Computational Statistics and Data Analysis 57 (2013) 518–535

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A simple generalisation of the Hill estimator

M. Fátima Brilhante [a], M. Ivette Gomes [b,*], Dinis Pestana [b]

[a] CEAUL and DM, Universidade dos Açores, Portugal

[b] CEAUL and DEIO, FCUL, Universidade de Lisboa, Portugal

ARTICLE INFO

Article history:
Received 15 March 2012
Received in revised form 21 July 2012
Accepted 22 July 2012
Available online 24 July 2012

Keywords:
Bias estimation
Bootstrap methodology
Heavy tails
Optimal levels
Semi-parametric estimation
Statistics of extremes

ABSTRACT

The classical Hill estimator of a positive extreme value index (EVI) can be regarded as the logarithm of the geometric mean, or equivalently the logarithm of the mean of order $p = 0$, of a set of adequate statistics. A simple generalisation of the Hill estimator is now proposed, considering a more general mean of order $p \geq 0$ of the same statistics. Apart from the derivation of the asymptotic behaviour of this new class of EVI-estimators, an asymptotic comparison, at optimal levels, of the members of such class and other known EVI-estimators is undertaken. An algorithm for an adaptive estimation of the tuning parameters under play is also provided. A large-scale simulation study and an application to simulated and real data are developed.

1. The new class of estimators and scope of the paper

Let us consider a sample of size n of independent, identically distributed (i.i.d.) random variables (r.v.'s), $X_1, \ldots, X_n$, with a common distribution function (d.f.) F. Let us denote by $X_{1:n} \leq \cdots \leq X_{n:n}$ the associated ascending order statistics (o.s.'s) and let us assume that there exist sequences of real constants $\{a_n > 0\}$ and $\{b_n \in \mathbb{R}\}$ such that the maximum, linearly normalised, i.e. $(X_{n:n} - b_n)/a_n$, converges in distribution to a non-degenerate r.v. Then, the limit distribution is necessarily of the type of the general *extreme value* (EV) d.f., given by

$$\mathrm{EV}_\gamma(x) = \begin{cases} \exp(-(1+\gamma x)^{-1/\gamma}), & 1+\gamma x > 0 \text{ if } \gamma \neq 0 \\ \exp(-\exp(-x)), & x \in \mathbb{R} \text{ if } \gamma = 0. \end{cases} \quad (1)$$

The d.f. F is said to belong to the max-*domain of attraction* of EV_γ, and we use the notation $F \in \mathcal{D}_{\mathcal{M}}\left(\mathrm{EV}_\gamma\right)$. The parameter γ is the *extreme value index* (EVI), the primary parameter of extreme events.

Let us denote by RV_a the class of regularly varying functions at infinity, with an index of regular variation equal to $a \in \mathbb{R}$, i.e. positive measurable functions $g(\cdot)$ such that for all $x > 0$, $g(tx)/g(t) \to x^a$, as $t \to \infty$ (see Bingham et al., 1987). The EVI measures the heaviness of the right *tail function*

$$\overline{F}(x) := 1 - F(x),$$

and the heavier the right tail, the larger γ is. In this paper we work with Pareto-type underlying d.f.'s, with a positive EVI, or equivalently, models such that $\overline{F}(x) = x^{-1/\gamma}L(x)$, $\gamma > 0$, with $L \in \mathrm{RV}_0$, a slowly varying function at infinity, i.e. a regularly varying function with an index of regular variation equal to zero. These heavy-tailed models are quite common in many

* Correspondence to: Departamento de Estatística e Investigação Operacional, Faculdade de Ciências de Lisboa, Bloco C6, Piso 4, 1749-016 Lisboa, Portugal. Tel.: +351 217500041.

E-mail addresses: fbrilhante@uac.pt (M. Fátima Brilhante), ivette.gomes@fc.ul.pt (M. Ivette Gomes), dinis.pestana@fc.ul.pt (D. Pestana).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2012.07.019

areas of application, like computer science, telecommunications, insurance, finance, bibliometrics and biostatistics, among others.

For Pareto-type models, the classical EVI-estimators are the Hill estimators (Hill, 1975), which are the averages of the log-excesses, given by

$$V_{ik} := \ln \frac{X_{n-i+1:n}}{X_{n-k:n}}, \quad 1 \le i \le k < n. \tag{2}$$

We thus have

$$\hat{\gamma}_n^H(k) \equiv H(k) := \frac{1}{k}\sum_{i=1}^{k} V_{ik}, \quad 1 \le k < n. \tag{3}$$

Note that with $F^{\leftarrow}(x) := \inf\{y : F(y) \ge x\}$ denoting the generalised inverse function of F, and

$$U(t) := F^{\leftarrow}(1 - 1/t), \quad t \ge 1,$$

the reciprocal quantile function, we can write the distributional identity $X = U(Y)$, with Y a unit Pareto r.v., i.e. a r.v. with d.f. $F_Y(y) = 1 - 1/y, y \ge 1$. For the o.s.'s associated with a random Pareto sample $(Y_1, \ldots, Y_n)$, we have the distributional identity $Y_{n-i+1:n}/Y_{n-k:n} = Y_{k-i+1:k}$, $1 \le i \le k$. Moreover, $kY_{n-k:n}/n \xrightarrow[n\to\infty]{p} 1$, i.e. $Y_{n-k:n} \overset{p}{\sim} n/k$. Consequently, and provided that $k = k_n$, $1 \le k < n$, is an intermediate sequence of integers, i.e. if

$$k = k_n \to \infty \quad \text{and} \quad k_n = o(n), \text{ as } n \to \infty, \tag{4}$$

we get

$$U_{ik} := \frac{X_{n-i+1:n}}{X_{n-k:n}} = \frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})} = \frac{U(Y_{n-k:n}Y_{k-i+1:k})}{U(Y_{n-k:n})} = Y_{k-i+1:k}^{\gamma}(1 + o_p(1)), \tag{5}$$

i.e. $U_{ik} \overset{p}{\sim} Y_{k-i+1:k}^{\gamma}$. Hence, we have the approximation $\ln U_{ik} \approx \gamma \ln Y_{k-i+1:k} = \gamma E_{k-i+1:k}$, $1 \le i \le k$, with E denoting a standard exponential r.v. The log-excesses, $V_{ik} = \ln U_{ik}$, $1 \le i \le k$, in (2), are thus approximately the k top o.s.'s of a sample of size k from an exponential parent with mean value γ. This justifies the Hill EVI-estimator, in (3).

