*STATISTICA, anno LXXX, n. 3, 2020*

# A REVIEW OF MORE THAN ONE HUNDRED PARETO-TAIL INDEX ESTIMATORS

Igor Fedotenkov [1]
*DG Joint Research Center, European Commission, Brussels, Belgium*
*RANEPA Institute for Industrial Economics, Russian Presidential Academy of National Economy and Public Administration, Moscow, Russia*

## 1. INTRODUCTION

Since the seminal work of Pareto (Pareto, 1897), the classical application of heavy-tailed distributions in economics has regarded allocations of income and wealth (Gastwirth, 1972; Cowell and Flachaire, 2007; Ogwang, 2011; Benhabib *et al.*, 2011; Toda, 2012; Benhabib *et al.*, 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 *et al.*, 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 *et al.*, 1998), seismology (Pisarenko and Sornette, 2003; Nordhaus, 2012) and many other fields (Newman, 2005).

Most statistical and econometric methods are based on laws of large numbers and central limit theorems. Kolmogorov's law of large numbers requires the existence of a first finite moment. Lyapunov's version of the central limit theorem assumes the existence of a finite moment of an order higher than two. If data comes from a heavy-tailed distribution, these assumptions may not 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

[1] Corresponding Author. E-mail: i.fedotenkov@gmail.com

one of the key problems in statistics. Many methods have been proposed and numerous modifications of existing methods made. The goal of this paper is to review the current tail-index estimators.

There are numerous tail-index estimators. They are based on various assumptions and have diverse asymptotic and finite-sample properties. Unfortunately, the literature in this field is unstructured. Hence, 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 when searching the literature and are not always aware of many developments in this field. As a consequence, a few similar estimators were derived independently by different authors. The goal of our paper is to correct this shortcoming. 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 - corrections okayed by the authors.[2] The paper also aims to provide nontechnical explanations of methods, so as to be understood by researchers with intermediate skills in statistics. The paper can be considered a reference source on tail index estimators for researchers from various fields of science.

An excellent review of the advantages in the extreme value theory was done 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 (2015); however, the most widely-known estimators are also briefly discussed in order 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 the asymptotic properties of several reduced-bias tail-index estimators. A few works performed Monte Carlo simulations in order to compare the 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 (less than ten). In our work, we perform simple Monte-Carlo simulations for more than ninety of the estimators reviewed in this paper. Computer codes for the estimators, written in R, are readily available on the author's web page.

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

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

### 1.1. Notation

Before proceeding, we introduce the notation used throughout the paper.

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

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

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

where $X_{(1)},...,X_{(n)}$ are the order statistics of the sample, $k$ denotes the number of largest observations treated as the tail, $j$ denotes an index, $1 \leq j \leq k$. A table with notations is presented in Appendix A.

### 1.2. Assumptions and k selection

A common assumption is that observations are i.i.d. However, a number of estimators reviewed here also allow for dependent observations. Many estimators are based on the 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 $\alpha$ 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 $\gamma$ (instead of $\alpha$) estimation. Sometimes $\gamma$ is also allowed to take on negative values. Note that even if $\hat{\gamma}$ is an unbiased estimator of $\gamma$, $\hat{\gamma}^{-1}$ is a biased estimator of $\alpha$.

Hall (1982) introduced the following class of distribution tails:

- A2: $1-F(x) = 1-F(x) = Cx^{-\alpha}[1+D_1x^{-\alpha}+...+D_mx^{-m\alpha}+o(x^{-m\alpha})]$ 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 $\rho$ 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, in assumption A4 this case is often omitted.

Sometimes the second-order parameter $\rho$ 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 allowing for positive $\alpha$ (and $\gamma$) 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 $\gamma$ 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 $\sigma$, $\sigma > 0$ is a scale parameter, and $\mu$, $\mu \in (-\infty, \infty)$ is a location parameter. There are many parametric methods which assume that an entire sample is drawn from a GPD. A GPD can also, however, 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 $a_n > 0$, $b_n$, $n = 1, 2, \ldots$ and $\gamma \in \mathbb{R}$,

  $\lim_{n\to\infty} P\left\{\frac{max(X_1, X_2, \ldots, X_n) - b_n}{a_n} \leq x\right\} = \exp\left(-(1+\gamma x)^{-1/\gamma}\right),\ 1 + \gamma x > 0.$

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

Although assumptions A1-A8 seem different, the parameter $\alpha$ ($\gamma$) 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: $k(n) = o(n)$ as $n \to \infty$.

If the methods use different assumptions, we specify them in the text. Otherwise, we suppose that B1-B2 hold.

The question of $k$ selection, which minimizes asymptotical mean-squared errors, was addressed in detail by Hall (1982). Under assumption A3, he showed that the optimal $k$ is proportional to $n^{-2\rho/(\alpha-2\rho)}$ for the Hill estimator. If the chosen $k$ is too high, the variance of the estimator increases. If $k$ is too low, the bias of the estimator grows. De Haan and Peng (1998) found optimal $k$ under assumption A4 for four popular estimators; however, the resulting expressions are rather complicated for their direct practical use. We refer to the work of Németh and Zempléni (2017) for a concise summary of a few 'practical' methods, which can be easily implemented numerically. A few graphical methods are also available (Drees *et al.*, 2000). It must be mentioned that nowadays attempts to dispense with assumptions B1-B2 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 approximating the $k$-th largest observations with a Pareto distribution as in assumption A1 and used a maximum likelihood estimator (MLE) for $\gamma$ estimation.

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

Although the Hill estimator is MLE, it is classified as a semi-parametric method because the Pareto distribution is only assumed for the limiting behavior of the tail.

Mason (1982) showed weak consistency of the Hill estimator under assumptions A1, B1 and B2. For strong consistency an additional technical assumption is also required (Deheuvels *et al.*, 1988). Its asymptotic normality was analysed in detail by Hall (1982); Haeusler and Teugels (1985); Csorgo and Mason (1985); Beirlant and Teugels (1989); De Haan and Resnick (1998) and others.

### 2.2. *Kernel estimator*

Csorgo *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 non-increasing 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 (see Eq. (87)), where $\mathbb{1}$ is a unit indicator function.

### 2.3. *Fraga Alves estimator*

Fraga Alves (1995) developed a simple estimator for positive $\gamma$. It uses fewer observations than the Hill estimator and its computation is faster.

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

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

### 2.4. *Aban & Meerschaert shifted Hill's estimator*

As the 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 $\alpha$ 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 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 Eq. (3). $l$ is a tuning parameter. Often $l$ is equalised to 1.

### 2.6. *Nuyts estimator*

Nuyts (2010) suggested using Simpson's rule to improve the Hill estimator. The tail index $\alpha$ is estimated numerically as a solution to the following equation:

$$\frac{1}{k}\sum_{i=0}^{k-1}\log\left(X_{(n-i)}\right)=\frac{1}{\alpha}+\frac{\log X_{(k+1)}X_{(k+1)}^{-\alpha}-\log X_{(n)}X_{(n)}^{-\alpha}}{X_{(k+1)}^{-\alpha}-X_{(n)}^{-\alpha}}. \tag{9}$$

### 2.7. *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 $K$, $\nu$ and $\Delta$ are positive but unknown constants. The estimator is:

$$\hat{\alpha}=\log 2\left[\log\frac{X_{(k(n))}-X_{(1)}}{X_{(k(n)/2)}-X_{(1)}}\right]^{-1}. \tag{10}$$

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

### 2.8. *Hall class of estimators*

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

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

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

The class of estimators is defined as

$$\hat{\alpha}=\left(\sum_{j=1}^{m+1}b_j\hat{\gamma}_n^H(\lfloor a_j k\rfloor)\right)^{-1}, \tag{11}$$

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

## 3. Quantile plots

First, quantile plots for $\gamma>0$ are discussed, in this section. Second, 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); Beirlant *et al.* (1996b) studied this method from various perspectives. Examine a scatter-plot with coordinates $\left(-\log\left(j/(n+1)\right), \log X_{(n-j+1)}\right)$, $j = 1, \dots, n$. If the right side of this plot is 'almost' linear (supposing, this is so for the last $k$ observations) its slope corresponds to $\gamma$, and it can be estimated by applying the ordinary least squares (OLS) with an intercept:

$$\hat{\gamma}_1 = \frac{\sum_{j=1}^{k} \log\left((k+1)/j\right) \log X_{(n-j+1)} - k^{-1} \sum_{j=1}^{k} \log\left((k+1)/j\right) \sum_{j=1}^{k} \log X_{(n-j+1)}}{\sum_{j=1}^{k} \log^2\left((k+1)/j\right) - k^{-1}\left(\sum_{j=1}^{k} \log\left((k+1)/j\right)\right)^2}. \tag{12}$$

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

### 3.2. Schultze and Steinebach estimators

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

$$\hat{\gamma}_2 = \frac{\sum_{j=1}^{k} \log(n/j) \log X_{(n-j+1)}}{\sum_{j=1}^{k} \log^2(n/j)}. \tag{13}$$

The other is similar to the estimator (12), but with reversed dependent and explanatory variables:

$$\hat{\gamma}_3 = \frac{\sum_{j=1}^{k} \log^2\left(X_{(n-j+1)}\right) - k^{-1}\left(\sum_{j=1}^{k} \log X_{(n-j+1)}\right)^2}{\sum_{j=1}^{k} \log\left((k+1)/j\right) \log X_{(n-j+1)} - k^{-1} \sum_{j=1}^{k} \log\left((k+1)/j\right) \sum_{j=1}^{k} \log X_{(n-j+1)}}. \tag{14}$$

### 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 $\hat{\gamma}_1 \le \hat{\gamma} \le \hat{\gamma}_3$.

### 3.4. Hüsler et al. estimator

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

$$\hat{\gamma} = \frac{\sum_{i=0}^{k-1} g\left(\frac{i+1}{k+1}\right) \log\left(\frac{i+1}{k+1}\right) \log\left(\frac{X_{(n-i)}}{X_{(n-k)}}\right)}{\sum_{i=0}^{k-1} g\left(\frac{i+1}{k+1}\right) \log\left(\frac{i+1}{k+1}\right)^2}, \tag{15}$$

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

### 3.5. Beirlant et al. estimators

Beirlant *et al.* (1996b) proposed the regression line to force passing through the anchor point

$$\left(-\log\left((k+1)(n+1)^{-1}\right), \log X_{(n-k)}\right). \tag{16}$$

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 (5). The advantage of this method is that $k$ can be chosen to minimize the weighted mean squared error.

### 3.6. Aban and Meerschaert estimators

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

$$\hat{\gamma} = k\hat{\gamma}_n^H/(k-1), \tag{17}$$

where $\gamma_n^H$ is the Hill estimator. Another estimator, suggested by Aban and Meerschaert, with a slightly higher variance is

$$\hat{\gamma} = \sum_{i=1}^{k} s_i \log X_{(n-i+1)}, \tag{18}$$

where

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

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

In a note, Aban and Meerschaert also suggested reducing the bias by using a regression of the following form:

$$\log X_{(n-j+1)} = \text{const} - \gamma \log\left(\frac{j-1/2}{n}\right), \quad j = 1, \ldots, k. \tag{19}$$

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

### 3.7. Gabaix and Ibragimov estimator

Apart from showing the consistency of the estimator (19), Gabaix and Ibragimov (2011) also considered a similar estimator with reversed dependent and explanatory variables; they showed that $-1/2$ is an optimal shift:

$$\log\left(\frac{j-1/2}{n}\right) = \text{const} - \alpha \log X_{(n-j+1)}, \quad j = 1, \ldots, k. \tag{20}$$

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

$$H(j-1) = \text{const} - \gamma \log X_{(n-j+1)}, \quad j = 1, \ldots, k, \tag{21}$$

$$\log X_{(n-j+1)} = \text{const} - \alpha H(j-1), \quad j = 1, \ldots, k. \tag{22}$$

### 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 taking 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 $\gamma$, $\rho$ and $b_{n,k}$ in the following model:

$$U_{j,k} = \left(\gamma + b_{n,k}\left(\frac{j}{k+1}\right)^{-\rho}\right) f_{j,k}, \quad 1 \le j \le k, \tag{23}$$

where $U_{j,k}$ is defined in Eq. (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 Eq. (23) 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 using quantile plots for $\gamma$, which is not limited by positive values. The method fits a linear regression through the $k$ points which correspond to $k$ largest $X$.

$$\big(-\log(j/n), \log UH_{j,n}\big), \quad j = 1, \ldots, n, \tag{24}$$

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 $\gamma \in \mathbb{R}$:

$$\hat{\gamma} = \frac{1}{k}\sum_{j=1}^{k}\left(\log\frac{k+1}{j} - \frac{1}{k}\sum_{i=1}^{k}\log\frac{k+1}{i}\right)\log UH_{j,n}. \tag{25}$$

They also generalized the Hill estimator:

$$\hat{\gamma} = \frac{1}{k}\sum_{i=1}^{k}\log\left(\frac{UH_{j,n}}{UH_{k+1,n}}\right). \tag{26}$$

However, this generalization 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 done with an OLS or a similar method. There are also several methods designed as minimal distance estimators with no relation to Q-Q plots.

