# A Class of Weighted Hill Estimators

Frederico Caeiro
Faculdade de Ciências e Tecnologia (FCT), Universidade Nova de Lisboa, Portugal , e-mail: fac@fct.unl.pt
Centro de Matemática e Aplicações (CMA), Universidade Nova de Lisboa, Portugal

Ayana Mateus
Faculdade de Ciências e Tecnologia (FCT), Universidade Nova de Lisboa, Portugal , e-mail: amf@fct.unl.pt
Centro de Matemática e Aplicações (CMA), Universidade Nova de Lisboa, Portugal

Louiza Soltane
Laboratory of Applied Mathematics, Mohamed Khider University, Biskra, Algeria, e-mail: l.soltane@univ-biskra.dz

**Abstract:** In Statistics of Extremes, the estimation of the extreme value index is an essential requirement for further tail inference. In this work we deal with the estimation of a strictly positive extreme value index from a model with a Pareto-type right tail. Under this framework we propose a new class of weighted Hill estimators, parameterized with a tuning parameter $a$. We derive their non-degenerate asymptotic behaviour and analyse the influence of the tuning parameter in such result. Their finite sample performance is analysed through a Monte Carlo simulation study. A comparison with other important extreme value index estimators from the literature is also provided.

This is an electronic postprint of the publication:
Caeiro, F., Mateus, A. and Soltane, L. (2021). A class of weighted Hill estimators. *Computational and Mathematical Methods.* 3(6), e1167.
doi: https://doi.org/10.1002/cmm4.1167
This reprint differs from the original in typographic details.

## 1 Introduction

Let $X_1, X_2, \ldots, X_n$ be a sample of $n$ independent observations from a common unknown distribution function (d.f.) $F$. Let us assume that the sample maximum, $M_n = \max(X_1, X_2, \ldots, X_n)$, after adequate linear normalization, converge in law to a non-degenerate limit distribution. Then, this limit distribution is definitely the Extreme Value distribution with d.f. given by

$$EV_\xi(x) = \begin{cases} \exp(-(1+\xi x)^{-1/\xi}), & 1+\xi x > 0 \quad \text{if } \xi \neq 0 \\ \exp(-\exp(-x)), & x \in \mathbf{R} \quad \text{if } \xi = 0, \end{cases}$$

and we write $F \in D(EV_\xi)$. The shape parameter, $\xi \in \mathbf{R}$, quantifies the heaviness of the right tail of $F$ and is usually refereed as the extreme value index (EVI). The aim of this paper is the estimation of the parameter $\xi$. This parameter dominates the tail behaviour and needs to be estimated in a precise way because other tail parameters such as an extreme quantile or a tail probability, depends on the value of the EVI. In the present paper we restrict to models $F \in D(EV_\xi)$ with $\xi > 0$. Models with a positive EVI are called Pareto-type or heavy tailed models. It is well-known (see [4, 23] among others) that Pareto-type models have a regularly varying tail with a negative index of regular variation equal to $-1/\xi$, i.e.,

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-1/\xi}, \quad \forall x > 0 \tag{1.1}$$

with $\overline{F}(x) = P(X > x)$ the survival function of $X$ and we write $\overline{F} \in RV_{-1/\xi}$ whenever this condition holds. The condition (1.1) is usually called the first-order condition. Consequently, we can write

$$\overline{F}(x) = x^{-1/\xi} l(x), \quad x \to \infty \tag{1.2}$$

with $l(x)$ a slowly varying function at infinity, i.e., $l(tx)/l(t) \to 1$, as $t \to \infty$, for all $x > 0$. This means that $\overline{F}$ tends to zero at a polynomial rate. For more details about the theory of regular variation see [6]. The strict Pareto model ([1]), also known as Pareto type I model, has a regularly varying survival function $\overline{F}(x) = (x/c)^{-1/\xi}$, $x > c > 0$ with index $-1/\xi$. The associated slowly varying function $l(x)$ is equal to $c^{1/\xi}$. Note that this is the only model with a constant slowly varying function. The Burr (type XII) distribution ([7, 38]), a member of the three parameter Singh-Maddala distribution ([43]), with survival function

$$\overline{F}(x) = (1 + x^a)^{-b}, \quad x > 0 \qquad (a, b > 0), \tag{1.3}$$

is an important model in the actuarial literature. This model has also a regularly varying right tail with index $-ab$ and (1.2) holds with a slowly varying function $l(x) = (1 + x^{-a})^{-b}$, $x > 0$.

Whenever applying Pareto-type models, the Hill estimator is frequently used to estimate the parameter $\xi > 0$. This estimator can be derived by maximum likelihood and is defined as the average of the log-excesses of the threshold $u = X_{n-k:n} > 0$,

$$\hat{\xi}^H(k) := \frac{1}{k} \sum_{i=1}^{k} \ln X_{n-i+1:n} - \ln X_{n-k:n}, \qquad k = 1, 2, \ldots, n-1, \tag{1.4}$$

where $X_{i:n}$, $i = 1, \ldots, n$, denotes the *i-th* ascending order statistics based on the sample of size $n$. This estimator can also be derived with other approaches (see [4], section 4.2). Consistency of the Hill estimator is guaranteed if $k$ is intermediate, that is, if $k = k_n \in [1, n-1]$ is a sequence of positive integers satisfying

$$k \to \infty \quad \text{and} \quad k/n \to 0, \quad \text{as} \qquad n \to \infty. \tag{1.5}$$

The asymptotic and finite sample properties of $\hat{\xi}^H(k)$, in (1.4), have been detailed studied by several authors. See the recent reviews in references [3, 29]. Despite being very popular it is difficult to use the Hill estimator due to its sensitivity in the choice of the level $k$.

In the strict Pareto case, if we choose $k = n-1$, the Hill estimator is the Maximum Likelihood estimator of $\xi$. However, if $F$ differs from the strict Pareto model, $\hat{\xi}^H(k)$ has often a large bias, if $k$ is too large. On the other hand, the variance of $\hat{\xi}^H(k)$ decreases as $k$ increases. Thus, to choose the level $k$, we have to deal with the well-known bias-variance trade-off problem. Typically, we construct a Hill plot, i.e., a graphic of $(k, \hat{\xi}^H(k))$, $1, 2, \ldots, n-1$ and choose the level $k$ as large as possible in the first stability or plateau region. However, while the procedure may yield decent estimates if the tail of $F$ is close to the strict Pareto distribution, in other cases it fails completely, because such region does not exist, and we are confronted with the so-called "Hill horror plot" ([41]). To illustrate such behaviour we show in Figure 1 a Hill (horror) plot based on a sample of 200 independent observations from the Burr distribution, in (1.3), with $a = 1$ and $b = 2$. The plot shows that the estimate of $\xi$, provided by the Hill estimator, varies greatly without any stable part.