We can write

$$H(k) = \sum_{i=1}^{k} \ln\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{1/k} = \ln\left(\prod_{i=1}^{k}\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{1/k}, \quad 1 \le i \le k < n,$$

the logarithm of the *geometric mean* of the statistics U_{ik}, given in (5). More generally, we now consider as basic statistics for the EVI estimation, the *mean of order p* (MOP) of U_{ik}, i.e. the class of statistics

$$A_p(k) = \begin{cases} \left(\dfrac{1}{k}\displaystyle\sum_{i=1}^{k} U_{ik}^p\right)^{1/p} & \text{if } p > 0 \\ \left(\displaystyle\prod_{i=1}^{k} U_{ik}\right)^{1/k} & \text{if } p = 0. \end{cases} \tag{6}$$

From (5), we can write $U_{ik}^p = Y_{k-i+1:k}^{\gamma p}(1 + o_p(1))$. Since

$$\mathbb{E}(Y^a) = \frac{1}{1-a} \quad \text{if } a < 1, \tag{7}$$

the law of large numbers enables us to say that if $p < 1/\gamma$,

$$A_p(k) \xrightarrow[n\to\infty]{p} \left(\frac{1}{1-\gamma p}\right)^{1/p}, \quad \text{i.e. } \frac{1 - A_p^{-p}(k)}{p} \xrightarrow[n\to\infty]{p} \gamma.$$

Hence the reason for the new class of MOP EVI-estimators,

$$\hat{\gamma}_n^{H_p}(k) \equiv H_p(k) := \begin{cases} \left(1 - A_p^{-p}(k)\right)/p & \text{if } p > 0 \\ \ln A_0(k) = H(k) & \text{if } p = 0, \end{cases} \tag{8}$$

with $A_p(k)$ given in (6), and with $H_0(k) \equiv H(k)$, given in (3). This class of MOP EVI-estimators depends on this tuning parameter $p \ge 0$, which makes it very flexible, and even able to overpass one of the simplest and one of the most efficient EVI-estimators in the literature, the corrected-Hill (CH) estimator in Caeiro et al. (2005), to be introduced in Section 2.2.

In this paper, after the introduction, in Section 2, of a few technical details in the field of *extreme value theory* (EVT) and a brief bibliographical study on various types of EVI-estimators, we deal in Section 3 with the asymptotic behaviour of the new class of MOP EVI-estimators, in (8). In Section 4 we compare asymptotically, at optimal levels, a large set of alternative classes of EVI-estimators, drawing some concluding remarks. In Section 5, we provide a method for the adaptive choice of the tuning parameters k and p, on the basis of the bootstrap methodology. Section 6 is dedicated to the finite sample properties of the new class of estimators comparatively with the behaviour of the aforementioned *CH* EVI-estimators, done through a large-scale simulation study. Finally, in Section 7, we illustrate the behaviour of the new class of MOP EVI-estimators, together with the adaptive choices provided in Sections 5 and 6, through an application to simulated random samples, and to sets of real data in the fields of insurance, finance and environment.

2. Preliminary results in the area of EVT

After a reference, in Section 2.1, to the most common first and second-order frameworks for heavy-tailed models, we briefly review, in Sections 2.2 and 2.3, the most popular EVI-estimators. Finally, in Section 2.4, we provide details on the asymptotic behaviour of those EVI-estimators. Recent reviews on the topic can be found in Beirlant et al. (2012) and Scarrott and MacDonald (2012).

2.1. A brief note on first and second-order conditions

In the area of *statistics of extremes* and whenever working with large values, a model F is usually said to be *heavy-tailed* whenever the right tail function $\overline{F}$ is a *regularly varying function* with a negative index of regular variation equal to $-1/\gamma$, $\gamma > 0$, or equivalently, the reciprocal quantile function U is of regular variation with an index γ, i.e.

$$F \in \mathcal{D}_{\mathcal{M}}^{+} := \mathcal{D}_{\mathcal{M}}\left(\mathrm{EV}_{\gamma}\right)_{\gamma>0} \Longleftrightarrow \overline{F} \in \mathrm{RV}_{-1/\gamma} \Longleftrightarrow U \in \mathrm{RV}_{\gamma}. \tag{9}$$

The first condition, in (9), was proved in Gnedenko (1943) and the second one in de Haan (1984).

The *second-order parameter* ρ (≤ 0) rules the rate of convergence in the first-order condition, in (9), and it is the non-positive parameter appearing in the limiting relation

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \begin{cases} \left(x^{\rho}-1\right)/\rho & \text{if } \rho < 0 \\ \ln x & \text{if } \rho = 0, \end{cases} \tag{10}$$

which is assumed to hold for every $x > 0$, and where $|A|$ must then be of regular variation with index ρ (Geluk and de Haan, 1987). This condition has been widely accepted as an appropriate condition to specify the right-tail of a Pareto-type distribution in a semi-parametric way. For reduced-bias estimators, and for technical simplicity, we often assume that we are working in the Hall–Welsh class of models (Hall and Welsh, 1985), with a right tail function,

$$\overline{F}(x) = \left(x/C\right)^{-1/\gamma}\left(1+\beta(x/C)^{\rho/\gamma}/\rho + o(x^{\rho/\gamma})\right), \quad \text{as } x \to \infty,$$

$C > 0$, $\beta \neq 0$ and $\rho < 0$. Equivalently, we can say that, with (β, ρ) a vector of second-order parameters, the general second-order condition in (10) holds with $A(t) = \gamma\beta t^{\rho}$, $\rho < 0$. Also,

$$U(t) = C\, t^{\gamma}\left(1+\gamma\,\beta\, t^{\rho}/\rho + o(t^{\rho})\right), \quad \text{as } t \to \infty. \tag{11}$$

Models like the log-gamma and the log-Pareto ($\rho = 0$) are thus excluded from this class. The standard Pareto is also excluded. But most heavy-tailed models used in applications, like the EV_{γ}, the Fréchet and the Student's t d.f.'s, among others, belong to the Hall–Welsh class.