### 4.1. *Vanderwale et al. estimator*

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

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

where $f(\cdot)$ denotes a density function of normalized tail observations: $X_{(n-j+1)}/X_{(n-k)}$. Denote $\theta = \{\gamma, \delta, \rho\}$. Then, assuming 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 of Eq. (27).

$$\varrho(\theta) = \frac{(1+\delta)^2}{(\gamma+2)\gamma} + 2\frac{\delta(1-\delta)(1-\gamma\rho)}{(\gamma+2-\gamma\rho)\gamma} + \frac{\delta^2(1-\gamma\rho)^2}{(\gamma+2-2\gamma\rho)\gamma},$$
$$(\hat{\theta}, \hat{w}) = \underset{\theta, w}{\operatorname{argmin}}\left[w^2\varrho(\theta) - \frac{2w}{k}\sum_{i=1}^{k} f_\theta\left(\frac{X_{(n-j+1)}}{X_{(n-k)}}\right)\right]. \tag{27}$$

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

### 4.2. Tripathi et al. estimators

Tripathi *et al.* (2014) improved the Hill estimator by selecting specific parametric forms, which generalize the Hill estimator, and by 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

$$\hat{\alpha}_1=\frac{k-3}{S_k}. \tag{28}$$

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

$$\hat{\alpha}_2=\frac{k-2}{S_k+\max(0,\log X_k)}. \tag{29}$$

The 'infimum' estimator is:

$$\hat{\alpha}_3=\frac{k-3}{S_k+\max(0,\log X_k)}. \tag{30}$$

The authors showed that these estimators surpass Hill estimator performance in terms of 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

Usually, bias-reduced estimators take the second-order parameter $\rho$ 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 $v_i=log\,U_i$, where $U_i$ is defined in Eq. (1) The MLE estimator of $\alpha$ is given by

$$\hat{\alpha}=\left[k^{-1}\sum_{i=1}^{k}U_i\exp\left\{-\hat{D}(i/n)^{-\hat{\rho}}\right\}\right]^{-1}. \tag{31}$$

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

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

The OLS estimate of $\alpha$ is received from

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

minimization with respect to $\mu$, $D$ and $\rho$, and

$$\hat{\alpha} = \exp(\Gamma'(1) - \hat{\mu}). \tag{32}$$

$\Gamma'(1) \approx -0.5772157$ is a derivative of gamma function at point 1. Alternatively, one can plug OLS estimates $\hat{\rho}$ and $\hat{D}$ into estimator (31). The weakness of these methods is that convergence problems often arise (Gomes and Oliveira, 2003).

Gomes and Martins (2004) simplified Feuerverger and Hall estimators by assuming $\rho = -1$. Furthermore, they approximated $\exp\{-Di/n\}$ as $1 - Di/n$, leading to the following approximation of Eq. (31):

$$\hat{\gamma}^{ML}(k) = \hat{\gamma}_n^H(k) - \left(\frac{1}{k}\sum_{i=1}^{k} iU_i\right)\frac{\sum_{i=1}^{k}(2i-k-1)U_i}{\sum_{i=1}^{k} i(2i-k-1)U_i}, \tag{33}$$

where $\hat{\gamma}_n^H(k)$ is the Hill estimator. The OLS estimator (32) is rewritten as

$$\hat{\gamma}^{LS}(k) = \exp\left\{\frac{2(2k+1)}{k(k-1)}\sum_{i=1}^{k}\log U_i - \Gamma'(1) - \frac{6}{k(k-1)}\sum_{i=1}^{k} i\log U_i\right\}, \tag{34}$$

Gomes *et al.* (2007a) 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 ($\gamma > 0$):

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

$$\gamma_n(k) = \hat{\gamma}_n^H - \frac{M_n(k) - 2(\hat{\gamma}_n^H(k))^2}{2\hat{\gamma}_n^H(k)\hat{\rho}_n}(1-\hat{\rho}_n), \tag{35}$$

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

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

$$\hat{\gamma}_n(k) = \hat{\gamma}_n^P(k) - \frac{\hat{\gamma}_n^P(k) - \hat{\gamma}_n^P(k/4)}{1 - 4^{\hat{\rho}}}, \tag{36}$$

where $\hat{\gamma}_n^P(k)$ is the Pickands estimator defined in Eq. (67). Despite the good theoretical asymptotic properties of the Peng estimators, our simulations show that the expressions under the logarithms in $\hat{\rho}$ estimation sometimes register negative values, leading to a failure 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 proposed estimating $\gamma(r)$ with the Hill estimator $r = 1, ..., k$, and then running a regression

$$\gamma(r) = \beta_0 + \beta_1 r + \epsilon(r), \quad r = 1, ..., k. \tag{37}$$

$\hat{\beta}_0$ 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 $\gamma$ and $q = d_1(n)/d_2(n)$ - is the ratio of biases of these estimators. They considered the following special cases:

$$\hat{\gamma}_1 = 2\hat{\gamma}^{(2)}(k) - \hat{\gamma}^{(1)}(k) \tag{38}$$

$$\hat{\gamma}_2 = 4\hat{\gamma}^{(3)}(k) - 3\gamma^{(1)}(k), \tag{39}$$

$$\hat{\gamma}_3 = 3\hat{\gamma}^{(2)}(k) - 2\gamma^{(3)}(k), \tag{40}$$

$$\hat{\gamma}_4 = 2\hat{\gamma}^{(1)}(k/2) - \gamma^{(1)}(k), \tag{41}$$

where $\hat{\gamma}^{(1)}(k)$ is the Hill estimator (Eq. (4)), $\hat{\gamma}^{(2)}(k)$: a Danielsson *et al.* estimator (Eq. (8)) with $l = 1$, and $\hat{\gamma}^{(3)}(k)$ is a Gomes and Martins estimator with $l = 2$ (Eq. (79)):

Estimator (39) is a simplified version of Peng's estimator (35) with $\rho = -1$. More complete versions of estimators (40) and (41) were studied by Gomes and Martins (2002); Gomes *et al.* (2002) under assumption A4:

$$\hat{\gamma}_3 = \frac{-(2-\hat{\rho})\hat{\gamma}^{(2)}(k) + 2\hat{\gamma}^{(3)}(k)}{\hat{\rho}}, \tag{42}$$

$$\hat{\gamma}_4 = \frac{\hat{\gamma}^{(1)}(k) - 2^{-\hat{\rho}}\hat{\gamma}^{(1)}(k/2)}{1-2^{-\hat{\rho}}}. \tag{43}$$

$\hat{\rho}$ 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 Eq. (3) and $\tau$ is a tuning parameter. This estimator of $\rho$ was introduced by Fraga Alves *et al.* (2003).

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

$$\hat{\gamma}^G_{n,\hat{\rho}}(k) = \frac{\hat{\gamma}_n - \left(n/(n-1)\right)^{\hat{\rho}}\bar{\hat{\gamma}}_n}{1-\left(n/(n-1)\right)^{\hat{\rho}}}, \quad \bar{\hat{\gamma}}_n = \frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_{n-1,i}(k), \tag{44}$$

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

The second estimator is a modification of estimator (43) in the case of the Fréchet model:

$$\hat{\gamma}^{G_2} = \frac{\hat{\gamma}_n(k) - \frac{log(1-k/n)}{log(1-k/(2n))}\hat{\gamma}_n(k/2)}{1-\frac{log(1-k/n)}{log(1-k/(2n))}}. \tag{45}$$

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

$$\hat{\gamma}^{G_1} = \frac{(2+k/n)\hat{\gamma}_n(k/2) - \hat{\gamma}_n(k)}{1+k/n}. \tag{46}$$

### 5.5. Gomes et al. (2005) Jackknife estimator

A different jackknife estimator was introduced by Gomes *et al.* (2005b). They introduced another generalised class of estimators:

$$\hat{\gamma}^{(s)}_n(k) = \frac{s^2}{k^s}\sum_{i=1}^{k} i^{s-1} log\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right), \quad s \geq 1. \tag{47}$$

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

$$\hat{\gamma}^G_n = -\frac{s(1-\hat{\rho})}{\hat{\rho}(s-1)}\left[\hat{\gamma}^{(1)}_n(k) - \frac{s-\hat{\rho}}{s(1-\hat{\rho})}\hat{\gamma}^{(s)}(k)\right]. \tag{48}$$

### 5.6. Gomes et al. (2007) Jackknife estimator

Gomes *et al.* (2007b) considered generalised estimators

$$\hat{\gamma}_1^{(s)}(k) = \frac{s}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{s-1}U_i, \quad s \geq 1 \tag{49}$$

$$\hat{\gamma}_2^{(s)}(k) = -\frac{s^2}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{s-1}\log\left(\frac{i}{k}\right)U_i, \quad s \geq 1 \tag{50}$$

proposing a number of jackknife estimators similar to previous papers. An optimal (variance reducing) combination of two estimators (49) (with different $s$) is

$$\hat{\gamma}^{GJ_1}(\hat{\rho}) = \frac{(1-\hat{\rho})^2\hat{\gamma}_1^{(1)}(k) - (1-2\hat{\rho})\hat{\gamma}_1^{(1-\hat{\rho})}(k)}{\hat{\rho}^2}. \tag{51}$$

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 (49) and (50) is

$$\hat{\gamma}^{GJ_2}(\hat{\rho}) = \frac{1}{\hat{\rho}}(\hat{s}\hat{\gamma}_1^{(\hat{s})}(k) - (\hat{s}-\hat{\rho})\gamma_2^{(\hat{s})}(k)). \tag{52}$$

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, an optimal jackknife estimator based on (50) with two different $s$ was also studied, but the authors do not provide an optimal combination. The authors also studied the case of $\hat{\rho}$ equalized to -1. In this case, the jackknife estimator with the lowest variance based on (49) 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 to reduce the bias of the Hill estimator:

$$\hat{\gamma}_{\hat{\beta},\hat{\rho}}(k) = \hat{\gamma}_n^H(k)\left(1 - \frac{\hat{\beta}}{1-\hat{\rho}}\left(\frac{n}{k}\right)^{\hat{\rho}}\right), \tag{53}$$

$$\hat{\gamma}_{\hat{\beta},\hat{\rho}}(k) = \hat{\gamma}_n^H(k)\exp\left(-\frac{\hat{\beta}}{1-\hat{\rho}}\left(\frac{n}{k}\right)^{\hat{\rho}}\right), \tag{54}$$

where $\hat{\gamma}_n^H(k)$ is the Hill estimator. $\rho$ and $\beta$ are second-order parameters. Consistent estimators for $\hat{\rho}$ can be found in subsections (5.2) or (5.4). $\beta$ 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 $k$ for an estimation of the second-order parameters than that used in the tail index estimation. 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.* (2005a) look for a BLUE in the presence of a second-order regular variation condition. They define $\boldsymbol{T} = \hat{\gamma}_i;\ i = k-m+1, \dots, k;\ 1 \le m \le k$, where $\hat{\gamma}_i$ is an estimator of $\gamma$. 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 the bias. Then, they look for a vector $\boldsymbol{a}' = (a_1, a_2, \dots, a_m)$ 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:

$$\boldsymbol{a} = \Sigma^{-1}\boldsymbol{P}(\boldsymbol{P}'\Sigma^{-1}\boldsymbol{P})^{-1}\begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

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

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

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

$$\hat{\gamma} = \frac{6}{k^2-1}\sum_{i=1}^{k-1} i\hat{\gamma}_n^H(i) - \frac{2k-1}{k+1}\hat{\gamma}_n^H(k), \tag{55}$$

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

$$\hat{\gamma} = \sum_{i=1}^{k+1} a_i(\hat{\rho}) \log X_{(n-i+1)}, \tag{56}$$

with

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

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

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 equal to the Hill estimator:

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

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 the number of observations $k_1$, which is higher than $k$, such as $k_1 = n^{1-\epsilon}$, with a (relatively) small $\epsilon$, such as 0.05.