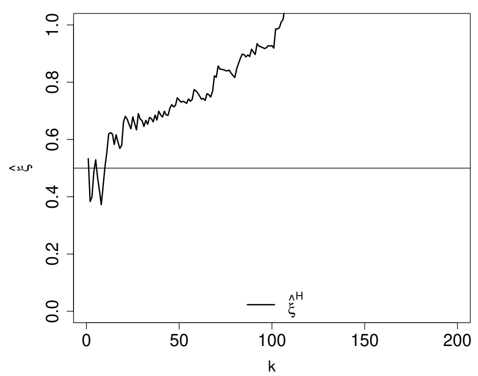


Figure 1: Hill plot of a sample of size $n = 200$ from a Burr model with $a = 1$ and $b = 2$. The EVI is equal to 0.5 (solid horizontal line).

Another popular approach consist in choosing an optimal $k$ level through the minimization of the asymptotic mean squared error (AMSE). This optimal value of $k$ strongly depends on the tail of the underlying model, or equivalently on the slowly varying function $l$ in (1.2). If the underlying model belongs to Hall's [33] sub-class of Pareto-type models with survival function,

$$\overline{F}(x) = \left(\frac{x}{c}\right)^{-\frac{1}{\xi}} \left(1 + \frac{\beta}{\rho}\left(\frac{x}{c}\right)^{\frac{\rho}{\xi}} + o\left(x^{\frac{\rho}{\xi}}\right)\right), \tag{1.6}$$

where $\xi > 0$ is the EVI, $c > 0$ is a first-order scale parameter and $\rho < 0$, $\beta \neq 0$ are second-order tail parameters, then the optimal $k$ ([34]) for the Hill estimator is given by

$$k_0 = \underset{k}{\operatorname{argmin}}\, \text{AMSE}\left(\hat{\xi}^H(k)\right) = \left(\frac{(1-\rho)^2}{(-2\rho)\beta^2}\right)^{\frac{1}{1-2\rho}} n^{\frac{-2\rho}{1-2\rho}}.$$

Note that this result cannot be directly used without the proper estimation of the vector of second-order parameters $(\rho, \beta)$. Moreover, the level $k_0$ leads to a biased estimate of the EVI. Other methods for selecting the threshold can be found in references [8, 13, 42].

The difficulty with the precise estimation of the EVI motivated researchers to introduce new estimators with better properties (see [5, 24, 39] for a non-exhaustive list of alternative EVI-estimators). An important alternative is the Moment estimator, proposed in Dekkers *et al.* [22]. This estimator is valid for any real EVI and is given by

$$\hat{\xi}^{\mathrm{M}}(k) := M_n^{(1)}(k) + 1 - \frac{1}{2}\left(1 - \frac{\left(M_n^{(1)}(k)\right)^2}{M_n^{(2)}(k)}\right)^{-1} = M_n^{(1)}(k) + \frac{1}{2}\left\{1 - \left(\frac{M_n^{(2)}(k)}{\left(M_n^{(1)}(k)\right)^2} - 1\right)^{-1}\right\}, \quad k = 1, 2, \ldots, n \tag{1.7}$$

where

$$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\ln X_{n-i+1:n} - \ln X_{n-k:n}\right)^j, \quad j > 0, \quad (M_n^{(1)}(k) \equiv \hat{\xi}^{\mathrm{H}}(k)), \tag{1.8}$$

are the moments of order $j$ of the log-excesses. It should be emphasized that the Moment estimator can also be sensitive to the choice of $k$ and, when $\xi > 0$, it has a greater asymptotic variance than the Hill estimator. Additional information about the distributional behaviour of the Moment estimator is available in references [9, 20, 32].

The reduction of bias is another topic that has deserved attention of the research community. This line of research started in the late nineties and is still important today. The first reduced bias (RB) EVI-estimators, usually called second-order reduced-bias estimators, had the classical bias-variance trade-off problem: the estimators had a null asymptotic bias but a larger asymptotic variance than the Hill estimator. Such a trade-off problem was overcome with an adequate estimation of second-order parameters leading to the so-called minimum-variance reduced-bias (MVRB) estimators. Gomes *et al.* [28] considered models with tail function in (1.6) and proposed the following MVRB weighted estimator of the log-excesses

$$\hat{\xi}^{WLE}(k) := \frac{1}{k}\sum_{i=1}^{k}\exp\left\{-\hat{\beta}\left(\frac{n}{k}\right)^{\hat{\rho}}\hat{\psi}_{ik}\right\}(\ln X_{n-i+1:n} - \ln X_{n-k:n}), \quad k = 1, 2, \ldots, n-1, \tag{1.9}$$

where

$$\hat{\psi}_{ik} = -\frac{(i/k)^{\hat{\rho}} - 1}{\hat{\rho}\ln(i/k)}, \quad 1 \leq i \leq k,$$

and $\hat{\rho}$ and $\hat{\beta}$ are adequate consistent estimators of the second-order tail parameters $\rho$ and $\beta$, respectively. Other MVRB EVI-estimators can be found in references [12, 15, 16, 28, 31, 35]. Moreover, MVRB estimators require a regular variation condition more restrictive than (1.1) and their performance may deteriorate in smaller samples if the limit relation (1.1) has a small rate of convergence. Thus, further research work on the estimation of the EVI is needed.

