y. (2010). On the tail index of a heavy tailed distribution. *Annals of the Institute of Statistical Mathematics 62*(2), 277–298.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika 43*(3/4), 353–360.

Rosen, K. T. and M. Resnick (1980). The size distribution of cities: an examination of the Pareto law and primacy. *Journal of Urban Economics 8*(2), 165–186.

Santos, P. A., M. Alves, and M. I. Gomes (2006). Peaks over random threshold methodology for tail index and high quantile estimation. *Revstat*, 227–247.

Schultze, J. and J. Steinebach (1996). On least squares estimates of an exponential tail coefficient. *Statistics & Risk Modeling 14*(4), 353–372.

Seekell, D. A. and M. L. Pace (2011). Does the Pareto distribution adequately describe the size-distribution of lakes? *Limnology and Oceanography 56*(1), 350–356.

Segers, J. (2001). Residual estimators. *Journal of Statistical Planning and Inference 98*(1-2), 15–27.

Segers, J. (2005). Generalized Pickands estimators for the extreme value index. *Journal of Statistical Planning and Inference 128*(2), 381–396.

Simon, H. A. and C. P. Bonini (1958). The size distribution of business firms. *The American Economic Review 48*(4), 607–617.

Smith, R. L. (1987). Estimating tails of probability distributions. *The Annals of Statistics 15*, 1174–1207.

Soo, K. T. (2005). Zipf's law for cities: A cross-country investigation. *Regional science and urban Economics 35*(3), 239–263.

Toda, A. A. (2012). The double power law in income distribution: Explanations and evidence. *Journal of Economic Behavior & Organization 84*(1), 364–381.

Toda, A. A. and K. Walsh (2015). The double power law in consumption and implications for testing euler equations. *Journal of Political Economy 123*(5), 1177–1200.

Tripathi, Y. M., S. Kumar, and C. Petropoulos (2014). Improved estimators for parameters of a Pareto distribution with a restricted scale. *Statistical Methodology 18*, 1–13.

Vaičiulis, M. (2009). An estimator of the tail index based on increment ratio statistics. *Lithuanian Mathematical Journal 49*(2), 222–233.

Vaičiulis, M. (2012). Asymptotic properties of generalized DPR statistic. *Lithuanian Mathematical Journal 52*(1), 95–110.

van Zyl, J. M. (2015). Estimation of the shape parameter of a generalized Pareto distribution based on a transformation to Pareto distributed variables. *Journal of Statistical Theory and Practice 9*(1), 171–183.

Vandewalle, B., J. Beirlant, A. Christmann, and M. Hubert (2007). A robust estimator for the tail index of Pareto-type distributions. *Computational Statistics & Data Analysis 51*(12), 6252–6268.

Victoria-Feser, M.-P. and E. Ronchetti (1994). Robust methods for personal-income distribution models. *Canadian Journal of Statistics 22*(2), 247–258.

Viharos, L. (1999). Weighted least-squares estimators of tail indices. *Probability and Mathematical Statistics 19*(2), 249–265.

Wang, C. and G. Chen (2016). A new hybrid estimation method for the generalized Pareto distribution. *Communications in Statistics-Theory and Methods 45*(14), 4285–4294.

Weiss, L. (1971). Asymptotic inference about a density function at an end of its range. *Naval Research Logistics 18*(1), 111–114.

Yun, S. (2000). A class of Pickands-type estimators for the extreme value index. *Journal of Statistical Planning and Inference 83*(1), 113–124.

Yun, S. (2002). On a generalized Pickands estimator of the extreme value index. *Journal of Statistical Planning and Inference 102*(2), 389–409.

Zhang, J. (2007). Likelihood moment estimation for the generalized Pareto distribution. *Australian & New Zealand Journal of Statistics 49*(1), 69–77.

Zhang, J. (2010). Improving on estimation for the generalized Pareto distribution. *Technometrics 52*(3), 335–339.

Zhang, J. and M. A. Stephens (2009). A new and efficient estimation method for the generalized Pareto distribution. *Technometrics 51*(3), 316–325.

Zhong, J. and X. Zhao (2012). Modeling complicated behavior of stock prices using discrete self-excited multifractal process. *Systems Engineering Procedia 3*, 110–118.

Zipf, G. K. (1941). *National unity and disunity; the nation as a bio-social organism.* Bloomington, Indiana: Principia Press.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology.* Addison-Wesley, Reading, Massachusetts.

Appendix

Table 1: Notation and abbreviations

Symbols	Explanations
X_i	i-th observation
$X_{(i)}$	i-th order statistics
n	Sample size
k	Number of largest observations under analysis
α	Tail index
γ	Extreme value index $\gamma = \alpha^{-1}$ (EVI)
$L(x)$	Slowly varying function, $\lim_{t\to\infty} L(t)/L(tx) \to 1$ as $t \to \infty$, $\forall\ x > 0$
$\mathbb{1}(u)$	Unit indicator function, $\mathbb{1}(u) = 1$ if u is TRUE, 0 otherwise.
$K(\cdot)$	Kernel function
ξ	Bandwidth.
$\lfloor u \rfloor$	the integer part of u.
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimator
GPD	Generalized Pareto distribution