### 5.10. Baek and Pipiras estimators

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

$$\operatorname{argmin}_{\beta_0,\alpha,\beta_1} \sum_{i=1}^{k}\left(\log\left(\frac{i}{k}\right) - \beta_0 + \alpha\log\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right) - \beta_1\left(\frac{X_{(n-k)}}{n-i+1}\right)\right)^2. \tag{58}$$

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

$$\operatorname{argmin}_{\beta_0,\alpha,\beta_1} \sum_{i=1}^{k}\left(\log\left(\frac{X_{(n-i+1)}}{X_{(n-k)}}\right) - \beta_0 + \frac{1}{\alpha}\log\left(\frac{i}{k}\right) - \beta_1\left(\frac{X_{(n-k)}}{n-i+1}\right)\right)^2. \tag{59}$$

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

$$\hat{\gamma} = \hat{\gamma}_H - \frac{\beta_1}{k} \sum_{i=1}^{k} (X^{-1}_{(n-i+1)} - X^{-1}_{(n-k+1)}), \tag{60}$$

where $\beta_1$ could be estimated from Eq. (59). Having adopted the methodologies of Gomes *et al.* (2000); Peng (1998), they received the following estimators:

$$\hat{\alpha}_G = \frac{2(2+\gamma_H)}{M^{(2)}_{k,n}} - \frac{2}{\hat{\gamma}_H}\sqrt{\frac{2}{M^{(2)}_{k,n}}}. \tag{61}$$

$$\hat{\alpha}_P = -(\gamma_H)^{(-2)} + 2(\gamma_H + 1)/M^{(2)}_{k,n}, \tag{62}$$

where $M^{(2)}_{k,n}$ is defined in Eq. (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:

$$\hat{\gamma}_B(k) = \frac{M^{(2)}_{k,n} - [M^{(1)}_{k,n}]^2}{k^{-1}\sum_{i=1}^{k}\log^2(n/i) - \left(k^{-1}\sum_{i=1}^{k}\log(n/i)\right)^2}. \tag{63}$$

Furthermore, they analysed its two biased-reduced modifications, which are similar, but not absolutely equivalent, to those introduced by Caeiro *et al.* (2005) (Eq. (53) and (54)):

$$\hat{\gamma}_{\hat{\beta},\hat{\rho}}(k) = \hat{\gamma}_B(k)\left(1 - \frac{\hat{\beta}}{(1-\hat{\rho})^2}\left(\frac{n}{k}\right)^{\hat{\rho}}\right), \tag{64}$$

$$\hat{\gamma}_{\hat{\beta},\hat{\rho}}(k) = \hat{\gamma}_B(k)\exp\left(-\frac{\hat{\beta}}{(1-\hat{\rho})^2}\left(\frac{n}{k}\right)^{\hat{\rho}}\right). \tag{65}$$

## 6. Estimators which allow for negative $\gamma$

The advantage of tail-index estimators based on extreme value distribution and GPD over those based on the Pareto is that $\gamma$ is not limited to positive values. Often estimators allow for $\gamma \in R$, negative values corresponding to distributions with a finite right endpoint. However, they usually have a higher asymptotic variance than estimators for positive $\gamma$. The consistency of these estimators is usually shown under assumptions A6 or A7. We separate the estimators into two groups: quantile-type and moment-type. But first, we discuss the Smith estimator, which does not fall into either of these groups.

### 6.1. Smith estimator

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

$$\hat{\gamma} = \frac{1}{m}\sum_{i=1}^{m}\log(1+Y_i/u), \tag{66}$$

where $u$ is a relatively high threshold, $Y_i = X_i - u$, where $i$ is the index of the $i$th exceedance and $m$ is the number of observations higher than $u$. A number of other estimators based on GPD is summarised further in the section of parametric estimators.

### 6.2. Quantile-type estimators

#### 6.2.1. Pickands estimator

Pickands III (1975) proposed a simple method considering the quartiles of the $k$ largest observations. Under the assumption that the tail of the distribution satisfies assumption A6, he found analytical expressions for the 3/4 and 1/2 quartiles and, substituting theoretical quartiles with their empirical counterparts, received the following tail-index estimator:

$$\hat{\gamma}_n^P(k) = \frac{1}{\log 2}\log\left(\frac{X_{(n-\lfloor k/4\rfloor)} - X_{(n-\lfloor k/2\rfloor)}}{X_{(n-\lfloor k/2\rfloor)} - X_{(n-k)}}\right), \tag{67}$$

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.

#### 6.2.2. Falk estimator

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

$$\hat{\gamma}(k,p) = p\hat{\gamma}_n^P([k/2]) + (1-p)\hat{\gamma}_n^P(k), \quad p \in [0,1] \tag{68}$$

The optimal level of $p$ depends on the parameter $\gamma$ itself:

$$p_{opt} = \big((2^{-2\gamma}+2)+2^{1-\gamma}\big)/\big(3(2^{-2\gamma}+2)+2^{2-\gamma}\big).$$

Therefore, $\gamma$ 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 that of normal distribution.

#### 6.2.3. Drees improvements of the Pickands estimator

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

$$\hat{\gamma} = \sum_{i=1}^{m_n} c_{n,i} \hat{\gamma}_n^P(4i), \tag{69}$$

where $m_n$ is an intermediate sequence and $c_{n,i}$, $i = 1, \ldots, m$ are weights $0 \leq c_{n,i} \leq 1$, which sum to unity. The author sets some restrictions on the choice of $m_n$ and $c_{n,i}$ to show asymptotic normality, and also discusses an optimal choice for weights; however, this choice depends on parameter $\gamma$ 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 $\hat{b} = \left(\hat{\gamma}_n^P(k) - \hat{\gamma}_n^P([k/2])\right)^2 / \left(\hat{\gamma}_n^P(k) - 2\hat{\gamma}_n^P(k/2) + \hat{\gamma}_n^P(k/4)\right)$. 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 (69) was refined as

$$\hat{\gamma} = \int \hat{\gamma}_n^P(\lfloor kt \rfloor)\nu(dt) - \frac{\left(\int \hat{\gamma}_n^P(\lfloor kt \rfloor) - \hat{\gamma}_n^P(\lfloor kt \rfloor)\nu(dt)\right)^2}{\int \hat{\gamma}_n^P(\lfloor kt \rfloor) - 2\hat{\gamma}_n^P(\lfloor kt/2 \rfloor) + \hat{\gamma}_n^P(\lfloor kt/4 \rfloor)\nu(dt)}, \tag{70}$$

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

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

#### 6.2.4. Yun estimator

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

$$\hat{\gamma}_{n,m}(u,v) = \frac{1}{\log v} \log \frac{X_{(m)} - X_{(\lfloor um \rfloor)}}{X_{(\lfloor vm \rfloor)} - X_{(\lfloor uvm \rfloor)}}, u, v > 0, \quad u, v \neq 1, \tag{71}$$

where $m \geq 1, \lfloor um \rfloor, \lfloor vm \rfloor, \lfloor uvm \rfloor \leq n$. Pickands estimator corresponds to $\hat{\gamma}_{n,m}(1/2, 1/2)$. With 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 should also mention that Fraga Alves (1995) had already analysed an estimator of the $\hat{\gamma}_{n,m}(c,c)$ form, and that Yun (2000) introduced a less general estimator $\hat{\gamma}_{n,m}(c, 4/c)$ $(1/4 < c < 1)$.

#### 6.2.5. Segers 2005 generalized Pickands estimator

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

$$\hat{\gamma}(c,\lambda) = \sum_{j=1}^{k}\left(\lambda\left(\frac{j}{k}\right) - \lambda\left(\frac{j-1}{k}\right)\right)\log(X_{(n-\lfloor cj\rfloor)} - X_{(n-j)}), \tag{72}$$

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

#### 6.2.6. Müller and Rufibach smooth estimators

Müller and Rufibach (2009) noted that order statistics $X_{(i)}$ can be rewritten as $F_n^{-1}(i/n)$, $i = 1,\ldots,n$, where $F_n(\cdot)$ is an empirical distribution function of $X$. They suggested substituting $F_n$ 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 estimators:

$$\hat{\gamma}_1(H) = \frac{1}{\log 2}\log\left(\frac{H^{-1}\big((n-r(H)+1)/n\big) - H^{-1}\big((n-2r(H)+1)/n\big)}{H^{-1}\big((n-2r(H)+1)/n\big) - H^{-1}\big((n-4r(H)+1)/n\big)}\right), \tag{73}$$

$$\hat{\gamma}_2(H) = \frac{1}{k-1}\sum_{j=2}^{k}\log\left(\frac{X_{(n)} - H^{-1}\big((n-j+1)/n\big)}{X_{(n)} - H^{-1}\big((n-k)/n\big)}\right), \tag{74}$$

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 $k = 4,\ldots,n$ and $\hat{\gamma}_2$ for $k = 3,\ldots,n-1$. If $H = F_n$, $\hat{\gamma}_1$ and $\hat{\gamma}_2$ boil down to Pickands' (Eq. (67)) and Falk's (Eq. (68)) estimators.

### 6.3. Moment-based estimators

#### 6.3.1. Moment estimator

Dekkers *et al.* (1989) proposed a moment estimator of the following form:

$$\hat{\gamma}_n^M(k) = M_{k,n}^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{\left(M_{k,n}^{(1)}\right)^2}{M_{k,n}^{(2)}}\right)^{-1}, \tag{75}$$

where $M_{k,n}^{(i)}$, $i = 1,2$, is defined in Eq. (3). To show the consistency of the moment estimator, 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).

*6.3.2. Gronenboom et al. kernel estimator*

Following Dekkers *et al.* generalization of the Hill estimator (75), Groeneboom *et al.* (2003) generalised a 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 Eq. (5) and

$$q_1 = \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^t K_h\left(\frac{i}{n}\right)(\log X_{(n-i+1)} - \log X_{(n-i)}). \tag{76}$$

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

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

*6.3.3. Fraga Alves et al. 2009 mixed moment estimator*

Fraga Alves *et al.* (2009) considered the conditions a distribution function should 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 $\gamma \in \mathbb{R}$:

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

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

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

$$\hat{\gamma}^{MM} = \frac{\hat{\phi}_n(k) - 1}{1 + 2\min(\hat{\phi}_n(k) - 1, 0)}. \tag{77}$$

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:

$$\hat{\gamma}_1^{(l)}(k) = \frac{M_{k,n}^{(l)}}{\Gamma(l+1)[\hat{\gamma}^{(1)}(k)]^{l-1}}, \tag{78}$$

$$\hat{\gamma}_2^{(l)}(k) = \left(\frac{M_{k,n}^{(l)}}{\Gamma(l+1)}.\right)^{\frac{1}{l}}, \tag{79}$$

where $l > 0$ and $M_{k,n}^{(l)}$ is defined in (3). If $l = 1$, they reduce to the Hill estimator. Estimator (79) 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:

$$\gamma_n^{(l)}(k) = \frac{\Gamma(\alpha)}{M_{k,n}^{(l-1)}(k)}\left(\frac{M_{k,n}^{(2l)}(k)}{\Gamma(2l+1)}\right)^{1/2}, \quad \alpha \geq 1, \tag{80}$$

$l$ is a tuning parameter and $M_{k,n}^{(0)}(k) = 1$. When $l = 1$, the Caeiro and Gomes estimator (80) and the Gomes and Martins estimator (79) (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 (80) may achieve a higher efficiency in comparison to the Hill estimator, if the 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*

The mean-of-order-p class of estimators was independently introduced by Brilhante *et al.* (2013); 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 Eq. (2). They proposed generalizing it to the following mean-of-order-p form:

$$\hat{\gamma}(p) = \begin{cases} \frac{1}{p}\left(1-\left(\frac{1}{k}\sum_{i=1}^{k} V_{i,k}^{p}\right)\right) & \text{if } p \leq 1/\gamma,\ p \neq 0, \\ \log\left(\prod_{i=1}^{k} V_{i,k}\right)^{1/k} & \text{if } p = 0. \end{cases} \tag{81}$$

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 $\rho$ is the second-order parameter.