To construct new estimators, we shall consider the following statistic

$$T_n(k) := \frac{1}{k}\sum_{i=1}^{k} w\left(\frac{i}{k+1}\right)(\ln X_{n-i+1:n} - \ln X_{n-k:n})^{\alpha}, \quad k = 1, 2, \ldots, n-1, \tag{1.10}$$

for some real parameter $\alpha > 0$ and some weight function $w$ such that $\int_0^1 w(1-u) \times [-\ln(1-u)^{\alpha}]du = 1$. This type of statistic was already considered in Ciuperca and Mercadier [18] and Hüsler *et al.*[37], among others, and belongs to the general class of kernel estimators studied by Csörgő et al. [19]. However, in this work, we do not assume a strictly positive weight function $w$. Here we shall restrict the weight function $w$ to a linear transformation of the empirical probabilities, $b_0 + b_1 i/(k+1)$ with $b_0$ and $b_1$ real parameters, and consider $\alpha = 1$ in (1.10). To make the weight function satisfy the aforementioned condition we need to impose the restriction $b_1 = -4(b_0 - 1) = -4a,\ a \in \mathbf{R}$. Therefore, the class of estimators introduced in the present paper is given by,

$$\hat{\xi}^{WH(a)}(k) := \frac{1}{k}\sum_{i=1}^{k} w_a\left(\frac{i}{k+1}\right)(\ln X_{n-i+1:n} - \ln X_{n-k:n}), \quad k = 1, 2, \ldots, n-1, \tag{1.11}$$

where

$$w_a\left(\frac{i}{k+1}\right) = 1 + a - 4a\frac{i}{k+1} \tag{1.12}$$

and $a \in \mathbf{R}$ denotes a tuning parameter. This class of estimators generalizes the Hill estimator ($a = 0$) in (1.4). The parameter $a$ may be used to control the distributional behaviour of the estimator and eventually to achieve a high efficiency relatively to the Hill estimator.

The remainder of this paper is organized as follows: Section 2 describes the theoretical properties of the estimators. Numerical performance is studied in Section 3. Finally, some concluding remarks are presented in Section 4.

# 2 Asymptotic Properties of the estimators

## 2.1 Second-order regular variation

To derive the asymptotic results, we need an additional assumption about the underlying distribution. It is well-known in extreme value theory that in the case of Pareto-type models condition (1.1) is equivalent to the following first-order regular variation representation

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^{\xi}, \tag{2.1}$$

for every $x > 0$, with $U(t) = F^{\leftarrow}(1 - 1/t)$, $t \geq 1$, the tail quantile function and $F^{\leftarrow}(t) = \inf\{x : F(x) \geq t\}$, $0 < t < 1$, the left-continuous generalized inverse function of the d.f. $F$.

To obtain the asymptotic bias of the EVI-estimators under consideration, we need to assume a second-order condition of regular variation. We thus assume the existence of a ultimately positive regular varying function $A$ with index $\rho \le 0$, i.e. $|A(t)| \in RV_\rho$ ([26, 20]), such that

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \xi \ln x}{A(t)} = \begin{cases} \frac{x^\rho - 1}{\rho}, & \text{if } \rho < 0, \\ \ln x, & \text{if } \rho = 0, \end{cases} \tag{2.2}$$

for every $x > 0$. The second-order parameter $\rho$ measures the rate of convergence of $U(tx)/U(t)$ to $x^\xi$. The smaller $\rho$ is, the faster $A(t)$ converges to 0 and smaller will be the deviation between the right tail of the underlying model and the strict Pareto distribution. This condition is not very restrictive since it is satisfied by the most common Pareto-type distributions. Hall's [33] class of Pareto-type models, with d.f. in (1.6), satisfies the second-order condition with $A(t) = \xi\beta t^\rho$, $\rho < 0$. This class includes, among others, the Fréchet, the Burr, Student's t, and the Generalized Pareto distributions.

**Remark 2.1** (Estimation of the second-order parameters)**.** *To use any MVRB EVI-estimator, such as the one in* (1.9)*, it is necessary to estimate the second-order tail parameters $\rho$ and $\beta$. Here we shall consider particular members of the class of estimators for the second-order parameter $\rho$ proposed in [25]. Such a class is defined as follow*

$$\hat\rho^{(\tau)}(k) := -\left|\frac{3\big(\mathrm{R}(k;\tau)-1\big)}{\big(\mathrm{R}(k;\tau)-3\big)}\right|, \quad \textit{with} \quad \mathrm{R}(k;\tau) := \begin{cases} \dfrac{\big(M_n^{(1)}(k)\big)^\tau - \big(M_n^{(2)}(k)/2\big)^{\tau/2}}{\big(M_n^{(2)}(k)/2\big)^{\tau/2} - \big(M_n^{(3)}(k)/6\big)^{\tau/3}}, & \textit{if } \tau \ne 0, \\ \dfrac{\ln\big(M_n^{(1)}(k)\big) - \frac12 \ln\big(M_n^{(2)}(k)/2\big)}{\frac12 \ln\big(M_n^{(2)}(k)/2\big) - \frac13 \ln\big(M_n^{(3)}(k)/6\big)}, & \textit{if } \tau = 0, \end{cases} \tag{2.3}$$

*and $M_n^{(j)}(k)$ defined in* (1.8)*. Consistency and asymptotic normality of the estimators in* (2.3) *were proved in [25]. The theoretical and simulated results in [25, 11, 17], together with their use in* RB *estimation, lead us to suggest the use of $\tau = 0$ for $\rho \in [-1, 0)$ and $\tau = 1$ for $\rho \in (-\infty, -1)$. Other estimators of the shape second-order parameter $\rho$ can be found in [11, 27, 18, 21, 36], among others. For the estimation of the parameter $\beta$, we consider the estimators introduced in [30],*

$$\hat\beta(k) = \hat\beta_{\hat\rho^{(\tau)}(k)}(k) := \left(\frac{k}{n}\right)^{\hat\rho} \frac{\left\{\left(\frac1k \sum_{i=1}^k \left(\frac ik\right)^{-\hat\rho}\right)\left(\frac1k \sum_{i=1}^k S_i\right) - \left(\frac1k \sum_{i=1}^k \left(\frac ik\right)^{-\hat\rho} S_i\right)\right\}}{\left(\frac1k \sum_{i=1}^k \left(\frac ik\right)^{-\hat\rho}\right)\left(\frac1k \sum_{i=1}^k \left(\frac ik\right)^{-\hat\rho} S_i\right) - \left(\frac1k \sum_{i=1}^k \left(\frac ik\right)^{-2\hat\rho} S_i\right)}. \tag{2.4}$$

*where $S_i := i\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\}$, $1 \le i \le k < n$ are the scaled log-spacings. The asymptotic behaviour of $\hat\beta_\rho(k)$ was obtained in [30] under the second-order framework in* (2.2)*. If* (2.2) *holds, with $A(t) = \xi\beta t^\rho$ , $\rho < 0$, $k = k_n$ is intermediate such that $\sqrt{k}A(n/k) \to \infty$, and assuming $\hat\rho - \rho = o_p(1/\ln n)$, then $\hat\beta(k)$ is consistent for $\beta$. In applications, we advise the estimation of $\rho$*

*and $\beta$ at the intermediate level $k_1 = \lfloor n^{1-\epsilon} \rfloor$ for some small $\epsilon > 0$, with $\lfloor x \rfloor$ denoting the integer part of $x$. The choice of $\epsilon$ is not crucial, and as a compromise between theoretical and practical considerations we take $\epsilon = 0.01$ ([10, 14]).*

## 2.2 Asymptotic normality of the estimators

In the following, we assume that the second-order condition in (2.2) is satisfied. For theoretical comparison of estimators, we first provide without proof the non-degenerated behaviour of the estimators previously known from the literature and mentioned in this paper.