2.2. Explicit EVI-estimators

Due to its simplicity, the most popular EVI-estimator, valid only for $\gamma \geq 0$, is the Hill estimator in (3). Apart from the Hill estimator, and with the notation

$$M_{k,n}^{(j)} := \frac{1}{k}\sum_{i=1}^{k} V_{ik}^{j}, \qquad L_{k,n}^{(j)} := 1 - \frac{1}{k}\sum_{i=1}^{k}\left(1-\frac{X_{n-k:n}}{X_{n-i+1:n}}\right)^{j}, \quad j \geq 1, \tag{12}$$

with V_{ik}, $1 \leq i \leq k$, defined in (2), we also consider in the asymptotic comparison at optimal levels performed in Section 4, the following classes of EVI-estimators.

- The moment (*Mo*) estimator (Dekkers et al., 1989), given by

$$\widehat{\gamma}_n^{Mo}(k) \equiv Mo(k) := M_{k,n}^{(1)} + \frac{1}{2}\left\{1-\left(M_{k,n}^{(2)}/\left(M_{k,n}^{(1)}\right)^2-1\right)^{-1}\right\}. \tag{13}$$

- The generalised Hill (GH) estimator (Beirlant et al., 1996), based on the Hill estimator in (3) and with the functional form

$$\widehat{\gamma}_n^{GH}(k) \equiv GH(k) := \hat{\gamma}_n^H(k) + \frac{1}{k}\sum_{i=1}^{k}\left\{\ln \hat{\gamma}_n^H(i) - \ln \hat{\gamma}_n^H(k)\right\}, \tag{14}$$

further studied in Beirlant et al. (2005).
- The mixed moment (MM) estimator (Fraga Alves et al., 2009), based on the statistics $M_{k,n}^{(1)}$ and $L_{k,n}^{(1)}$ in (12), and given by

$$\widehat{\gamma}_n^{MM}(k) \equiv MM(k) := \frac{\hat{\varphi}_{k,n} - 1}{1 + 2\min\left(\hat{\varphi}_{k,n} - 1, 0\right)}, \quad \text{with } \hat{\varphi}_{k,n} := \frac{M_{k,n}^{(1)} - L_{k,n}^{(1)}}{\left(L_{k,n}^{(1)}\right)^2}. \tag{15}$$

In the simulation study we consider

- the simplest class of CH EVI-estimators, the one introduced in Caeiro et al. (2005),

$$\widehat{\gamma}_n^{CH}(k) \equiv \widehat{\gamma}_{n,\hat{\beta},\hat{\rho}}^{CH}(k) \equiv CH(k) := \hat{\gamma}_n^H(k)\left(1 - \hat{\beta}(n/k)^{\hat{\rho}}/(1-\hat{\rho})\right). \tag{16}$$

The estimators in (16) can be second-order minimum-variance reduced-bias (MVRB) estimators, for adequate levels k and an adequate external estimation of the vector of second-order parameters, (β, ρ), in (11), i.e., the use of $\widehat{\gamma}_n^{CH}(k)$, and an adequate estimation of (β, ρ), enables the elimination of the dominant component of bias of the Hill estimator, $\widehat{\gamma}_n^H(k)$, keeping its asymptotic variance. For details on algorithms for the (β, ρ)-estimation, see Gomes and Pestana (2007a,b) and Gomes et al. (2008). We have so far suggested the use of the ρ-estimators in Fraga Alves et al. (2003) and the β-estimators in Gomes and Martins (2002).

The estimators in (13)–(15) are valid for $\gamma \in \mathbb{R}$, but are considered only for $\gamma \geq 0$.

2.3. Maximum likelihood EVI-estimators

As mentioned in de Haan and Ferreira (2006), the class of d.f.'s $F \in \mathcal{D}_{\mathcal{M}}(\mathrm{EV}_\gamma)$, for $\gamma > 0$ (or, more generally, for $\gamma \in \mathbb{R}$), cannot be parameterised with a finite number of parameters, and consequently, there does not exist a maximum-likelihood (ML) estimator for γ in such a wide class of models. There exists however an estimator, introduced by Smith (1987), usually denoted as the ML estimator. Such an estimator was based on the excesses over a deterministic high level u, but can easily be rephrased on the basis of the excesses over the high random threshold $X_{n-k:n}$,

$$W_{ik} := X_{n-i+1:n} - X_{n-k:n}, \quad 1 \leq i \leq k < n. \tag{17}$$

For models in (9), αW_{ik} is well approximated by $Y_{k-i+1:k}^{\gamma} - 1$, with Y a unit Pareto r.v., i.e. the k excesses, in (17), are approximately distributed as the whole set of k top o.s.'s associated with a sample of size k from a *generalised Pareto* d.f., $GP(x; \gamma, \alpha) = 1 - (1 + \alpha x)^{-1/\gamma}, x > 0$ $(\alpha, \gamma > 0)$, a re-parameterisation due to Davison (1984). The solution of the associated ML equations gives rise to an explicit expression for the ML estimator of γ, a function of the ML implicit estimator $\hat{\alpha}_{ML}$ of α and the sample of excesses,

$$\widehat{\gamma}_n^{ML}(k) \equiv \hat{\gamma}_{n,\hat{\alpha}_{ML}}^{ML}(k) \equiv ML(k) := \frac{1}{k}\sum_{i=1}^{k}\ln(1 + \hat{\alpha}_{ML} W_{ik}). \tag{18}$$

A comprehensive study of the asymptotic properties of the ML estimator in (18) has been undertaken in Drees et al. (2004).