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

$$\hat{\gamma}(p, \beta, \rho, \phi) = \hat{\gamma}(p)\Big(1 - \frac{\beta(1-\phi)}{1-\rho-\phi}\Big(\frac{n}{k}\Big)^\rho\Big) \tag{82}$$

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

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 yielded the optimal choice of $k$: $\hat{k}^* = \big((1-\hat{\rho})n^{-\hat{\rho}}/(\hat{\beta}\sqrt{-2\hat{\rho}})\big)^{2/(1-2\hat{\rho})}$.

### 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 $\beta > 0$, the tail-index estimator reduces to one of the estimators studied by Gomes and Martins (2001) (Eq. (79)). $f(x) = x^{-p}$ with $p > -1/\gamma$ yields the following estimator:

$$\hat{\gamma} = \frac{1}{p}\Big[\Big(\frac{1}{k}\sum_{j=1}^{k}\Big(\frac{X_{(n-k)}}{X_{(n-j+1)}}\Big)^p\Big)^{-1} - 1\Big] \tag{83}$$

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:

$$\hat{\gamma}(g, l) = \frac{\frac{1}{k}\sum_{i=1}^{k} g\big(\frac{i}{k+1}\big)\Big[\log\frac{X_{(n-i+1)}}{X_{(n-k)}}\Big]^l}{\int_0^1 g(x)(-\log(x))^l\, dx}, \tag{84}$$

where $l > 0$ and $g(x)$ is a positive, non-increasing and integrable weight function defined on $(0,1)$. It is assumed that $\delta > 0.5$ satisfying $\int_0^1 g(x)x^{-\delta}dx < \infty$ and $0 < \int_0^1 g(x)(1-x)^{-\delta}dx < \infty$ exists. $\hat{\gamma}(1,1)$ is the Hill estimator (Eq. (4)), $\hat{\gamma}(1,l)$ - Gomes and Martins (2001) (Eq. (79)), $\hat{\gamma}(g,1)$ - corresponds to the weighted least squares estimator of Hüsler *et al.* (2006) (Eq. (15)).

### 7.6. *Generalisation of a few estimators applying a Box-Cox transformation*

Paulauskas and Vaičiulis (2013) applied a Box-Cox transformation to the Hill and a few other estimators:

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

Define

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

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

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

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

The tail index estimators are received from the relation:

$$\gamma^{(l)}(k,r) = \frac{\lambda^{(l)}(k,r)}{1 + r\lambda^{(l)}(k,r)}, \quad l = 1,...,3. \tag{85}$$

The estimator $\gamma_n^1(k,r)$ generalizes the Hill estimator (Eq. (4)) and it is similar to the Segers estimator (83). With the appropriate choice of $r$, this estimator has a lower asymptotic variance compared to the classical Hill estimator. $\gamma_n^2(k,r)$ generalizes moment estimator (Eq. (75)), and $\gamma_n^3(k,r)$ - Danielsson *et al.* estimator (Eq. (8)).

In a 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:

$$\hat{\gamma}_4 = \frac{2G_n(k,r,1)}{2rG_n(k,r,1)+1+\sqrt{4rG_n(k,r,1)+1}} \tag{86}$$

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

where $\hat{\gamma}^{MR}(k)$ 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 assumption A8. For a review of classical methods for parameter estimation of specific stable distributions please see Mittnik and Rachev (1993). Below, we provide a few more general methods. First, it is important to note that in case of stable distribution tail indexes $0 < \alpha < 2$ correspond to heavy-tailed distributions with corresponding tail indexes; however, 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:

$$\hat{\gamma}_n^{HR}(k) = \frac{\log X_{(n)} - \log X_{(n-k+1)}}{\log k}. \tag{87}$$

The weakness of this estimator is that it converges to its limit at 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:

$$\hat{\gamma}_n^{BB}(k) = -\frac{\log X_{(n-\lfloor \nu k \rfloor+1)} - \log X_{(n-k+1)}}{\log \nu}, \quad 0 < \nu < 1. \tag{88}$$

$\nu$ is a fixed constant.

### 8.3. De Haan and Pereira estimator

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

$$\hat{\beta}_n = \frac{kX^2_{(n-k)}}{\sum_{i=1}^{n-k} X^2_{(i)}} \tag{89}$$

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

### 8.4. Fan estimator

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

$$\hat{\gamma} = \binom{n}{m}^{-1} \sum_{1 \le i_1 < ... < i_m \le n} h(X_{i_1}, ..., X_{i_m}). \tag{90}$$

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:

$$\hat{\gamma} = \frac{\log_+ \sum_{i=1}^{n} (X_i - \bar{X}_n)^2}{2 \log n}, \tag{91}$$

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

### 8.6. Politis estimator

A similar method based on the divergence speed of a 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 $\alpha$ is designed as follows:

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

$$\hat{\alpha} = \frac{2}{\hat{\mu}+1}, \tag{92}$$

where $\bar{Y} = n^{-1}\sum_{j=1}^{k} Y_j$ 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, the Politis estimator is not scale invariant. Also, different permutations of the data may lead to different results. To deal with this problem, the author suggests applying the estimator for a number of permutations and taking 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. (91)) in the following way:

$$\hat{\gamma}_r = \frac{\log \sum_{i=1}^{n} X_i^{2r}}{2r \log n}, \tag{93}$$

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

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

where $S_n(X^2) = \sum_{i=1}^{n} X_i^2$ and $S_{\sqrt{n}}(X^2)$ is the sum of the first $\sqrt{n}$ squared observations.

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

$$\hat{\gamma}^{SCEN} = \frac{1}{M}\sum_{m=1}^{M} \hat{\gamma}_{b^2}^{CEN(m)}, \tag{94}$$

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

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

$$\hat{\gamma}_{b^2}^{RCEN} = \frac{1}{b}\sum_{j=1}^{b} \frac{\log S_{b^2}(X^2) - \log S_b^{(j)}(X^2)}{2\log b}. \tag{95}$$

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

$$\hat{\gamma}^{SRCEN} = \frac{1}{M}\sum_{m=1}^{M} \hat{\gamma}_{b^2}^{RCEN(m)}. \tag{96}$$

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 $\tilde{\alpha}$

$$\sum_{j=1}^{k_n} \psi_c(U_j, \tilde{\alpha}_n(c)) = 0, \tag{97}$$

where $U_j$ is defined in Eq. (1) and

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

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:

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

$$\sum_{i=1}^{n} \psi_v(X_i, \alpha) = 0, \tag{98}$$

the estimator is the value of $\alpha$, which solves Eq. (98). $v$ 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 quantile plot regressions exceeding a threshold.

$$\sum_{i=1}^{k} w(X_{(n-i+1)}; \theta) \frac{\partial}{\partial \theta} \log f_\theta(X_i) = 0, \tag{99}$$

where $f_\theta(X_i)$ is the density function of the right tail of the distribution with parameters $\theta$ ($\alpha \in \theta$) 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 (a lower $c$ gives 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, see 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, $q = 0$ can also be applied.[3] Apart from the Hill estimator, they also applied this idea to the moment estimator (75). Similarly, Fraga Alves *et al.* (2009) introduced a PORT modification of their estimator (77) and Gomes *et al.* (2016) developed PORT methodology for mean-of-order-p class of estimators (81). One of the estimators studied by Gomes and Henriques-Rodrigues (2016) was a PORT version of the Caeiro *et al.* (2005) estimator (53). Li *et al.* (2008) developed PORT methodology for the Caeiro and Gomes estimator (80).

## 11. Block estimators

### *11.1. Davydov, Paulauskas, Račkauskas (DPR) estimator*

LePage *et al.* (1981) studied asymptotic properties of order statistics in case of the domain of attraction of a non-Gaussian stable law. Based on their findings, Davydov *et al.* (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

[3] However, Gomes *et al.* (2008) expressed caution about using $q = 0$.

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 $M_j^{(1)}$ and $M_j^{(2)}$ - the second largest observation in the same group. Also denote

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

They showed that

$$\hat{\alpha} = \frac{Z_T}{1 - Z_T} \tag{100}$$

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). To improve the finite sample properties of this estimator it is wise to apply it to different permutations of the sample, and then take an arithmetic mean or median of the estimates.

### *11.2. Vaičiulis' bias reduced DPR*

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

$$Z_{T,r} = \frac{1}{t}\sum_{j=1}^{T} f_r\left(\frac{M_j^{(2)}}{M_j^{(1)}}\right), \quad f_r(b) = \log^r \frac{1}{b}.$$

$$\hat{\alpha}_l = \frac{\sum_{r=1}^{l}(-1)^{r+1}\left(\Gamma(r+1)\right)^{-1} Z_{T,r}}{\sum_{r=2}^{l+1}(-1)^{r}\left(\Gamma(r+1)\right)^{-1} Z_{T,r}}, \quad l \in \boldsymbol{N}. \tag{101}$$

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

### *11.3. Generalised DPR estimator*

Paulauskas and Vaičiulis (2011) suggested applying a Box-Cox transformation to the DPR estimator (Eq. (100)):

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

where $f_r()$ is a Box-Cox transformation defined in section 7.6.

$$\hat{\alpha}_r = \frac{1 - rZ_T}{Z_T}. \tag{102}$$

If $1-F(x)=C_1x^{-\alpha}+C_2x^{-\beta}+o(x^{-\beta})$, with $0<\alpha<\beta\le\infty$, 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, the 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.

### 11.4. Vaičiulis (2009) estimator

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

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

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

$$\hat{\alpha}\approx\left(2.0024-0.4527IR_{n,m}+0.4246IR_{n,m}^2-0.1386IR_{n,m}^3\right)cos\left(\frac{\pi IR_{n,m}}{2}\right). \quad (103)$$

### 11.5. Qi estimator

Qi (2010) adopted 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, the Hill estimator for each group is applied, and their arithmetic mean is taken:

$$\hat{\gamma}=\frac{1}{\sum_{i=1}^{T}k_i}\sum_{i=1}^{T}\sum_{j=1}^{k_i}\left(\log X_{(m_i-j+1)}^{(i)}-\log X_{(m_i-k_i)}^{(i)}\right), \quad (104)$$

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.