**Theorem 2.1.** *Under the validity of the second-order condition in* (2.2)*, and for intermediate sequences $k = k_n$, the EVI-estimators $\hat{\xi}^H(k)$ and $\hat{\xi}^M(k)$, respectively defined in* (1.4) *and* (1.7)*, generally denoted by $\hat{\xi}^\bullet(k)$, the asymptotic distributional identity*

$$\hat{\xi}^\bullet(k) \stackrel{d}{=} \xi + \frac{\sigma_\bullet}{\sqrt{k}} Z_k^\bullet + b_\bullet A(n/k) + o_p(A(n/k)), \tag{2.5}$$

*holds, where $\stackrel{d}{=}$ denotes equality in distribution, $Z_k^\bullet$ is an asymptotically standard normal random variable (r.v.) and $b_\bullet$ and $\sigma_\bullet$ given by*

$$\begin{aligned} b_H &= \frac{1}{1-\rho}, \qquad \sigma_H^2 = \xi^2 \\ b_M &= \frac{\xi(1-\rho)+\rho}{\xi(1-\rho)^2}, \qquad \sigma_M^2 = 1+\xi^2. \end{aligned} \tag{2.6}$$

*Moreover, further assuming that the second-order parameters $\rho$ and $\beta$ are estimated at a higher level $k_1 = O(n^{1-\epsilon})$, with $\epsilon$ small, $(\hat{\rho}-\rho)\ln n = o_p(1)$ and $(\hat{\beta}-\beta)/\beta \stackrel{d}{=} -(\hat{\rho}-\rho)\ln(n/k_1)(1+o_p(1))$, a condition that holds for several estimators of the parameter $\beta$ and $A(t) = \xi\beta t^\rho$, $\rho < 0$, the asymptotic distributional identity in* (2.5) *also holds for $\hat{\xi}^{WLE}(k)$ in* (1.9) *with*

$$b_{WLE} = 0, \quad \sigma_{WLE}^2 = \sigma_H^2 = \xi^2.$$

**Theorem 2.2.** *Under the validity of the second-order condition in* (2.2) *and assuming $k$ is an intermediate sequence of positive integers, the asymptotic distributional expansion*

$$\hat{\xi}^{WH(a)}(k) \stackrel{d}{=} \xi + \frac{\sigma_a}{\sqrt{k}} Z_k^{(a)} + b(\rho,a)A(n/k)(1+o_p(1)), \tag{2.7}$$

*holds, where $Z_k^{(a)} \sim N(0,1)$ is an asymptotically standard normal r.v.,*

$$\sigma_a^2 = \xi^2 \left(1 + \frac{a^2}{3}\right), \tag{2.8}$$

*and*

$$b(\rho,a) = \frac{1+a}{1-\rho} - \frac{2a}{2-\rho}. \tag{2.9}$$

*Consequently, if we choose $k$ such that $\sqrt{k}\,A(n/k) \to \lambda \in \mathbf{R}$,*

$$\sqrt{k}(\hat{\xi}^{WH(a)}(k) - \xi) \stackrel{d}{\longrightarrow} N\left(\lambda\, b(\rho,a), \sigma_a^2\right), \qquad as \quad n \to \infty. \tag{2.10}$$

*Proof.* Let $Y_{i:n}$, $1 \le i \le n$ be the set of ascending order statistics of $n$ i.i.d. standard pareto r.v.'s with distribution function $1 - 1/y$, $y > 1$ and let $U(t)$ denote the tail quantile function of the r.v. $X$. Since $X_{i:n} \stackrel{d}{=} U(Y_{i:n})$, $1 \le i \le n$ and $Y_{n-i+1:n}/Y_{n-k:n} \stackrel{d}{=} Y_{k-i+1:k}$, $1 \le i \le k$, we may write

$$\ln \frac{X_{n-i+1:n}}{X_{n-k:n}} \stackrel{d}{=} \ln \frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})} = \ln \frac{U(\frac{Y_{n-i+1:n}}{Y_{n-k:n}} Y_{n-k:n})}{U(Y_{n-k:n})} \stackrel{d}{=} \ln \frac{U(Y_{k-i+1:k} Y_{n-k:n})}{U(Y_{n-k:n})}.$$

Next, under the second-order framework in (2.2) we obtain

$$\ln \frac{X_{n-i+1:n}}{X_{n-k:n}} \stackrel{d}{=} \xi \ln(Y_{k-i+1:k}) + \frac{Y_{k-i+1:k}^{\rho} - 1}{\rho} A(Y_{n-k:n})(1 + o_p(1)).$$

Furthermore, since $A \in RV_\rho$ and $\frac{k}{n} Y_{n-k:n} \stackrel{p}{\to} 1$ we have $A(Y_{n-k:n})/A(n/k) \stackrel{p}{\to} 1$, as $n \to \infty$. Consequently, from the fact that $E_{i:k} \stackrel{d}{=} \ln Y_{i:k}$ where $E_1, E_2, \ldots, E_k$ are independent exponentially distributed r.v.'s with mean value 1, we obtain

$$\hat{\xi}^{WH(a)}(k) \stackrel{d}{=} \frac{1}{k} \sum_{i=1}^{k} w_a\left(\frac{i}{k+1}\right) \xi E_{k-i+1:k} + \frac{1}{k} \sum_{i=1}^{k} w_a\left(\frac{i}{k+1}\right) \frac{(Y_{k-i+1:k})^{\rho} - 1}{\rho} A(n/k)(1 + o_p(1))$$

where $w_a(\cdot)$ is defined in (1.12). Using the asymptotic result for linear functions of order statistics (Arnold *et al.* [2], theorem 8.6.1), with the notation