Remark 1. A simple heuristic estimator of α is $1/X_{n-k:n}$. If we consider $\hat{\alpha} = 1/X_{n-k:n}$ and the excesses W_{ik}, $1 \leq i \leq k$, in (17), $1 + \hat{\alpha} W_{ik} = X_{n-i+1:n}/X_{n-k:n}$. Then, $\widehat{\gamma}_{n,\hat{\alpha}}^{ML}(k) = \frac{1}{k}\sum_{i=1}^{k}\{\ln X_{n-i+1:n} - \ln X_{n-k:n}\}$ is the classical Hill estimator in (3).

2.4. Asymptotic behaviour of the EVI-estimators

Under the validity of the second-order condition in (10), trivial adaptations of the results in de Haan and Peng (1998), Beirlant et al. (2005), Caeiro et al. (2005), de Haan and Ferreira (2006) and Fraga Alves et al. (2009) enable us to restate the following theorem, already stated in Gomes and Henriques-Rodrigues (2010). Let the notation $\mathcal{N}\left(\mu, \sigma^2\right)$ stand for a normal r.v. with mean value μ and variance σ^2.

Theorem 1 (*Gomes and Henriques-Rodrigues, 2010, Theorem 2.1*). *Assume that condition (10) holds. Let $k = k_n$ be such that (4) holds, and let us additionally assume that we are working with values of k such that $\lambda := \lim_{n\to\infty} \sqrt{k}\,A(n/k)$ is finite. We can then guarantee that*

$$\sqrt{k}\left(\widehat{\gamma}_n^{\bullet}(k) - \gamma\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \mathcal{N}\left(\lambda b_\bullet, \sigma_\bullet^2\right),$$

where

$$b_H = \frac{1}{1-\rho}, \qquad b_M = b_{GH} = \frac{\gamma - \gamma\rho + \rho}{\gamma(1-\rho)^2}, \qquad b_{MM} = b_{ML} = \frac{(1+\gamma)(\gamma+\rho)}{\gamma(1-\rho)(1+\gamma-\rho)},$$
$$\sigma_H^2 = \gamma^2, \qquad \sigma_M^2 = \sigma_{GH}^2 = 1+\gamma^2 \quad \text{and} \quad \sigma_{MM}^2 = \sigma_{ML}^2 = (1+\gamma)^2.$$

If we further assume to be working in the Hall–Welsh class of models in (11), and estimate β and ρ consistently through $\hat\beta$ and $\hat\rho$, with $\hat\rho - \rho = o_p(1/\ln n)$, we get $b_{CH} = 0$ and $\sigma_{CH}^2 = \sigma_H^2 = \gamma^2$.

3. Asymptotic behaviour of MOP EVI-estimators

To have consistency of the Hill estimator, in (3), in all $\mathcal{D}_{\mathcal{M}}^+$, we need to work with *intermediate* values of k, i.e. a sequence of integers $k = k_n$, $1 \le k < n$, such that (4) holds. Under the second-order framework, in (10), the asymptotic distributional representation

$$H(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}} Z_k + \frac{1}{1-\rho} A(n/k)(1+o_p(1)) \tag{19}$$

holds (de Haan and Peng, 1998), where, with $\{E_i\}$ a sequence of i.i.d. standard exponential r.v.'s,

$$Z_k = \sqrt{k}\left(\sum_{i=1}^{k} E_i/k - 1\right) \tag{20}$$

is an asymptotically standard normal r.v.

We now state the main theorem in this paper.

Theorem 2. *Under the validity of the first-order condition, in (9), and for intermediate sequences $k = k_n$, i.e. if (4) holds, the class of estimators $H_p(k)$, in (8), is consistent for the estimation of γ, provided that $p < 1/\gamma$.*

If we moreover assume the validity of the second-order condition in (10), the asymptotic distributional representation

$$H_p(k) \overset{d}{=} \gamma + \frac{\sigma_p(\gamma)\, Z_k^{(p)}}{\sqrt{k}} + b_p(\gamma|\rho)\, A(n/k) + o_p(A(n/k)) \tag{21}$$

holds for all $p < 1/(2\gamma)$ and $\rho \le 0$, with $Z_k^{(p)}$ asymptotically standard normal,

$$\sigma_p(\gamma) := \frac{\gamma(1-p\gamma)}{\sqrt{1-2p\gamma}} \quad \text{and} \quad b_p(\gamma|\rho) := \frac{1-p\gamma}{1-p\gamma-\rho}. \tag{22}$$

Proof. As we have seen before, on the basis of (7) and the law of large numbers, the statistics in (8) are consistent for the estimation of γ for all $p < 1/\gamma$. With Y denoting again a unit Pareto r.v., and working under the second-order framework in (10), we can write

$$\begin{aligned} p \ln A_p(k) &= \ln\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})}\right)^p\right) \\ &= \ln\left(\frac{1}{k}\sum_{i=1}^{k}\left(Y_i^{\gamma}\left(1 + A(n/k)\,(Y_i^{\rho}-1)/\rho + o_p(A(n/k))\right)\right)^p\right) \\ &= \ln\left(\frac{1}{k}\sum_{i=1}^{k} Y_i^{p\gamma}\left(1 + pA(n/k)\,(Y_i^{\rho}-1)/\rho + o_p(A(n/k))\right)\right). \end{aligned}$$

Consequently,

$$p \ln A_p(k) = \ln\left(\frac{1}{k}\sum_{i=1}^{k} Y_i^{p\gamma} + pA(n/k)\,\frac{1}{k}\sum_{i=1}^{k}\frac{Y_i^{p\gamma}(Y_i^{\rho}-1)}{\rho} + o_p(A(n/k))\right).$$