### 11.6. Fialova et al. estimator

Fialová *et al.* (2004) suggested an estimator which takes all observations into account; however, it requires some preliminary knowledge of the tail index. The sample size $n$

is randomly partitioned into $T$ non-overlapping sub-samples of size $m$: $(X_1^{(1)},..,X_r^{(1)})$, ..., $(X_1^{(T)},..,X_r^{(T)})$. Denote $(\bar{X}^{(1)},...,\bar{X}^{(T)})$ as a vector of arithmetic means of the sub-samples. Denote $\hat{F}_{\bar{X}_n}^{(T)}(x) = T^{-1}\sum_{i=1}^{T}\mathbb{1}(\bar{X}^{(i)} \leq x)$ the distribution function of the sub-sample means. Also, suppose there is preliminary information that tail index $\alpha$ is lower than a certain value $\alpha_0$. I.e. the tail is heavier than that of a Pareto distribution with the tail index $\alpha_0$. Then, choose $x_T = T^{\frac{1-\delta}{\alpha_0}}$, for a fixed $\delta$, $0 < \delta < 1$ and calculate

$$\tilde{\alpha}(x_T) = -\frac{\log\left(1-\hat{F}_{\bar{X}_n}^{(T)}(x_T)\right)}{\log x_T},$$
$$\hat{\alpha}(x_T) = \tilde{\alpha}(x_T)\mathbb{1}\left[0 < F_{\bar{X}_n}^{(T)}(x_T) < 1\right] + \tilde{\alpha}_0\mathbb{1}\left[F_{\bar{X}_n}^{(T)}(x_T) = 0 \quad or \quad 1\right]. \tag{105}$$

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

### 11.7. *Jurečková and Picek estimator*

Jurečková (2000) and Jurečková and Picek (2001) suggested a semiparametric test for a hypothesis that the right tail of a distribution is heavier than that of a Pareto distribution with the tail index $\alpha_0$. Jurečková and Picek (2004) reversed the question, and suggested using 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 subsample $j$ as $X_{(m)}^j$. Denote $\hat{F}^*(a) = T^{-1}\sum_{j=1}^{T}\mathbb{1}[X_{(m)}^j \leq a]$ - 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

$$\hat{\alpha}_T = \frac{1}{2}(\hat{\alpha}_T^+ + \hat{\alpha}_T^-), \tag{106}$$

where

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

## 12. Parametric methods

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

### 12.1. 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 $\sigma$ 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 $\gamma$ estimates. As a final estimator, they suggested using

$$\hat{\gamma}=\text{median}(H) \tag{107}$$

### 12.2. Finkelstein et al. estimator

Finkelstein *et al.* (2006) proposed estimating parameter $\alpha$ by solving an equation

$$n^{-1}\sum_{i=1}^{n}\left(\frac{C}{X_i}\right)^{\hat{\alpha}t}=\frac{1}{t+1}, \quad t>0 \tag{108}$$

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 $\alpha$, the left side of Eq. (108) is distributed like the arithmetic mean of uniform random variables raised to power $t$. Having substituted this mean with the mathematical expectation, which equals $(1+t)^{-1}$, Eq. (108) is received. Brzezinski (2016) argued that this estimator has desirable properties as a compromise between ease of use and robustness against outliers in a small-sample setting.

### 12.3. McElroy parametric estimators

McElroy (2007) used the properties of $\text{Var}(\log|X|)$ for the 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 $\hat{V}=\widehat{\text{Var}}(\log|X|)$. $\hat{V}$ is an empirical estimate of logged $|X|$ variance. Denote, $g(\alpha)=\text{Var}(\log|X|)$. Hence, if $g^{-1}(\cdot)$ exists, $\hat{\alpha}=g^{-1}(\hat{V})$.

For stable distribution the tail index estimator can be expressed as

$$\hat{\alpha}=\frac{2}{\sqrt{1+\frac{4\hat{V}-\Psi_2(1/2)}{\Psi_2(1)}}}, \tag{109}$$

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

$$g(\alpha)=\frac{1}{2}\big(\Psi_2(\alpha/2)+\Psi_2(1/2)\big). \tag{110}$$

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

$$\hat{\alpha} = \frac{\sqrt{6}}{\sqrt{\hat{V}\widehat{\text{Kur}}\log|X|}}, \tag{111}$$

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

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

$$\hat{\alpha} = \frac{1}{\sqrt{\hat{V} - \Psi_2(1/2)/4}}. \tag{112}$$

Estimator (112) allows for serially correlated data.

### 12.4. 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: the MLE, the method of moments, and the 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 here because it may provide parameters resulting in arbitrarily large values and the 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 $\gamma$ estimator:

$$\gamma = 0.5(1 - (\bar{X})^2/s^2), \tag{113}$$

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 normality is shown under the assumption of the existence of the fourth finite moment. The third estimator presented by Hosking and Wallis is the method of probability weighted moments. It has the following form:

$$\hat{\gamma}_{PWM} = \bar{X}/(2w - \bar{X}), \tag{114}$$

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

Dupuis and Tsao (1998) modified this estimator for negative $\gamma$, to exclude cases where the right endpoint of the distribution is estimated to be lower than the largest observation.

### 12.5. Zhang estimators

Zhang (2007) created a mix of the MLE and the 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

$$\hat{\gamma} = n^{-1}\sum_{i=1}^{n}\log(1-\hat{b}X_i). \quad (115)$$

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

### 12.6. Wang and Chen estimator

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

$$G(b) = -n^{-1}\sum_{i=1}^{n}\Big\{(2i-1)\log[g_i(b)] + (2n+1-2i)\log[1-g_i(b)]\Big\}. \quad (116)$$

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

with respect to $b$. Next, $\hat{b}$ is plugged into Eq. (115).

### 12.7. Van Zyl (2012) estimator

Van Zyl (2012) noted that for a given $\gamma$ the functional form of assumption A6 can be rewritten as

$$\frac{\big(1-F(x)\big)^{-\gamma}-1}{\gamma} = \frac{x}{\sigma} - \frac{\mu}{\sigma}.$$

Denote $\hat{F}(X_{(i)}) = r/(n+1)$ and $\hat{F}^{(\gamma)}(X_{(i)})$ - the Box-Cox transformation of $1-F(x)$, which is given on the left side of the equation above. Next, for a number of $\gamma$ the regression

$$\hat{F}^{(\gamma)}(X_{(i)}) = \beta_0 + \beta_1 X_{(i)} + \epsilon_i, \quad i=1,\ldots,n. \quad (117)$$

is estimated. $\beta_1 = 1/\sigma$, $\beta_2 = -\mu/\sigma$. $\epsilon_i$ are residuals of the model. It is assumed that they are Laplace distributed. The shape parameter $\gamma$ is chosen which performs best with respect to the profile likelihood.

### 12.8. Van Zyl (2015) 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 $\mu$, $\sigma$ and $\gamma$ 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).

$$\hat{\gamma}_Z = \frac{1}{n}\sum_{i=0}^{n-1} \log\left[Z_i/\mu\right]. \tag{118}$$

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

## 13. Monte Carlo experiment

In this section, we compare the performance of more than 90 estimators on finite samples; to that end we carry out four experiments. Results are presented in Tables 1-4. First, we applied the estimators for i.i.d. samples drawn from a Pareto distribution with scale and shape parameters equal to 1; n=100 and 10,000; $k$ or a similar parameter was equalized to $2\sqrt{n}$, i.e. 20 and 200. In PORT versions of the estimators we set $m = qn + 1$ with $q = 0.1$. Second, we generated random variables as $X = X_1 + X_2$, where $X_1$ denotes a random variable drawn from a Pareto distribution with scale and shape parameters equal to 1, and $X_2$ - is independently drawn from a Pareto distribution with the shape parameter equal to 2; the scale parameter still being equalized to 1. In all cases, the 'true' tail index equals to 1, and in the second case, the second-order parameter equals to 2. We performed 1000 Monte-Carlo simulations for each estimator. In Tables 1-4 we present standard deviations of estimates from the true value of the tail index.

Excellent performance was shown by the Schultze & Steinebach estimator (13), whose errors are lower on average than those of the other estimators. Good performance was also found with the De Haan and Pereira estimator (89) and the Gomes *et al.* (2016) PORT version of the mean-of-order-p estimator (81). Good small sample properties were shown by the Tripathi *et al.* (29), Meerschaert & Scheffer (91) and Jurečková & Picek (106) estimators. There are many well-performing estimators in large samples. The McElroy & Politis (96) and DPR (100) estimators perform well in large samples with and without a finite second-order parameter. The Hill estimator also performs well.

Tables 1, 2, 3, and 4 present time (in minutes) needed to perform Monte-Carlo simulations for every estimator. The experiments were performed on a computer with i7-3630QM (2.4GHz) processor and 6GB installed memory (RAM). For most estimators, Monte-Carlo simulations took around three minutes. The least time consuming estimators are: De Haan and Resnick (87), Hosking and Wallis (113), Pickands (67), Falk (68). Monte-Carlo simulations for most jackknife-type estimators lasted around four minutes. The computational time for block estimators is considerably longer: 47-51 minutes. This is due to the fact that computations were made for 500 random permutations of each sample.

*TABLE 1*

*Standard deviation from the true tail index.*

| Estimators | Eq. | *Pareto distribution* n=100 | *Pareto distribution* n=10000 | *Pareto, second-order* n=100 | *Pareto, second-order* n=10000 | time (min) | Notes |
|---|---|---|---|---|---|---|---|
| Hill | (4) | 0.26 | 0.07 | 0.43 | 0.09 | 2.99 | |
| Fraga Alves | (6) | 0.20 | 0.06 | 0.52 | 0.12 | 2.96 | c=2.5 |
| Aban& Meerschaert | (7) | 1.96 | 0.16 | 2.70 | 0.16 | 2.91 | |
| Danielsson et al. | (8) | 0.39 | 0.10 | 0.51 | 0.10 | 2.90 | l=2 |
| Nuyts | (9) | 3.05 | 0.81 | 2.81 | 1.10 | 2.96 | |
| Weiss | (10) | 0.32 | 0.10 | 0.45 | 0.71 | 3.10 | |
| Zipf | (12) | 0.30 | 0.10 | 0.40 | 0.09 | 2.92 | |
| Schultze&Steinebach | (13) | 0.12 | 0.02 | 0.14 | 0.02 | 2.91 | |
| Schultze&Steinebach | (14) | 0.30 | 0.11 | 0.38 | 0.10 | 2.91 | |
| Beirlant et al. | (16) | 0.24 | 0.08 | 0.35 | 0.08 | 3.05 | |
| Aban&Meerschaert | (17) | 0.27 | 0.07 | 0.49 | 0.09 | 2.95 | |
| Aban&Meerschaert | (19) | 0.36 | 0.10 | 0.52 | 0.11 | 3.02 | |
| Gabaix&Ibragimov | (20) | 0.33 | 0.10 | 0.49 | 0.10 | 3.02 | |
| Gabaix&Ibragimov | (21) | 0.31 | 0.09 | 0.53 | 0.10 | 2.94 | |
| Gabaix&Ibragimov | (22) | 0.29 | 0.09 | 0.43 | 0.10 | 3.05 | |
| Beirlant et al. | (24) | 18.3 | 0.21 | 9.35 | 0.23 | 3.15 | |
| Beirlant et al. | (25) | 53.7 | 0.32 | 23.4 | 0.30 | 2.95 | |
| Beirlant et al. | (26) | 0.64 | 0.12 | 0.77 | 0.13 | 2.92 | |
| Vandewalle et al. | (27) | 56.6 | 0.20 | 35.6 | 0.24 | 3.35 | |
| Tripathi et al. | (28) | 0.24 | 0.07 | 0.31 | 0.08 | 2.95 | |
| Tripathi et al. | (29) | 0.22 | 0.07 | 0.22 | 0.07 | 3.00 | |
| Tripathi et al. | (30) | 0.25 | 0.07 | 0.22 | 0.07 | 3.04 | |
| Gomes&Martins | (33) | 0.27 | 0.07 | 0.44 | 0.05 | 2.80 | |
| Gomes&Martins | (34) | 0.90 | 0.20 | 0.97 | 0.18 | 2.85 | |
| Peng | (36) | 175.5 | 1.67 | 32.6 | 1.30 | 3.07 | |