$$L_k^{(a)} := \frac{1}{k} \sum_{i=1}^{k} J\left(\frac{i}{k+1}\right) E_{i:k} = \frac{1}{k} \sum_{i=1}^{k} \xi w_a\left(1 - \frac{i}{k+1}\right) E_{i:k} = \frac{1}{k} \sum_{i=1}^{k} w_a\left(\frac{i}{k+1}\right) \xi E_{k-i+1:k},$$

we have

$$Z_k^{(a)} := \sqrt{k}\, \frac{L_k^{(a)} - \mu_a}{\sigma_a} \underset{n\to\infty}{\stackrel{d}{\longrightarrow}} N(0,1), \tag{2.11}$$

with

$$\mu_a := \int_0^\infty x J(1 - e^{-x}) e^{-x} dx = \int_0^\infty \xi x w_a(e^{-x}) e^{-x} dx = \xi, \tag{2.12}$$

and

$$\sigma_a^2 = 2 \iint_{0<x<y<\infty} J(1-e^{-x}) J(1-e^{-y}) [(1-e^{-x}) e^{-y}] dx dy = 2 \iint_{0<x<y<\infty} \xi^2 w_a(e^{-x}) w_a(e^{-y}) [(1-e^{-x}) e^{-y}] dx dy,$$

equal to the result given in (2.8). Moreover, applying again the asymptotic result for linear functions of order statistics, the following result hold

$$\frac{1}{k} \sum_{i=1}^{k} w_a\left(\frac{i}{k+1}\right) \frac{(Y_{k-i+1:k})^{\rho} - 1}{\rho} \stackrel{p}{\longrightarrow} b(\rho, a)$$

with $b(\rho, a)$ given in (2.9) and the asymptotic distributional expansion in (2.7) follows straightforwardly. Finally, the limit result (2.10) is immediate since

$$\sqrt{k}\left(\hat{\xi}^{WH(a)}(k) - \xi\right) \stackrel{d}{=} \sigma_a Z_k^{(a)} + \sqrt{k} b(\rho, a) A(n/k)(1 + o_p(1)). \tag{2.13}$$

□

Note that, for a fixed $\rho < 0$, the function $b(\rho, a)$ in (2.9) is strictly decreasing and unbounded. Hence it is advisable to choose the value of the tuning parameter $a$ in a limited interval of the real line. Also, whenever $\rho < 0$, there is always a value $a_0$ that nullifies the asymptotic bias component in (2.9), even when $\sqrt{k}\, A(n/k) \to \lambda \neq 0$, finite. Such value $a_0$ is always positive and explicitly given by

$$a_0 = a_0(\rho) = \frac{\rho - 2}{\rho}, \quad \rho < 0. \tag{2.14}$$

Consequently, if the aim is the reduction of bias, it is thus advised to only use positive values of the tuning parameter $a$. In table 1 we present the value of $a_0$, with 2 decimal places, for some selected values of $\rho$.

Table 1: Values of $a_0$ for some selected values of $\rho$

| $\rho$ | $-3.00$ | $-2.50$ | $-2.00$ | $-1.50$ | $-1.25$ | $-1.00$ | $-0.75$ | $-0.25$ | $-0.10$ | $-0.05$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $a_0$ | 1.67 | 1.8 | 2 | 2.33 | 2.6 | 3 | 3.67 | 9 | 21 | 41 |

In Figure 2 we illustrate the sample path of $\hat{\xi}^H(k)$ and of $\hat{\xi}^{WH(3)}(k)$, for one sample of size $n = 2000$ from the Burr model, in (1.3), with $a = 2$ and $b = 1$, to which corresponds $\xi = 0.5$ and $\rho = -1$. For such a sample and values of $k$ smaller than 800, the sample path of $\hat{\xi}^{WH(3)}(k)$ shows generally a small bias.

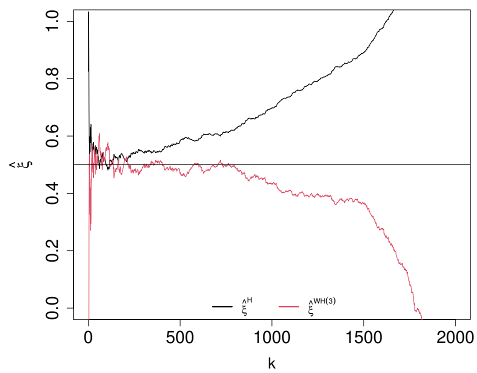


Figure 2: Hill and Weighted Hill plot for one generated sample of size $n = 2000$ from a Burr model with $a = 2$, $b = 1$ $\xi = 0.5$ and $\rho = -1$ .

The asymptotic variance component $\sigma_a^2$ in (2.8) is a convex function with a global minimum equal to $\xi^2$ at $a = 0$. The plots of the asymptotic bias and variance components, $b(\rho, a)$ and $\sigma_a^2/\xi^2$, as a function of $a \in [0, 6]$, are provided in Figure 3.

**Remark 2.2.** *According with the asymptotic distributional representation in* (2.7)*, the Asymptotic Mean Squared Error (AMSE) is*

$$AMSE\left(\hat{\xi}^{WH(a)}(k)\right) = \frac{\sigma_a^2}{k} + b^2(\rho, a) A^2(n/k)$$

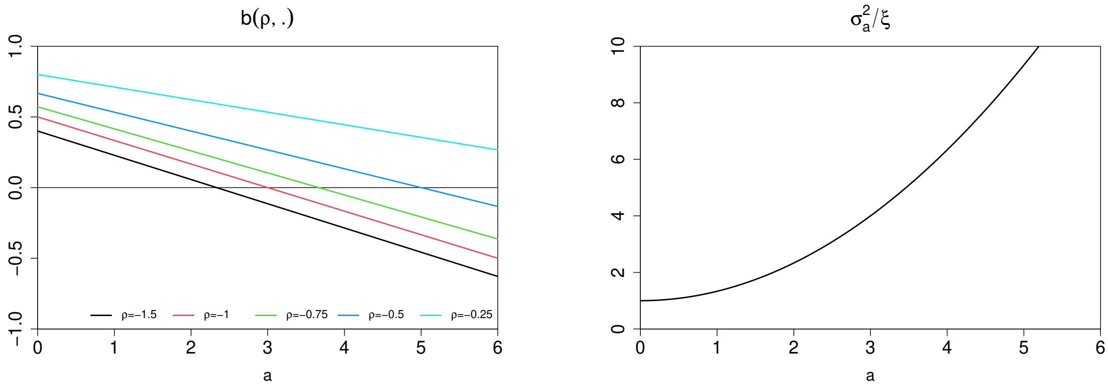


Figure 3: The asymptotic bias and variance components $b(\rho, a)$ (left) and $\sigma_a^2$ (right), as a function of the tuning parameter $a$, with $0 \le a \le 6$.