On the basis of (7), and for $a < 1/2$, $\mathbb{V}\mathrm{ar}(Y^a) = a^2/\big((1-a)^2(1-2a)\big)$. We thus know that for $p < 1/(2\gamma)$,

$$\frac{\sqrt{k}(1-p\gamma)\sqrt{1-2p\gamma}\left(\frac{1}{k}\sum_{i=1}^{k} Y_i^{p\gamma} - \frac{1}{1-p\gamma}\right)}{p\gamma} =: Z_k^{(p)} \tag{23}$$

is asymptotically standard normal, and we can write

$$\frac{1}{k}\sum_{i=1}^{k} Y_i^{p\gamma} = \frac{1}{1-p\gamma} + \frac{p\gamma Z_k^{(p)}}{\sqrt{k}(1-p\gamma)\sqrt{1-2p\gamma}}.$$

Also, and now for $p < 1/\gamma$,

$$\mathbb{E}\left(Y^{p\gamma}(Y^{\rho}-1)/\rho\right) = \frac{1}{(1-p\gamma)(1-p\gamma-\rho)}.$$

Let us go back to the EVI-estimators in (8):

$$\begin{aligned} pH_p(k) &= 1-\exp\left(-p\ln A_p(k)\right)\\ &= 1-1\Big/\left(\frac{1}{k}\sum_{i=1}^{k} Y_i^{p\gamma} + pA(n/k)\,\frac{1}{k}\sum_{i=1}^{k} Y_i^{p\gamma}(Y_i^{\rho}-1)/\rho + o_p(A(n/k))\right)\\ &= 1-(1-p\gamma)\Big/\left(1+\frac{p\gamma Z_k^{(p)}}{\sqrt{k}\sqrt{1-2p\gamma}} + \frac{pA(n/k)}{1-p\gamma-\rho} + o_p(A(n/k))\right). \end{aligned}$$

We can thus further write

$$\begin{aligned} pH_p(k) &= 1-(1-p\gamma)\left(1-\frac{p\gamma Z_k^{(p)}}{\sqrt{k}\sqrt{1-2p\gamma}} - \frac{pA(n/k)}{1-p\gamma-\rho} + o_p(A(n/k))\right)\\ &= p\gamma + \frac{p\gamma(1-p\gamma)Z_k^{(p)}}{\sqrt{k}\sqrt{1-2p\gamma}} + \frac{p(1-p\gamma)A(n/k)}{1-p\gamma-\rho} + o_p(A(n/k)), \end{aligned}$$

i.e. (21) follows, with $\sigma_p(\gamma)$ and $b_p(\gamma|\rho)$ given in (22). □

Remark 2. For $p = 0$, $Z_k^{(0)} \equiv Z_k$, with Z_k and $Z_k^{(p)}$ given in (20) and (23), respectively, and on the basis of (21), we get for $H_0(k) \equiv H(k)$, in (3), the particular result in (19), as derived in de Haan and Peng (1998).

Remark 3. Note that, for any $\gamma > 0$, the asymptotic standard deviation $\sigma_p(\gamma)$, in (22), is increasing in $p \geq 0$. In Fig. 1, we present such a standard deviation, as a function of p.

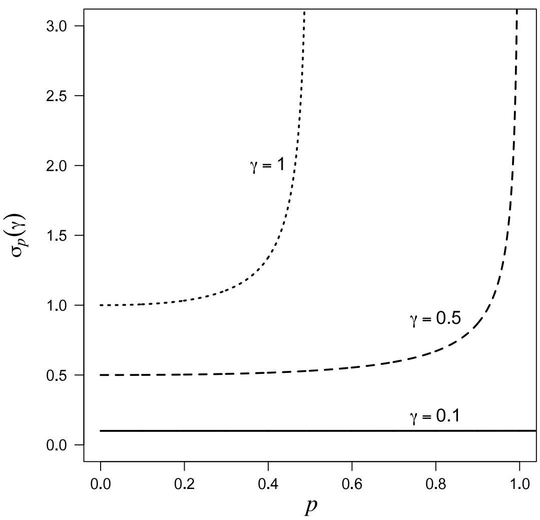

Fig. 1. The asymptotic standard deviation $\sigma_p(\gamma)$ for $\gamma = 0.1, 0.5$ and 1, as a function of $p \geq 0$.

Remark 4. On the other side, also for any $\gamma > 0$, $\rho < 0$ and $p \neq (1-\rho)/\gamma$, the asymptotic bias $b_p(\gamma|\rho)$, also in (22), is decreasing in p. Such a performance is shown in Fig. 2.

These aforementioned results claim for an asymptotic comparison, at optimal levels of the class of EVI-estimators in (8), a topic to be dealt with next, in Section 4.

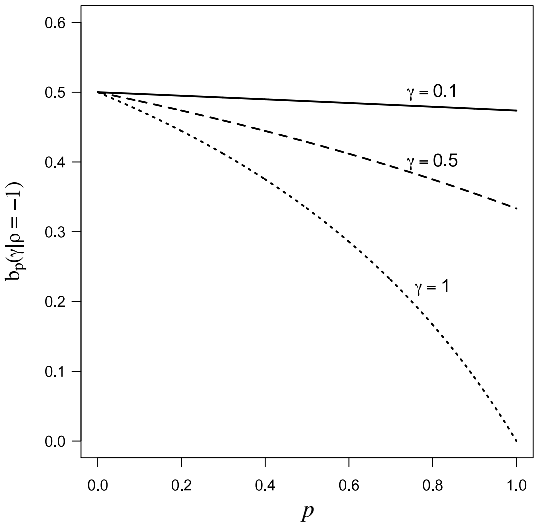

Fig. 2. The asymptotic bias ruler, $b_p(\gamma|\rho)$, for $\rho = -1$ and $\gamma = 0.1, 0.5, 1$, as a function of $p \geq 0$.