*TABLE 2*
*Standard deviation from the true tail index (continued).*

| Estimators | Eq. | *Pareto distribution* n=100 | n=10000 | *Pareto, second-order* n=100 | n=10000 | time (min) | Notes |
|---|---|---|---|---|---|---|---|
| Huisman et al. | (37) | 1.22 | 0.32 | 1.37 | 0.32 | 3.31 | |
| Gomes et al. | (38) | 0.79 | 0.17 | 0.85 | 0.16 | 3.86 | |
| Gomes et al. | (39) | 0.45 | 0.16 | 0.44 | 0.16 | 3.85 | |
| Gomes et al. | (40) | 0.68 | 0.16 | 0.80 | 0.17 | 3.78 | |
| Gomes et al. | (41) | 10.7 | 0.18 | 3.96 | 0.17 | 3.78 | |
| Gomes&Martins. | (42) | 124.5 | 3.53 | 93.5 | 12.7 | 3.89 | |
| Gomes&Martins. | (43) | 0.26 | 0.08 | 0.45 | 0.09 | 3.86 | |
| Gomes et al. | (44) | 6.61 | 6.15 | 5.48 | 4.26 | 953.8 | |
| Gomes et al. | (45) | 8.11 | 0.17 | 6.63 | 0.17 | 4.02 | |
| Gomes et al. | (46) | 1.51 | 0.16 | 4.62 | 0.17 | 3.98 | |
| Gomes et al. | (47) | 0.27 | 0.08 | 0.37 | 0.09 | 3.96 | s=2 |
| Gomes et al. | (48) | 28.1 | 0.20 | 53.4 | 0.20 | 3.91 | s=2 |
| Gomes et al. | (49) | 0.27 | 0.08 | 0.43 | 0.09 | 2.87 | l=1.5 |
| Gomes et al. | (50) | 0.32 | 0.08 | 0.47 | 0.09 | 2.86 | l=1.5 |
| Gomes et al. | (51) | 36.6 | 6.88 | 13.4 | 3.87 | 2.91 | |
| Gomes et al. | (52) | 74.2 | 7.27 | 25.4 | 5.16 | 2.91 | |
| Gomes et al. | (52) | 12.7 | 0.17 | 8.39 | 0.15 | 2.88 | $\rho = -1$ |
| Caeiro et al. | (53) | 10.9 | 7.16 | 53.6 | 23.3 | 2.90 | |
| Gomes&Henr.-Rodr. | (53) | 28.9 | 6.70 | 22.6 | 6.47 | 2.90 | PORT, q=0.1 |
| Caeiro et al. | (54) | 16.2 | 15.9 | 9.77 | 8.74 | 2.90 | |
| Gomes et al. | (55) | 2.79 | 0.16 | 3.69 | 0.14 | 4.02 | |
| Gomes et al. | (56) | 19.7 | 7.36 | 41.7 | 5.94 | 3.89 | |
| Beirlant et al. | (57) | 29.7 | 0.27 | 7.75 | 1.76 | 30.2 | |
| Baek&Pipiras | (58) | 0.79 | 0.18 | 1.11 | 0.20 | 3.16 | |
| Baek&Pipiras | (59) | 1.31 | 0.20 | 1.78 | 0.22 | 3.12 | |

*TABLE 3*
*Standard deviation from the true tail index (continued).*

| Estimators | Eq. | *Pareto distribution* n=100 | n=10000 | *Pareto, second-order* n=100 | n=10000 | time (min) | Notes |
|---|---|---|---|---|---|---|---|
| Baek&Pipiras | (60) | 0.28 | 0.07 | 0.42 | 0.09 | 3.22 | |
| Baek&Pipiras | (61) | 0.62 | 0.16 | 0.52 | 0.17 | 3.03 | |
| Baek&Pipiras | (62) | 0.56 | 0.16 | 0.84 | 0.17 | 3.07 | |
| Brito et al. | (63) | 0.42 | 0.14 | 0.56 | 0.14 | 2.82 | |
| Brito et al. | (64) | 60.2 | 5.79 | 0.55 | 0.13 | 3.21 | |
| Brito et al. | (65) | 9.53 | 6.87 | 0.71 | 0.15 | 3.22 | |
| Smith | (66) | 5.87 | 0.16 | 1.06 | 0.15 | 3.06 | |
| Pickands | (67) | 30.8 | 0.94 | 141.4 | 0.58 | 2.84 | |
| Falk | (68) | 36.2 | 0.46 | 37.8 | 0.38 | 2.80 | $p = 5/13$ |
| DEdH | (75) | 7.27 | 0.11 | 12.3 | 0.12 | 3.01 | |
| Fraga Alves et al. | (77) | 1.15 | 0.15 | 4.88 | 0.15 | 3.00 | |
| Fraga Alves et al. | (77) | 0.64 | 0.15 | 1.05 | 0.15 | 3.00 | PORT, $q = 0.1$ |
| Gomes&Martins | (78) | 0.26 | 0.07 | 0.46 | 0.08 | 3.07 | l=0.5 |
| Gomes&Martins | (79) | 0.25 | 0.08 | 0.46 | 0.09 | 3.05 | l=0.5 |
| Caeiro&Gomes | (80) | 0.44 | 0.28 | 0.39 | 0.28 | 4.03 | |
| Mean-of-order-p | (81) | 0.27 | 0.08 | 0.47 | 0.09 | 3.11 | (Brilhante *et al.*, 2014) |
| Mean-of-order-p | (81) | 5.68 | 1.51 | 12.3 | 6.10 | 2.99 | (Caeiro *et al.*, 2016) |
| Gomes et al. | (81) | 0.22 | 0.07 | 0.26 | 0.08 | 3.00 | PORT, q=0.1 |
| Segers | (83) | 0.27 | 0.08 | 0.46 | 0.09 | 2.90 | p=0.7 |
| Vaičiulis&Paulauskas | (85) | 0.40 | 0.10 | 0.56 | 0.11 | 2.88 | r=-0.1, l=1 |
| Vaičiulis&Paulauskas | (85) | 29.4 | 0.22 | 13.1 | 0.22 | 2.86 | r=-0.1, l=2 |
| Vaičiulis&Paulauskas | (85) | 0.59 | 0.17 | 0.80 | 0.19 | 3.00 | r=-0.1, l=3 |
| Vaičiulis&Paulauskas | (86) | 0.39 | 0.10 | 0.53 | 0.11 | 3.01 | r=0.1, u=0.1 |
| De Haan & Resnick | (87) | 0.34 | 0.20 | 0.42 | 0.19 | 2.80 | |
| Bacro & Brito | (88) | 2.92 | 0.26 | 4.38 | 0.29 | 2.95 | $\nu = 0.9$ |

TABLE 4
*Standard deviation from the true tail index (continued).*

| Estimators | Eq. | *Pareto distribution* | | *Pareto, second-order* | | time (min) | Notes |
|---|---|---|---|---|---|---|---|
| | | n=100 | n=10000 | n=100 | n=10000 | | |
| De Haan & Pereira | (89) | 0.16 | 0.06 | 0.24 | 0.11 | 2.93 | |
| Fan | (90) | 0.36 | 0.29 | 0.41 | 0.31 | 5.90 | |
| Meerschaert & Scheffer | (91) | 0.20 | 0.12 | 0.20 | 0.12 | 2.91 | $k=n$ |
| Politis | (92) | 0.28 | 0.24 | 0.36 | 0.25 | 63.1 | $k=n$ |
| McElroy & Politis | (93) | 0.20 | 0.12 | 0.19 | 0.12 | 2.78 | $r=1.5$ |
| McElroy & Politis | (94) | 0.91 | 0.11 | 0.70 | 0.14 | 2.86 | $k=n$ |
| McElroy & Politis | (95) | 0.36 | 0.21 | 0.44 | 0.23 | 2.90 | $k=n$ |
| McElroy & Politis | (96) | 0.23 | 0.07 | 0.34 | 0.10 | 2.89 | $k=n$ |
| DPR | (100) | 0.54 | 0.07 | 0.98 | 0.10 | 47.8 | |
| Generalized DPR | (102) | 0.30 | 0.08 | 0.50 | 0.09 | 48.8 | r=-0.1 |
| Vaičiulis | (103) | 0.25 | 0.08 | 0.32 | 0.10 | 3.43 | |
| Qi | (104) | 0.56 | 0.12 | 1.05 | 0.28 | 49.9 | |
| Fialova et al. | (105) | 0.59 | 0.25 | 0.81 | 0.26 | 3.04 | |
| Jurečková&Picek | (106) | 0.22 | 0.10 | 0.19 | 0.10 | 51.4 | $\delta=0.05$ |
| Hosking&Wallis | (113) | 20.5 | 1.34 | 44.2 | 1.39 | 2.84 | |
| van Zyl | (117) | 0.25 | 0.07 | 0.45 | 0.09 | 2.87 | |
| Brito&Freitas | | 0.30 | 0.10 | 0.36 | 0.10 | 3.07 | |

The most time-consuming estimator is the Gomes *et al.* (44) jackknife estimator. The corresponding Monte-Carlo simulations lasted more than fifteen hours because all possible permutations must be used. In general, greater time consumption in our simulations was not compensated by greater accuracy in the results.

## ACKNOWLEDGEMENTS

I would like to thank Remigijus Leipus, Anton Skrobotov and J. Martin Van Zyl for their help in finding the literature and Marius Vaičiulis for his comments about the structure of the paper. I am grateful to Tucker Sprague McElroy, Constantinos Petropoulos, Martin Van Zyl and Dimitris N. Politis for conferring with me about their methods. Their consultation helped me to avoid inaccuracies present in the original papers. Furthermore, I would like to thank an anonymous referee for his/her constructive comments and suggestions. All remaining errors are my own.

## APPENDIX

### A. NOTATION AND ABBREVIATIONS

*TABLE 5*
*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 |
| $\alpha$ | Tail index |
| $\gamma$ | 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 |
| $\xi$ | Bandwidth |
| $\lfloor u \rfloor$ | the integer part of $u$ |
| ML | Maximum Likelihood |
| MLE | Maximum Likelihood Estimator |
| GPD | Generalized Pareto distribution |

REFERENCES

I. B. ABAN, M. M. MEERSCHAERT (2001). *Shifted Hill's estimator for heavy tails*. Communications in Statistics-Simulation and Computation, 30, no. 4, pp. 949–962.

I. B. ABAN, M. M. MEERSCHAERT (2004). *Generalized least-squares estimators for the thickness of heavy tails*. Journal of Statistical Planning and Inference, 119, no. 2, pp. 341–352.

F. AGTERBERG (1995). *Multifractal modeling of the sizes and grades of giant and supergiant deposits*. International Geology Review, 37, no. 1, pp. 1–8.

R. L. AXTELL (2001). *Zipf distribution of US firm sizes*. Science, 293, no. 5536, pp. 1818–1820.

J. N. BACRO, M. BRITO (1995). *Weak limiting behaviour of a simple tail Pareto-index estimator*. Journal of Statistical Planning and inference, 45, no. 1-2, pp. 7–19.

C. BAEK, 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, no. 7, pp. 1957–1967.

J. BEIRLANT, G. DIERCKX, Y. GOEGEBEUR, G. MATTHYS (1999). *Tail index estimation and an exponential regression model*. Extremes, 2, no. 2, pp. 177–200.

J. BEIRLANT, G. DIERCKX, A. GUILLOU (2005). *Estimation of the extreme-value index and generalized quantile plots*. Bernoulli, 11, no. 6, pp. 949–970.

J. BEIRLANT, G. DIERCKX, A. GUILLOU, C. STĂRICĂ (2002). *On exponential representations of log-spacings of extreme order statistics*. Extremes, 5, no. 2, pp. 157–180.

J. BEIRLANT, F. FIGUEIREDO, M. I. GOMES, B. VANDEWALLE (2008). *Improved reduced-bias tail index and quantile estimators*. Journal of Statistical Planning and Inference, 138, no. 6, pp. 1851–1870.

J. BEIRLANT, A. GUILLOU (2001). *Pareto index estimation under moderate right censoring*. Scandinavian Actuarial Journal, 2001, no. 2, pp. 111–125.

J. BEIRLANT, A. GUILLOU, G. DIERCKX, A. FILS-VILLETARD (2007). *Estimation of the extreme value index and extreme quantiles under random censoring*. Extremes, 10, no. 3, pp. 151–174.

J. BEIRLANT, J. L. TEUGELS (1989). *Asymptotic normality of Hill's estimator*. In J. HÜSLER, R. REISS (eds.), *Extreme Value Theory. Lecture Notes in Statistics*, Springer, New York, vol. 51, pp. 148–155.

J. BEIRLANT, P. VYNCKIER, J. L. TEUGELS (1996a). *Excess functions and estimation of the extreme-value index*. Bernoulli, 2, no. 4, pp. 293–318.

J. BEIRLANT, P. VYNCKIER, J. L. TEUGELS (1996b). *Tail index estimation, Pareto quantile plots regression diagnostics*. Journal of the American Statistical Association, 91, no. 436, pp. 1659–1667.

J. BENHABIB, A. BISIN, M. LUO (2017). *Earnings inequality and other determinants of wealth inequality*. American Economic Review, 107, no. 5, pp. 593–97.