*Figure 4 illustrate the behaviour of the AMSE for a sample of size* $n = 1000$ *from a model under the second-order condition in* (2.2) *with* $\xi = 1$, $\rho = -1$ *and* $A(t) = 1/t$. *In this example and with an adequate choice of the level* $k$, $\hat{\xi}^{WH(a)}(k)$ *can achieve a smaller AMSE than the Hill estimator.*

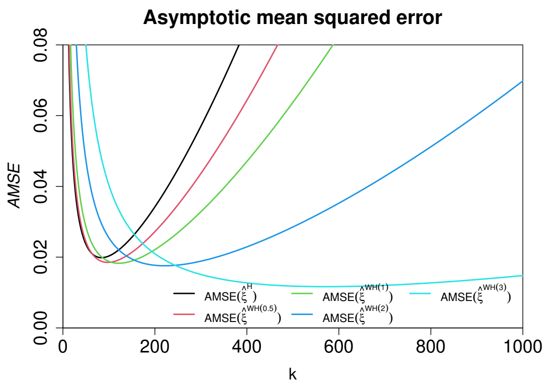


Figure 4: Plot of the Asymptotic Mean Squared Error of $\hat{\xi}^H(k)$ and $\hat{\xi}^{WH(a)}(k)$, $a = 0.5, 1, 2, 3$ as a function of $k$ ($n = 1000$) for a model under the second order condition with $\xi = 1$ and $\rho = -1$ and $A(t) = 1/t$.

**Remark 2.3.** *Let* $F$ *belong to the sub-class of Pareto-type models with survival function given in* (1.6). *Through the minimization of the AMSE, the optimal value of* $k$ *to consider in the estimator* (1.11) *is*

$$k_0^{WH(a)} = \operatorname*{argmin}_k AMSE\left(\hat{\xi}^{WH(a)}(k)\right) = \left(\frac{(1-\rho)^2(2-\rho)^2(1+a^2/3)^2}{(-2\rho)(2-\rho+a\rho)^2\beta^2}\right)^{\frac{1}{1-2\rho}} n^{\frac{-2\rho}{1-2\rho}}, \qquad \text{if} \quad a \neq \frac{\rho-2}{\rho}.$$

*Moreover, if* $a = (\rho - 2)/\rho$ *the dominant component of asymptotic bias, in* (2.9), *is zero, and a third-order condition is needed to compute the optimal value of* $k$. *Such condition falls outside* *the scope of this paper. For further details on the third-order condition, for heavy tails, see [25, 10], among others. To close the gap between theory and applications to real data, we advise to choose $k$ as large as possible in the first stability region of the estimates of the $\hat{\xi}^{WH(a)}(k)$. Such region should be much easier to detect, than with the Hill estimator, if the tuning parameter $a$ is close to $(\rho-2)/\rho$.*

## 3 Finite sample properties of the estimators

This section provide some results based on a Monte Carlo simulation study to assess the finite-sample performance of all EVI-estimators considered in section 1 of this paper. For the class of weighted Hill estimators in (1.11) we considered the following values of the tuning parameter: $a = 1.5, 2, 2.5, 3$ and $3.5$. Notice that the simulated results for $\hat{\xi}^{WH(a)}(k)$ are not very relevant in practise, since the true value of the second-order parameter $\rho$ in unknown. Thus, we also advance with a data driven method for selecting the parameter $a$ and considered in this simulation study the estimator

$$\hat{\xi}^{WH^*}(k) = \hat{\xi}^{WH(\hat{a})}(k), \tag{3.1}$$

such that $\hat{a} = a_0(\hat{\rho})$ with $a_0(\rho)$ and $\hat{\rho}$ given respectively in (2.14) and (2.3). Numerical calculations were made in R software [40]. The distributions from which we will generate the samples are as follows:

- The Burr distribution in (1.3), conveniently parameterized so that $\xi = 0.5$ and $\rho = -0.75$;
- The Fréchet model, with d.f. $F(x) = 1 - \exp(-x^{-1/\xi}),\ x > 0,$ with $\xi = 0.5$ and $\rho = -1$;
- The Log-gamma model with d.f. $F(x) = 1 - x^{-1/\xi}(1 + \ln(x)/\xi),\ x > 1$ with $\xi = 0.5$. This distribution does not belong to Hall's class in (1.6), although the second-order condition in (2.2) holds, with $\rho = 0$.

The finite sample performance of the estimators are investigated with 5000 samples of size $n = 500$. For each sample, the estimates $\hat{\xi}^{\bullet}(k)$ are first computed for $k = 1, 2, \ldots, n-1$, next we computed the Monte Carlo estimates of the mean value (E) and root mean squared error (RMSE). The simulated mean value (left figure) and root mean squared error (right figure) for $n = 500$ and all possible values of the level $k$ are given in Figures 5, 6 and 7, for the Burr, the Fréchet and the log-gamma distributions.

For the Burr and the Fréchet models we observe that the empirical behaviour of the different estimators is very similar:

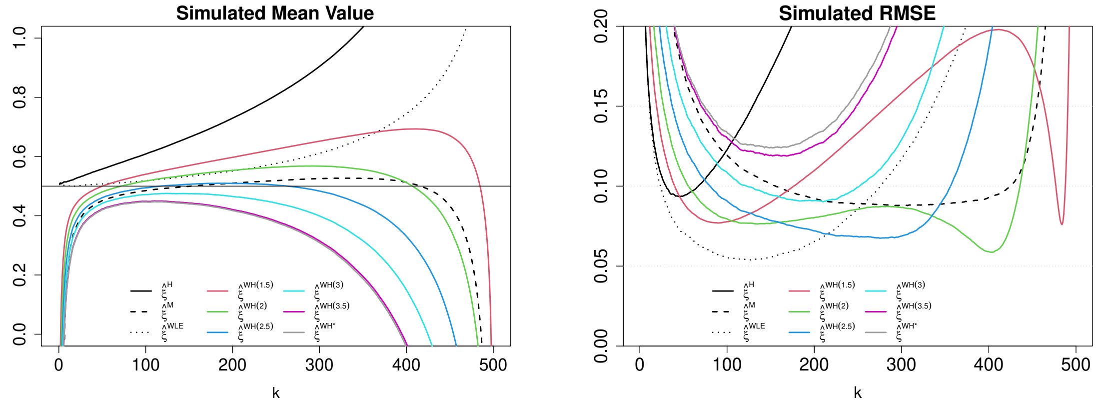


Figure 5: Simulated Mean values (left) and RMSEs (right) of the EVI-estimators under study for samples of size $n = 500$ from a Burr parent with $\xi = 0.5$ and $\rho = -0.75$.