4. Asymptotic comparison at optimal levels

We next proceed to the comparison of 'classical' EVI-estimators at their optimal levels. This is again done in a way similar to the one used in de Haan and Peng (1998), Gomes and Martins (2001), Gomes et al. (2005, 2007, 2011a), Gomes and Neves (2008) and Gomes and Henriques-Rodrigues (2010). Let us assume that $\widehat{\gamma}_n^\bullet(k)$ denotes any arbitrary semi-parametric EVI-estimator, for which we have the asymptotic distributional representation

$$\widehat{\gamma}_n^\bullet(k) = \gamma + \frac{\sigma_\bullet Z_k^\bullet}{\sqrt{k}} + b_\bullet A(n/k) + o_p(A(n/k)), \tag{24}$$

for any intermediate sequence of integers $k = k_n$, and where $Z_k^\bullet$ is asymptotically standard normal. Then, $\sqrt{k}\big(\widehat{\gamma}_n^\bullet(k) - \gamma\big) \xrightarrow{d} \mathcal{N}(\lambda_A b_\bullet, \sigma_\bullet^2)$ provided that k is such that $\sqrt{k}\,A(n/k) \to \lambda_A$, finite, as $n \to \infty$. We then write $\mathrm{Bias}_\infty\big(\widehat{\gamma}_n^\bullet(k)\big) := b_\bullet\, A(n/k)$ and $\mathrm{Var}_\infty\big(\widehat{\gamma}_n^\bullet(k)\big) := \sigma_\bullet^2/k$. The so-called *asymptotic mean square error* (AMSE) is then given by

$$\mathrm{AMSE}\,\big(\widehat{\gamma}_n^\bullet(k)\big) := \sigma_\bullet^2/k + b_\bullet^2\, A^2(n/k).$$

Regular variation theory (Bingham et al., 1987) enabled Dekkers and de Haan (1993) to show that, whenever $b_\bullet \neq 0$, there exists a function $\varphi(n) = \varphi(n, \gamma, \rho)$, such that

$$\lim_{n\to\infty} \varphi(n)\ \mathrm{AMSE}\,\big(\widehat{\gamma}_{n0}^\bullet\big) = \big(\sigma_\bullet^2\big)^{-\frac{2\rho}{1-2\rho}}\, \big(b_\bullet^2\big)^{\frac{1}{1-2\rho}} =: \mathrm{LMSE}\,\big(\widehat{\gamma}_{n0}^\bullet\big),$$

where $\widehat{\gamma}_{n0}^\bullet := \widehat{\gamma}_n^\bullet(k_{0|\bullet}(n))$ and $k_{0|\bullet}(n) := \arg\min_k \mathrm{MSE}\,\big(\widehat{\gamma}_n^\bullet(k)\big)$. Moreover, if we slightly restrict the second-order condition in (10), assuming that $A(t) = \gamma\beta t^\rho,\ \rho < 0$, just as happens for the class in (11), we can write

$$k_{0|\bullet}(n) := \arg\min_k \mathrm{MSE}\,\big(\widehat{\gamma}_n^\bullet(k)\big) = \left(\frac{\sigma_\bullet^2\, n^{-2\rho}}{b_\bullet^2\gamma^2\beta^2(-2\rho)}\right)^{1/(1-2\rho)} (1 + o(1)). \tag{25}$$

We again consider the following definition.

Definition 1. Given two biased estimators $\widehat{\gamma}_n^{(1)}(k)$ and $\widehat{\gamma}_n^{(2)}(k)$, for which a distributional representation of the type of the one in (24) holds, with constants (σ_1, b_1) and (σ_2, b_2), $b_1, b_2 \neq 0$, respectively, both computed at their optimal levels, the *asymptotic root efficiency* (AREFF) of $\widehat{\gamma}_{n0}^{(1)}$ relatively to $\widehat{\gamma}_{n0}^{(2)}$ is

$$\mathrm{AREFF}_{1|2} \equiv \mathrm{AREFF}_{\widehat{\gamma}_{n0}^{(1)}|\widehat{\gamma}_{n0}^{(2)}} := \sqrt{\frac{\mathrm{LMSE}\big(\widehat{\gamma}_{n0}^{(2)}\big)}{\mathrm{LMSE}\big(\widehat{\gamma}_{n0}^{(1)}\big)}} = \left(\left(\frac{\sigma_2}{\sigma_1}\right)^{-2\rho}\left|\frac{b_2}{b_1}\right|\right)^{\frac{1}{1-2\rho}}. \tag{26}$$

Remark 5. Note that the AREFF indicator, in (26), has been conceived so that the highest the AREFF indicator is, the better is the first estimator.

4.1. Asymptotic comparison of MOP EVI-estimators at optimal levels

Let us now turn back to the MOP EVI-estimators $H_p(k)$ in (8). We have

$$\mathrm{LMSE}(H_{p0}) = \left(\frac{\gamma^2(1-p\gamma)^2}{1-2p\gamma}\right)^{-\frac{2\rho}{1-2\rho}} \left(\frac{1-p\gamma}{1-p\gamma-\rho}\right)^{\frac{2}{1-2\rho}}.$$

For every (γ, ρ) there is thus always a positive p-value, p_0, such that

$$\mathrm{LMSE}(H_{p0}) < \mathrm{LMSE}(H_{00}) = \mathrm{LMSE}(H_0), \quad \text{for any } p \in (0, p_0).$$

To measure the performance of H_{p0}, we have computed the AREFF-indicator, in (26), now denoted as follows:

$$\mathrm{AREFF}_{p|0} = \left(\left(\frac{\sqrt{1-2p\gamma}}{1-p\gamma}\right)^{-2\rho} \left|\frac{1-p\gamma-\rho}{(1-\rho)(1-p\gamma)}\right|\right)^{\frac{1}{1-2\rho}}. \tag{27}$$

We can reparameterise $\mathrm{AREFF}_{p|0}$, so that we have a dependence on two parameters only, the second-order parameter ρ and the parameter $a = p\gamma < 1/2$. In Fig. 3, we picture the values of

$$\mathrm{AREFF}^*_{a|0} = \left(\left(\frac{\sqrt{1-2a}}{1-a}\right)^{-2\rho} \left|\frac{1-a-\rho}{(1-\rho)(1-a)}\right|\right)^{\frac{1}{1-2\rho}}. \tag{28}$$