J. BENHABIB, A. BISIN, S. ZHU (2011). *The distribution of wealth and fiscal policy in economies with finitely lived agents*. Econometrica, 79, no. 1, pp. 123–157.

J. BERAN, D. SCHELL (2012). *On robust tail index estimation*. Computational Statistics & Data Analysis, 56, no. 11, pp. 3430–3443.

J. BERAN, D. SCHELL, M. STEHLÍK (2014). *The harmonic moment tail index estimator: Asymptotic distribution and robustness*. Annals of the Institute of Statistical Mathematics, 66, no. 1, pp. 193–220.

V. BRAZAUSKAS, R. SERFLING (2000). *Robust and efficient estimation of the tail index of a single-parameter Pareto distribution*. North American Actuarial Journal, 4, no. 4, pp. 12–27.

M. BRILHANTE, M. GOMES, D. PESTANA (2014). *The mean-of-order p extreme value index estimator revisited*. In A. PACHECO, R. SANTOS, M. OLIVEIRA, C. PAULINO (eds.), *New Advances in Statistical Modeling and Application*, Springer, Berlin, pp. 163–175.

M. F. BRILHANTE, M. I. GOMES, D. PESTANA (2013). *A simple generalisation of the Hill estimator*. Computational Statistics & Data Analysis, 57, no. 1, pp. 518–535.

M. BRITO, L. CAVALCANTE, A. C. M. FREITAS (2016). *Bias-corrected geometric-type estimators of the tail index*. Journal of Physics A: Mathematical and Theoretical, 49, no. 21, pp. 1–38.

M. BRITO, A. C. M. FREITAS (2003). *Limiting behaviour of a geometric-type estimator for tail indices*. Insurance: Mathematics and Economics, 33, no. 2, pp. 211–226.

M. BRZEZINSKI (2016). *Robust estimation of the Pareto tail index: a Monte Carlo analysis*. Empirical Economics, 51, no. 1, pp. 1–30.

F. CAEIRO, M. I. GOMES (2002). *A class of asymptotically unbiased semi-parametric estimators of the tail index*. Test, 11, no. 2, pp. 345–364.

F. CAEIRO, M. I. GOMES (2006). *A new class of estimators of a "scale" second order parameter*. Extremes, 9, no. 3-4, pp. 193–211.

F. CAEIRO, M. I. GOMES, J. BEIRLANT, T. DE WET (2016). *Mean-of-order p reduced-bias extreme value index estimation under a third-order framework*. Extremes, 19, no. 4, pp. 561–589.

F. CAEIRO, M. I. GOMES, D. PESTANA (2005). *Direct reduction of bias of the classical Hill estimator*. Revstat, 3, no. 2, pp. 113–136.

T. CHANEY (2008). *Distorted gravity: the intensive and extensive margins of international trade*. American Economic Review, 98, no. 4, pp. 1707–1721.

G. CIUPERCA, C. MERCADIER (2010). *Semi-parametric estimation for heavy tailed distributions*. Extremes, 13, no. 1, pp. 55–87.

F. A. COWELL, E. FLACHAIRE (2007). *Income distribution and inequality measurement: The problem of extreme values*. Journal of Econometrics, 141, no. 2, pp. 1044–1072.

S. CSORGO, P. DEHEUVELS, D. MASON (1985). *Kernel estimates of the tail index of a distribution*. The Annals of Statistics, 13, pp. 1050–1077.

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

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

J. DANIELSSON, D. W. JANSEN, C. G. DE VRIES (1996). *The method of moments ratio estimator for the tail shape parameter*. Communications in Statistics-Theory and Methods, 25, no. 4, pp. 711–720.

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

Y. DAVYDOV, V. PAULAUSKAS, A. RAČKAUSKAS (2000). *More on p-stable convex sets in Banach spaces*. Journal of Theoretical Probability, 13, no. 1, pp. 39–64.

L. DE HAAN, T. T. PEREIRA (1999). *Estimating the index of a stable distribution*. Statistics & Probability Letters, 41, no. 1, pp. 39–55.

L. DE HAAN, S. RESNICK (1998). *On asymptotic normality of the Hill estimator*. Stochastic Models, 14, no. 4, pp. 849–866.

L. DE HAAN, S. I. RESNICK (1980). *A simple asymptotic estimate for the index of a stable distribution*. Journal of the Royal Statistical Society. Series B, 42, no. 1, pp. 83–87.

L. D. DE HAAN, L. PENG (1998). *Comparison of tail index estimators*. Statistica Neerlandica, 52, no. 1, pp. 60–70.

L. F. M. DE HAAN (1970). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*. Mathematisch Centrum, Amsterdam, NL.

P. DEHEUVELS, E. HAEUSLER, D. M. MASON (1988). *Almost sure convergence of the Hill estimator*. In *Mathematical Proceedings of the Cambridge Philosophical Society*. vol. 104, pp. 371–381.

A. L. DEKKERS, J. H. EINMAHL, L. DE HAAN (1989). *A moment estimator for the index of an extreme-value distribution*. The Annals of Statistics, 17, no. 4, pp. 1833–1855.

H. DREES (1995). *Refined Pickands estimators of the extreme value index*. The Annals of Statistics, 23, no. 6, pp. 2059–2080.

H. DREES (1996). *Refined Pickands estimators with bias correction*. Communications in Statistics-Theory and Methods, 25, no. 4, pp. 837–851.

H. DREES (1998a). *A general class of estimators of the extreme value index*. Journal of Statistical Planning and Inference, 66, no. 1, pp. 95–112.

H. DREES (1998b). *On smooth statistical tail functionals*. Scandinavian Journal of Statistics, 25, no. 1, pp. 187–210.

H. DREES, L. DE HAAN, S. RESNICK, *et al.* (2000). *How to make a Hill plot*. The Annals of Statistics, 28, no. 1, pp. 254–274.

D. DUPUIS, M. TSAO (1998). *A hybrid estimator for generalized Pareto and extreme-value distributions*. Communications in Statistics-Theory and Methods, 27, no. 4, pp. 925–941.

D. J. DUPUIS, S. MORGENTHALER (2002). *Robust weighted likelihood estimators with an application to bivariate extreme value problems*. Canadian Journal of Statistics, 30, no. 1, pp. 17–36.

D. J. DUPUIS, M.-P. VICTORIA-FESER (2006). *A robust prediction error criterion for Pareto modelling of upper tails*. Canadian Journal of Statistics, 34, no. 4, pp. 639–658.

M. FALK (1994). *Efficiency of convex combinations of Pickands estimator of the extreme value index*. Journal of Nonparametric Statistics, 4, no. 2, pp. 133–147.

Z. FAN (2004). *Estimation problems for distributions with heavy tails*. Journal of Statistical Planning and Inference, 123, no. 1, pp. 13–40.

R. FERRIERE, B. CAZELLES (1999). *Universal power laws govern intermittent rarity in communities of interacting species*. Ecology, 80, no. 5, pp. 1505–1521.

A. FEUERVERGER, P. HALL (1999). *Estimating a tail exponent by modelling departure from a Pareto distribution*. The Annals of Statistics, 27, no. 2, pp. 760–781.

A. FIALOVÁ, J. JURECKOVÁ, J. PICEK (2004). *Estimating Pareto tail index based on sample means*. REVSTAT–Statistical Journal, 2, no. 1, pp. 75–100.

M. FINKELSTEIN, H. G. TUCKER, J. ALAN VEEH (2006). *Pareto tail index estimation revisited*. North American Actuarial Journal, 10, no. 1, pp. 1–10.

M. FRAGA ALVES (1995). *Estimation of the tail parameter in the domain of attraction of an extremal distribution*. Journal of Statistical Planning and Inference, 45, no. 1-2, pp. 143–173.

M. FRAGA ALVES (2001). *A location invariant Hill-type estimator*. Extremes, 4, no. 3, pp. 199–217.

M. FRAGA ALVES, M. I. GOMES, L. DE HAAN (2003). *A new class of semi-parametric estimators of the second order parameter*. Portugaliae Mathematica, 60, no. 2, pp. 193–214.

M. I. FRAGA ALVES, M. I. GOMES, L. DE HAAN, C. NEVES (2009). *Mixed moment estimator and location invariant alternatives*. Extremes, 12, no. 2, pp. 149–185.

X. GABAIX (2009). *Power laws in economics and finance*. Annual Review of Economics, 1, no. 1, pp. 255–294.

X. GABAIX, P. GOPIKRISHNAN, V. PLEROU, H. E. STANLEY (2003). *A theory of power-law distributions in financial market fluctuations*. Nature, 423, no. 6937, pp. 267–270.

X. GABAIX, R. IBRAGIMOV (2011). *Rank- 1/2: a simple way to improve the OLS estimation of tail exponents*. Journal of Business & Economic Statistics, 29, no. 1, pp. 24–39.

X. GABAIX, A. LANDIER (2008). *Why has CEO pay increased so much?* The Quarterly Journal of Economics, 123, no. 1, pp. 49–100.

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

M. I. GOMES, M. I. F. ALVES, P. A. SANTOS (2008). *PORT Hill and moment estimators for heavy-tailed models*. Communications in Statistics - Simulation and Computation, 37, no. 7, pp. 1281–1306.

M. I. GOMES, F. CAEIRO, F. FIGUEIREDO (2004). *Bias reduction of a tail index estimator through an external estimation of the second-order parameter*. Statistics, 38, no. 6, pp. 497–510.

M. I. GOMES, F. FIGUEIREDO, S. MENDONÇA (2005a). *Asymptotically best linear unbiased tail estimators under a second-order regular variation condition*. Journal of Statistical Planning and Inference, 134, no. 2, pp. 409–433.

M. I. GOMES, A. GUILLOU (2015). *Extreme value theory and statistics of univariate extremes: A review*. International Statistical Review, 83, no. 2, pp. 263–292.

M. I. GOMES, L. HENRIQUES-RODRIGUES (2016). *Competitive estimation of the extreme value index*. Statistics & Probability Letters, 117, pp. 128–135.

M. I. GOMES, L. HENRIQUES-RODRIGUES, B. MANJUNATH (2016). *Mean-of-order-p location-invariant extreme value index estimation*. Revstat, 14, no. 3, pp. 273–296.

M. I. GOMES, M. J. MARTINS (2001). *Generalizations of the Hill estimator–asymptotic versus finite sample behaviour*. Journal of Statistical Planning and Inference, 93, no. 1-2, pp. 161–180.

M. I. GOMES, M. J. MARTINS (2002). *Asymptotically unbiased estimators of the tail index based on external estimation of the second order parameter*. Extremes, 5, no. 1, pp. 5–31.

M. I. GOMES, M. J. MARTINS (2004). *Bias reduction and explicit semi-parametric estimation of the tail index*. Journal of Statistical Planning and Inference, 124, no. 2, pp. 361–378.

M. I. GOMES, M. J. MARTINS, M. NEVES (2000). *Alternatives to a semi-parametric estimator of parameters of rare events - the jackknife methodology*. Extremes, 3, no. 3, pp. 207–229.

M. I. GOMES, M. J. MARTINS, M. NEVES (2002). *Generalized jackknife semi-parametric estimators of the tail index*. Portugaliae Mathematica, 59, no. 4, pp. 393–408.

M. I. GOMES, M. J. MARTINS, M. NEVES (2007a). *Improving second order reduced bias extreme value index estimation*. Revstat, 5, no. 2, pp. 177–207.

M. I. GOMES, C. MIRANDA, C. VISEU (2007b). *Reduced-bias tail index estimation and the jackknife methodology*. Statistica Neerlandica, 61, no. 2, pp. 243–270.

M. I. GOMES, O. OLIVEIRA (2003). *Maximum likelihood revisited under a semi-parametric context-estimation of the tail index*. Journal of Statistical Computation and Simulation, 73, no. 4, pp. 285–301.

M. I. GOMES, H. PEREIRA, M. C. MIRANDA (2005b). *Revisiting the role of the jackknife methodology in the estimation of a positive tail index*. Communications in Statistics-Theory and Methods, 34, no. 2, pp. 319–335.

P. GROENEBOOM, H. LOPUHAÄ, P. DE WOLF (2003). *Kernel-type estimators for the extreme value index*. The Annals of Statistics, 31, no. 6, pp. 1956–1995.