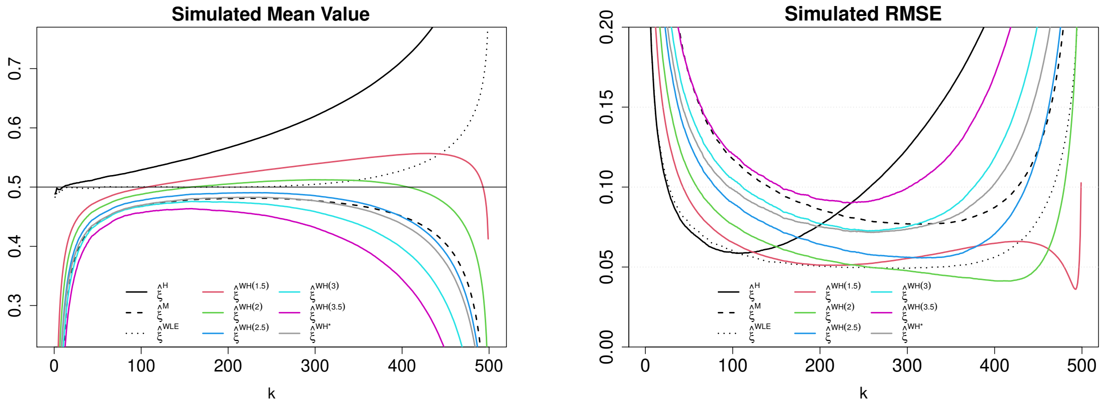


Figure 6: Simulated Mean values (left) and RMSEs (right) of the EVI-estimators under study for samples of size $n = 500$ from a Fréchet parent with $\xi = 0.5$.

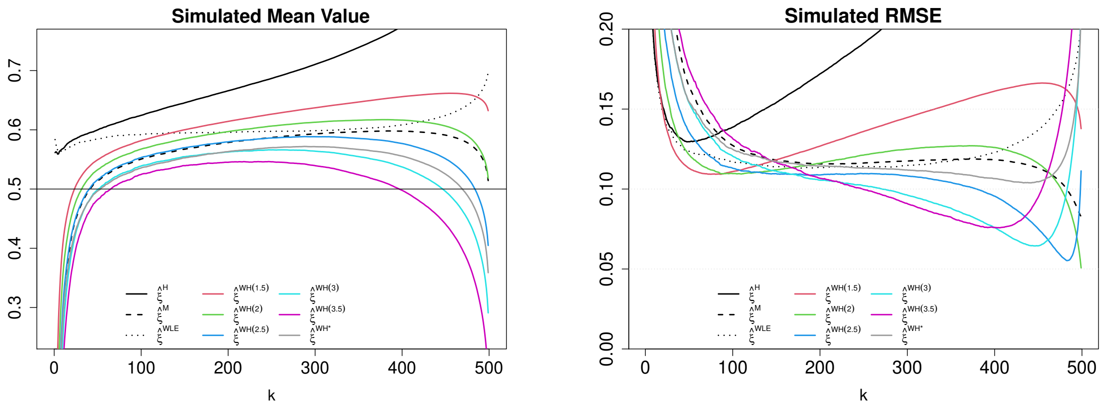


Figure 7: Simulated Mean values (left) and RMSEs (right) of the EVI-estimators under study for samples of size $n = 500$ from a log-gamma parent with $\xi = 0.5$.

- The Hill estimator has a positive increasing positive bias and a peaked RMSE, as a function of $k$; If $k < 50$ then the Hill estimator has one of the smallest RMSE. However, if $k$ is large, this estimator has the largest RMSE.

- The Moment estimator provides a small bias for values of $k$ between 100 and 400. Unfortunately the RMSE of this estimator is only (slightly) smaller than the RMSE of the Hill estimator for the Burr distribution. For the Fréchet model, the RMSE of the Moment estimator is always greater than the smallest RMSE of the Hill estimator.

- The WLE estimator, in (1.9), shows a good performance for both Burr and the Fréchet models. This estimator provides small bias and a similar or smaller RMSE than the Hill estimator. This confirms the theoretical variance and bias components of those estimators, presented in Theorem 2.1.

- The Weighted Hill estimator, with an adequate value for the tuning parameter $a$, is highly competitive against the other alternative EVI-estimators. In terms of bias it is the best estimator if $a = 2.5$. Regarding RMSE, this estimator is the second-best for the Burr model and the best one for the Fréchet model. Notice that for small values of the parameter $a$, the RMSE of the WH estimator can have a local or absolute minimum for $k$ close to the value $n$. Those values of the RMSE should be ignored because they result from a change in the bias signal and since they are not associated with a region of stability of the estimates, they are irrelevant in practical applications.

- Regarding the WH* estimator, in (3.1), the data-driven choice of the parameter $a$ provides good estimates in term of bias. In terms of RMSE there is no clear advantage over the Hill and WLE estimators. These results suggest, for future research work, the study of other data-driven choice of the parameter $a$.

For the log-gamma model, almost all estimators evidence a substantial bias and a high RMSE. Nevertheless, the Weighted Hill estimator with $a = 3$ or $a = 3.5$ outperform the other estimators. Since the log-gamma model does not belong to the class introduced by Hall, we have no theoretical support for such a choice of the parameter of the Weighted Hill estimator.

## 4 Conclusion

In this paper we introduced a new class of Weighted Hill estimators of the EVI. We analyzed the asymptotic limiting distribution of the new class of estimators assuming the validity of a second order framework and illustrated their performance with a Monte Carlo simulation study. A comparison with other important estimators from the literature was also provided. The results indicate that the new Weighted Hill estimator, with a suitable choice of the parameter $a$, is a reliable estimator and, for some models, can outperform several well known estimators.