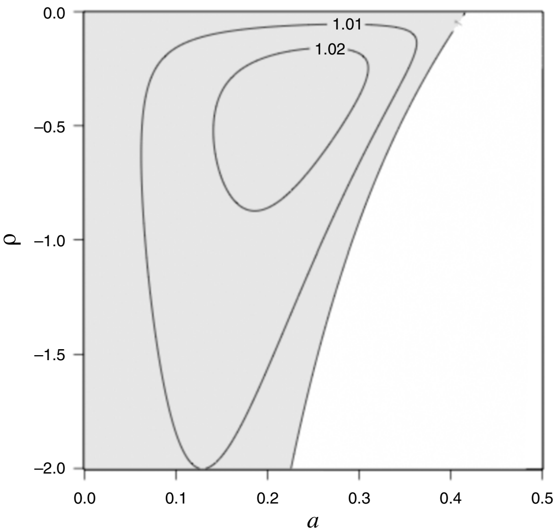

Fig. 3. The indicator $\mathrm{ARREF}^*_{a|0}$, in (28), as a function of (a, ρ).

The gain in efficiency is not terribly high, but, at optimal levels, there is a wide region of the (a, ρ)-plane where the new class of MOP EVI-estimators performs better than the Hill estimator. Let $p_M := \arg\sup_p \mathrm{AREFF}_{p|0}$. Note again that $\mathrm{AREFF}_{p|0}$, in (27), depends on (p, γ) through $p\gamma$. There thus exists a function $\varphi(\rho)$ such that $p_M = \varphi(\rho)/\gamma$. Moreover, as derived in Brilhante et al. (submitted for publication), we have

$$\varphi(\rho) = 1 - \rho/2 - \sqrt{\rho^2 - 4\rho + 2}/2. \tag{29}$$

Thus $\mathrm{AREFF}_{p_M|0}$ depends only on ρ and $\mathrm{AREFF}_{p_M|0} > 1$ if $\rho < 0$, being equal to 1 only if $\rho = 0$.

4.2. An overall comparison of EVI-estimators at optimal levels

As detected in Section 4.1, at optimal levels, the MOP EVI-estimator, H_{p_M0}, can beat the Hill EVI-estimator, H_{00}, in the whole (γ, ρ)-plane. But it can be beaten by the *Mo* EVI-estimator, unless γ is small. The *MM*-estimator in (15) (asymptotically equivalent to the *ML* estimator, unless $\gamma + \rho = 0$ and $(\gamma, \rho) \neq (0, 0)$), can also outperform the *Mo* estimator at optimal levels, in a region around $\gamma + \rho = 0$. In Fig. 4 we exhibit the comparative behaviour of all 'classical' EVI-estimators under consideration.

ρ \ γ	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90	2.00
0.00	MM	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MOP
-0.10	MOP	ML	MM	MM	MM	MM	MM	MM	MM	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo
-0.20	MOP	MOP	ML	MM	MM	MM	MM	MM	MM	MM	MM	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo
-0.30	MOP	MOP	MOP	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	Mo	Mo	Mo	Mo	Mo	Mo	Mo	Mo
-0.40	MOP	MOP	MOP	Mo	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	Mo	Mo	Mo	Mo	Mo	Mo
-0.50	MOP	MOP	MOP	Mo	Mo	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	Mo	Mo	Mo	Mo
-0.60	MOP	MOP	MOP	MOP	Mo	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	Mo	Mo
-0.70	MOP	MOP	MOP	MOP	Mo	Mo	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
-0.80	MOP	MOP	MOP	MOP	Mo	Mo	Mo	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
-0.90	MOP	MOP	MOP	MOP	MOP	Mo	Mo	Mo	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
-1.00	MOP	MOP	MOP	MOP	MOP	Mo	Mo	Mo	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM
-1.10	MOP	MOP	MOP	MOP	MOP	Mo	Mo	Mo	MOP	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM	MM
-1.20	MOP	MOP	MOP	MOP	MOP	Mo	Mo	Mo	MOP	MOP	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM	MM
-1.30	MOP	MOP	MOP	MOP	MOP	MOP	Mo	Mo	MOP	MOP	MOP	MM	MM	ML	MM	MM	MM	MM	MM	MM	MM
-1.40	MOP	MOP	MOP	MOP	MOP	MOP	Mo	Mo	MOP	MOP	MOP	MOP	MM	MM	ML	MM	MM	MM	MM	MM	MM
-1.50	MOP	MOP	MOP	MOP	MOP	MOP	Mo	Mo	MOP	MOP	MOP	MOP	MM	MM	MM	ML	MM	MM	MM	MM	MM
-1.60	MOP	MOP	MOP	MOP	MOP	MOP	Mo	Mo	MOP	MOP	MOP	MOP	MOP	MM	MM	MM	ML	MM	MM	MM	MM
-1.70	MOP	MOP	MOP	MOP	MOP	MOP	Mo	Mo	MOP	MOP	MOP	MOP	MOP	MOP	MM	MM	MM	ML	MM	MM	MM
-1.80	MOP	MOP	MOP	MOP	MOP	MOP	Mo	Mo	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MM	MM	MM	ML	MM	MM
-1.90	MOP	MOP	MOP	MOP	MOP	MOP	MOP	Mo	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MM	MM	MM	ML	MM
-2.00	MOP	MOP	MOP	MOP	MOP	MOP	MOP	Mo	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MOP	MM	MM	MM	MM	ML

Fig. 4. Comparative overall behaviour of the classical EVI-estimators under consideration.

As expected, none of the estimators can always dominate the alternatives, but the MOP EVI-estimators have a nice performance in a region of the (γ, ρ)-plane quite usual in the real world.