E. HAEUSLER, J. L. TEUGELS (1985). *On asymptotic normality of Hill's estimator for the exponent of regular variation*. The Annals of Statistics, 13, no. 1, pp. 743–756.

P. HALL (1982). *On some simple estimates of an exponent of regular variation*. Journal of the Royal Statistical Society. Series B, 44, no. 1, pp. 37–42.

P. HALL, A. WELSH (1985). *Adaptive estimates of parameters of regular variation*. The Annals of Statistics, 13, pp. 331–341.

B. M. HILL (1975). *A simple general approach to inference about the tail of a distribution*. The Annals of Statistics, 3, no. 5, pp. 1163–1174.

J. HINLOOPEN, C. VAN MARREWIJK (2012). *Power laws and comparative advantage*. Applied Economics, 44, no. 12, pp. 1483–1507.

J. R. M. HOSKING, J. R. WALLIS (1987). *Parameter and quantile estimation for the generalized Pareto distribution*. Technometrics, 29, no. 3, pp. 339–349.

F. HUEBNER, D. LIU, 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, vol. 1, pp. 471–476.

R. HUISMAN, K. G. KOEDIJK, C. J. M. KOOL, F. PALM (2001). *Tail-index estimates in small samples*. Journal of Business & Economic Statistics, 19, no. 2, pp. 208–216.

J. HÜSLER, D. LI, S. MÜLLER (2006). *Weighted least squares estimation of the extreme value index*. Statistics & Probability Letters, 76, no. 9, pp. 920–930.

J. JUREČKOVÁ (2000). *Test of tails based on extreme regression quantiles*. Statistics & Probability Letters, 49, no. 1, pp. 53–61.

J. JUREČKOVÁ, J. PICEK (2001). *A class of tests on the tail index*. Extremes, 4, no. 2, pp. 165–183.

J. JUREČKOVÁ, J. PICEK (2004). *Estimates of the tail index based on nonparametric tests*. In M. HUBERT, G. PISON, A. STRUYF, S. VAN AELST (eds.), *Theory and Applications of Recent Robust Methods*, Springer, pp. 141–152.

T. KAIZOJI (2003). *Scaling behavior in land markets*. Physica A: Statistical Mechanics and its Applications, 326, no. 1-2, pp. 256–264.

S. KANG, J. SONG (2017). *Parameter and quantile estimation for the generalized Pareto distribution in peaks over threshold framework*. Journal of the Korean Statistical Society, 46, no. 4, pp. 487–501.

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

M. KRATZ, S. I. RESNICK (1996). *The qq-estimator and heavy tails*. Stochastic Models, 12, no. 4, pp. 699–724.

R. LEPAGE, M. WOODROOFE, J. ZINN (1981). *Convergence to a stable distribution via order statistics*. The Annals of Probability, 9, no. 4, pp. 624–632.

J. LI, Z. PENG, S. NADARAJAH (2008). *A class of unbiased location invariant Hill-type estimators for heavy tailed distributions*. Electronic Journal of Statistics, 2, pp. 829–847.

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

D. M. MASON (1982). *Laws of large numbers for sums of extreme values*. The Annals of Probability, 10, no. 3, pp. 754–764.

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

T. MCELROY, D. N. POLITIS (2007). *Moment-based tail index estimation*. Journal of Statistical Planning and Inference, 137, no. 4, pp. 1389–1406.

M. M. MEERSCHAERT, H.-P. SCHEFFLER (1998). *A simple robust estimation method for the thickness of heavy tails*. Journal of Statistical Planning and Inference, 71, no. 1, pp. 19–34.

S. MITTNIK, S. T. RACHEV (1993). *Modeling asset returns with alternative stable distributions*. Econometric Reviews, 12, no. 3, pp. 261–330.

S. MÜLLER, K. RUFIBACH (2009). *Smooth tail-index estimation*. Journal of Statistical Computation and Simulation, 79, no. 9, pp. 1155–1167.

U. K. MÜLLER, Y. WANG (2017). *Fixed-k asymptotic inference about tail properties*. Journal of the American Statistical Association, 112, no. 519, pp. 1334–1343.

E. A. NADARAYA (1964). *On estimating regression*. Theory of Probability & Its Applications, 9, no. 1, pp. 141–142.

L. NÉMETH, A. ZEMPLÉNI (2017). *Regression estimator for the tail index*. arXiv preprint arXiv:1708.04815.

M. E. NEWMAN (2005). *Power laws, Pareto distributions and Zipf's law*. Contemporary Physics, 46, no. 5, pp. 323–351.

W. D. NORDHAUS (2012). *Economic policy in the face of severe tail events*. Journal of Public Economic Theory, 14, no. 2, pp. 197–219.

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

T. OGWANG (2011). *Power laws in top wealth distributions: Evidence from Canada*. Empirical Economics, 41, no. 2, pp. 473–486.

V. PARETO (1897). *Cours d'e'conomie politique*. Lausanne, Rouge.

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

V. PAULAUSKAS, M. VAIČIULIS (2011). *Several modifications of DPR estimator of the tail index*. Lithuanian Mathematical Journal, 51, no. 1, pp. 36–50.

V. PAULAUSKAS, M. VAIČIULIS (2012). *Estimation of the tail index in the max-aggregation scheme*. Lithuanian Mathematical Journal, 52, no. 3, pp. 297–315.

V. PAULAUSKAS, M. VAIČIULIS (2013). *On an improvement of Hill and some other estimators*. Lithuanian Mathematical Journal, 53, no. 3, pp. 336–355.

V. PAULAUSKAS, M. VAIČIULIS (2017a). *A class of new tail index estimators*. Annals of the Institute of Statistical Mathematics, 69, no. 2, pp. 461–487.

V. PAULAUSKAS, M. VAIČIULIS (2017b). *Comparison of the several parameterized estimators for the positive extreme value index*. Journal of Statistical Computation and Simulation, 87, no. 7, pp. 1342–1362.

L. PENG (1998). *Asymptotically unbiased estimators for the extreme-value index*. Statistics & Probability Letters, 38, no. 2, pp. 107–115.

J. PICKANDS III (1975). *Statistical inference using extreme order statistics*. The Annals of Statistics, 3, no. 1, pp. 119–131.

V. PISARENKO, D. SORNETTE (2003). *Characterization of the frequency of extreme earthquake events by the generalized Pareto distribution*. Pure and Applied Geophysics, 160, no. 12, pp. 2343–2364.

D. N. POLITIS (2002). *A new approach on estimation of the tail index*. Comptes Rendus Mathematique, 335, no. 3, pp. 279–282.

Y. QI (2010). *On the tail index of a heavy tailed distribution*. Annals of the Institute of Statistical Mathematics, 62, no. 2, pp. 277–298.

M. H. QUENOUILLE (1956). *Notes on bias in estimation*. Biometrika, 43, no. 3/4, pp. 353–360.

K. T. ROSEN, M. RESNICK (1980). *The size distribution of cities: An examination of the Pareto law and primacy*. Journal of Urban Economics, 8, no. 2, pp. 165–186.

P. A. SANTOS, M. ALVES, M. I. GOMES (2006). *Peaks over random threshold methodology for tail index and high quantile estimation*. Revstat, 4, no. 3, pp. 227–247.

J. SCHULTZE, J. STEINEBACH (1996). *On least squares estimates of an exponential tail coefficient*. Statistics & Risk Modeling, 14, no. 4, pp. 353–372.

D. A. SEEKELL, M. L. PACE (2011). *Does the Pareto distribution adequately describe the size-distribution of lakes?* Limnology and Oceanography, 56, no. 1, pp. 350–356.

J. SEGERS (2001). *Residual estimators*. Journal of Statistical Planning and Inference, 98, no. 1-2, pp. 15–27.

J. SEGERS (2005). *Generalized Pickands estimators for the extreme value index*. Journal of Statistical Planning and Inference, 128, no. 2, pp. 381–396.

H. A. SIMON, C. P. BONINI (1958). *The size distribution of business firms*. The American Economic Review, 48, no. 4, pp. 607–617.

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

K. T. SOO (2005). *Zipf's law for cities: A cross-country investigation*. Regional science and urban Economics, 35, no. 3, pp. 239–263.

A. A. TODA (2012). *The double power law in income distribution: Explanations and evidence*. Journal of Economic Behavior & Organization, 84, no. 1, pp. 364–381.

A. A. TODA, K. WALSH (2015). *The double power law in consumption and implications for testing Euler equations*. Journal of Political Economy, 123, no. 5, pp. 1177–1200.

Y. M. TRIPATHI, S. KUMAR, C. PETROPOULOS (2014). *Improved estimators for parameters of a Pareto distribution with a restricted scale*. Statistical Methodology, 18, pp. 1–13.

M. VAIČIULIS (2009). *An estimator of the tail index based on increment ratio statistics*. Lithuanian Mathematical Journal, 49, no. 2, pp. 222–233.

M. VAIČIULIS (2012). *Asymptotic properties of generalized DPR statistic*. Lithuanian Mathematical Journal, 52, no. 1, pp. 95–110.

J. M. VAN ZYL (2012). *A median regression model to estimate the parameters of the three-parameter generalized Pareto distribution*. Communications in Statistics-Simulation and Computation, 41, no. 4, pp. 544–553.

J. M. VAN ZYL (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, no. 1, pp. 171–183.

B. VANDEWALLE, J. BEIRLANT, A. CHRISTMANN, M. HUBERT (2007). *A robust estimator for the tail index of Pareto-type distributions*. Computational Statistics & Data Analysis, 51, no. 12, pp. 6252–6268.

M.-P. VICTORIA-FESER, E. RONCHETTI (1994). *Robust methods for personal-income distribution models*. Canadian Journal of Statistics, 22, no. 2, pp. 247–258.

L. VIHAROS (1999). *Weighted least-squares estimators of tail indices*. Probability and Mathematical Statistics, 19, no. 2, pp. 249–265.

C. WANG, G. CHEN (2016). *A new hybrid estimation method for the generalized Pareto distribution*. Communications in Statistics-Theory and Methods, 45, no. 14, pp. 4285–4294.

L. WEISS (1971). *Asymptotic inference about a density function at an end of its range*. Naval Research Logistics, 18, no. 1, pp. 111–114.

S. YUN (2000). *A class of Pickands-type estimators for the extreme value index*. Journal of Statistical Planning and Inference, 83, no. 1, pp. 113–124.

S. YUN (2002). *On a generalized Pickands estimator of the extreme value index*. Journal of Statistical Planning and Inference, 102, no. 2, pp. 389–409.

J. ZHANG (2007). *Likelihood moment estimation for the generalized Pareto distribution*. Australian & New Zealand Journal of Statistics, 49, no. 1, pp. 69–77.

J. ZHANG (2010). *Improving on estimation for the generalized Pareto distribution*. Technometrics, 52, no. 3, pp. 335–339.

J. ZHANG, M. A. STEPHENS (2009). *A new and efficient estimation method for the generalized Pareto distribution*. Technometrics, 51, no. 3, pp. 316–325.

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

G. K. ZIPF (1941). *National Unity and Disunity; the Nation as a Bio-Social Organism*. Principia Press, Bloomington, Indiana.

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

SUMMARY

Heavy-tailed distributions are often encountered in economics, finance, biology, telecommunications, geology, etc. The heaviness of a tail is measured by a tail index. Numerous methods for tail index estimation have been proposed. This paper reviews more than one hundred Pareto (and equivalent) tail index estimators. It focuses on univariate estimators for non-truncated data. We discuss the basic features of these estimators and provide their analytical expressions. As samples from heavy-tailed distributions are often analysed by researchers from various sciences, the paper provides nontechnical explanations of the methods, so as to be understood by researchers with intermediate skills in statistics. We also discuss the strengths and weaknesses of the estimators, if known. The main focus of the paper is semi-parametric estimators; however, a number of parametric estimators under-represented in previous reviews are also discussed. The paper can be viewed as a catalog or a reference work on Pareto-tail index estimators. A Monte-Carlo comparison of more than 90 estimators is presented.

*Keywords*: Heavy tails; Pareto distribution; Tail index; Review.