## Acknowledgments

This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project UIDB/00297/2020 (Centro de Matemática e Aplicações).

## References

[1] B. C. Arnold. *Pareto distributions.* Monographs on Statistics and Applied Probability. Chapman and Hall/CRC, 2015.

[2] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *A first course in order statistics*, volume 54. Siam, 1992.

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

[4] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels. *Statistics of Extremes: Theory and Applications.* John Wiley & Sons, 2004.

[5] J. Beirlant, K. Herrmann, and J. Teugels. Estimation of the extreme value index. In *Extreme Events in Finance*, pages 97–115. John Wiley & Sons, Inc., Oct. 2016.

[6] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27. Cambridge university press, 1989.

[7] I. W. Burr. Cumulative frequency functions. *The Annals of Mathematical Statistics*, 13(2):215–232, June 1942.

[8] C. Caballero-Megido, J. Hillier, D. Wyncoll, L. Bosher, and B. Gouldby. Technical note: comparison of methods for threshold selection for extreme sea levels. *Journal of Flood Risk Management*, 11(2):127–140, 2017.

[9] I. Cabral, F. Caeiro, and M. I. Gomes. On the comparison of several classical estimators of the extreme value index. *Communications in Statistics - Theory and Methods*, pages 1–18, Apr. 2020.

[10] F. Caeiro and M. I. Gomes. Minimum-variance reduced-bias tail index and high quantile estimation. *Revstat – Statistical Journal*, 6(1):1–20, 2008.

[11] F. Caeiro and M. I. Gomes. Bias reduction in the estimation of a shape second-order parameter of a heavy tail model. *Journal of Statistical Computational and Simulation*, 85(17):3405–3419, 2015.

[12] F. Caeiro and M. I. Gomes. Revisiting the maximum likelihood estimation of a positive extreme value index. *Journal of Statistical Theory and Practice*, 9(1):200–218, 2015.

[13] F. Caeiro and M. I. Gomes. Threshold selection in extreme value analysis. In D. K. Dey and J. Yan, editors, *Extreme Value Modeling and Risk Analysis*, pages 69–86. Chapman and Hall/CRC, 2015.

[14] F. Caeiro, M. I. Gomes, and L. Henriques-Rodrigues. Reduced-bias tail index estimators under a third order framework. *Communications in Statistics–Theory and Methods*, 38(7):1019–1040, 2009.

[15] F. Caeiro, M. I. Gomes, and D. Pestana. Direct reduction of bias of the classical hill estimator. *Revstat – Statistical Journal*, 3(2):113–136, 2005.

[16] F. Caeiro, L. Henriques-Rodrigues, M. I. Gomes, and I. Cabral. Minimum-variance reduced-bias estimation of the extreme value index: A theoretical and empirical study. *Computational and Mathematical Methods*, 2(4), May 2020.

[17] F. Caeiro and D. Prata Gomes. Adaptive estimation of a tail shape second order parameter: A computational comparative study. In T. Simos, Z. Kalogiratou, and T. Monovasilis, editors, *AIP Conference Proceedings*, volume 1702, page 030005. AIP Publishing, 2015.

[18] G. Ciuperca and C. Mercadier. Semi-parametric estimation for heavy tailed distributions. *Extremes*, 13:55–87, 2010.

[19] S. Csörgő, P. Deheuvels, and D. Mason. Kernel estimates of the tail index of a distribution. *The Annals of Statistics*, 13(3):1050–1077, 1985.

[20] L. de Haan and A. Ferreira. *Extreme Value Theory*. Springer New York, 2006.

[21] T. de Wet, Y. Goegebeur, and M. R. Munch. Asymptotically unbiased estimation of a second order tail parameter. *Statistics & Probability Letters*, 82(3):565–573, 2012.

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

[23] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events: For Insurance and Finance*. Springer, New York, 1997.

[24] I. Fedotenkov. A review of more than one hundred pareto-tail index estimators. *Statistica*, 80(3):245–299, 2020.

[25] M. I. Fraga Alves, M. I. Gomes, and L. De Haan. A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica*, 60(2):193–213, 2003.

[26] J. Geluk and L. de Haan. *Regular Variation, Extensions and Tauberian Theorems.* CWI Tract 40. Centre for Mathematics and Computer Science, Amsterdam, The Netherlands, 1987.

[27] Y. Goegebeur, J. Beirlant, and T. De Wet. Kernel estimators for the second order parameter in extreme value statistics. *Journal of Statistical Planning and Inference*, 140(9):2632–2652, 2010.

[28] M. I. Gomes, L. de Haan, and L. H. Rodrigues. Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):31–52, 2008.

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

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

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

[32] M. I. Gomes and C. Neves. Asymptotic comparison of the mixed moment and classical extreme value index estimators. *Statistics & Probability Letters*, 78(6):643–653, Apr. 2008.

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

[34] P. Hall and A. Welsh. Adaptive estimates of parameters of regular variation. *The Annals of Statistics*, 13(1):331–341, 1985.

[35] L. Henriques-Rodrigues and M. I. Gomes. Location-invariant reduced-bias tail index estimation under a third-order framework. *Journal of Statistical Theory and Practice*, 12(2):206–230, July 2017.

[36] L. Henriques-Rodrigues, M. I. Gomes, M. I. Fraga Alves, and C. Neves. Port-estimation of a shape second-order parameter. *Revstat – Statistical Journal*, 12(3):299–328, 2014.

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

[38] C. Kleiber and S. Kotz. *Statistical size distributions in economics and actuarial sciences.*, volume 470. John Wiley & Sons, 2003.

[39] J. Panaretos and Z. Tsourti. Extreme value index estimators and smoothing alternatives: A critical review. In J. Panaretos, editor, *Stochastic Musings: Perspectives From the Pioneers of the Late 20th Century*, pages 141–160. Laurence Erlbaum, USA, 2003.

[40] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2020.

[41] S. I. Resnick. Heavy tail modeling and teletraffic data: special invited paper. *The Annals of Statistics*, 25(5):1805–1869, 1997.

[42] C. Scarrott and A. MacDonald. A review of extreme value threshold estimation and uncertainty quantification. *Revstat – Statistical Journal*, 10(1):33–60, 2012.

[43] S. K. Singh and G. S. Maddala. A function for size distribution of incomes. *Econometrica*, 44(5):963, Sept. 1976.