Note that in the region $\gamma + \rho \neq 0$ and $\gamma \neq -\rho/(1-\rho)$, the *CH*-estimators, in (16), overpass at optimal levels all other classical estimators under consideration. They are thus not included in Fig. 4 so that we can see the comparative behaviour of the non reduced-bias EVI-estimators. The *MM*, the *CH* and the *ML* estimators, in (15), (16) and (18), respectively, are all second-order reduced-bias estimators in the region $\gamma + \rho = 0$ (where $b_{ML} = b_{MM} = b_{CH} = 0$), and consequently, are expected to outperform at optimal levels any of the other EVI-estimators. The *MM* and the *ML* estimators have an asymptotic variance equal to $(1+\gamma)^2 > \gamma^2$, the asymptotic variance of *CH*. However, this does not mean too much. All depends on the dominant component of bias … and it is without doubt a challenge for further research, out of the scope of this paper. A similar comment applies to the behaviour of the *M*, the *GH* and the *CH*-estimators in the region $\gamma = -\rho/(1-\rho)$ (where $b_M = b_{GH} = b_{CH} = 0$). Again, despite of the fact that the *M* and the *GH* estimators have an asymptotic variance equal to $1 + \gamma^2 > \gamma^2$, the asymptotic variance of *CH*, all depends on the comparative behaviour of the mean square errors.

5. An adaptive choice of *p* and *k*

A reasonably sophisticated algorithm, that has proved to work properly in many situations, is the double-bootstrap algorithm. The basic framework for such algorithm is next provided. For the new class of MOP EVI-estimators $H_p(k)$, in (8),

$$k_{0|p}(n) = \arg\min_k \mathrm{MSE}(H_p(k)) = k_{A|p}(n)(1 + o(1)), \tag{30}$$

with

$$k_{A|p}(n) := \arg\min_k \mathrm{AMSE}\big(H_p(k)\big). \tag{31}$$

For any admissible p, and provided that we can assure the asymptotic normality of the estimator under play, i.e. if $p < 1/(2\gamma)$, the bootstrap methodology can thus enable us to consistently estimate the optimal sample fraction (OSF), $k_{0|p}(n)/n$, with $k_{0|p}(n)$ given in (30), on the basis of a consistent estimator of $k_{A|p}(n)$, in (31), in a way similar to the one used in Gomes and Oliveira (2001), for the classical adaptive Hill EVI estimation, performed through $H(k) \equiv H_0(k)$, in (3), and in Gomes et al. (2011b, in press-a), for second-order reduced-bias estimation. With the notation $\lfloor x \rfloor$ for the integer part of x, we use again the auxiliary statistics

$$T_{k,n} \equiv T(k|H_p) \equiv T_{k,n|p} := H_p(\lfloor k/2 \rfloor) - H_p(k), \quad k = 2, \ldots, n-1, \tag{32}$$

which converge in probability to zero, for any intermediate k, and have an asymptotic behaviour strongly related with the asymptotic behaviour of $H_p(k)$. Indeed, under the above-mentioned second-order framework in (10), we get, for all $p \geq 0$,

$$T(k|H_p) \stackrel{d}{=} \frac{\sigma_p(\gamma)\, P_k^{(p)}}{\sqrt{k}} + b_p(\gamma|\rho)(2^\rho - 1)\, A(n/k)(1 + o_p(1)),$$

with $P_k^{(p)}$ asymptotically standard normal, and $\big(\sigma_p(\gamma), b_p(\gamma|\rho)\big)$ given in (22).

Consequently, denoting $k_{0|T}(n) := \arg\min_k \mathrm{MSE}(T_{k,n})$, we have

$$k_{0|p}(n) = k_{0|T}(n) \times (1 - 2^\rho)^{\frac{2}{1-2\rho}}\, (1 + o(1)). \tag{33}$$

Given the random sample $\underline{X}_n = (X_1, \ldots, X_n)$ from any unknown model F, and the functional in (32), $T_{k,n} =: \phi_k(\underline{X}_n)$, $1 < k < n$, consider for any $n_1 = O(n^{1-\epsilon})$, $0 < \epsilon < 1$, the bootstrap sample $\underline{X}^*_{n_1} = (X_1^*, \ldots, X^*_{n_1})$, from the model

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} I_{[X_i \leq x]},$$

the empirical d.f. associated with the available sample, $\underline{X}_n$. Next, associate to the bootstrap sample the corresponding bootstrap auxiliary statistic, $T^*_{k_1,n_1} := \phi_{k_1}(\underline{X}^*_{n_1})$, $1 < k_1 < n_1$. With $k^*_{0|T}(n_1) = \arg\min_{k_1} \mathrm{MSE}\big(T^*_{k_1,n_1}\big)$,

$$\frac{k^*_{0|T}(n_1)}{k_{0|T}(n)} = \left(\frac{n_1}{n}\right)^{-\frac{2\rho}{1-2\rho}} (1 + o(1)).$$

Consequently, for another sample size, $n_2 = \lfloor n_1^2/n \rfloor + 1$, we have

$$\left(k^*_{0|T}(n_1)\right)^2 / k^*_{0|T}(n_2) = k_{0|T}(n)(1 + o(1)), \quad \text{as } n \to \infty. \tag{34}$$

On the basis of (34), we are now able to first consistently estimate $k_{0|T}$, and next $k_{0|p}$ through (33), on the basis of any estimate $\hat\rho$ of the second-order parameter ρ. With $\hat k^*_{0|T}$ denoting the sample counterpart of $k^*_{0|T}$, $\hat\rho$ an adequate ρ-estimate, and $c_\rho = (1 - 2^\rho)^{\frac{2}{1-2\rho}}$, we thus have the k_0-estimate

$$\hat k^*_{0|p} \equiv \hat k^*_{0|p}(n; n_1) := \min\Big(n - 1, \big\lfloor c_{\hat\rho}\, (\hat k^*_{0|T}(n_1))^2 / \hat k^*_{0|T}([n_1^2/n] + 1) \big\rfloor + 1\Big). \tag{35}$$

The adaptive estimate of γ is then given by

$$H_p^* \equiv H^*_{p,n,n_1|T} := H_p(\hat k^*_{0|p}(n; n_1)). \tag{36}$$

5.1. A double-bootstrap algorithm for an adaptive MOP EVI-estimation

We now proceed with the description of an algorithm for the adaptive estimation of γ. In Steps 2–4, we reproduce the algorithm provided in Gomes and Pestana (2007b) for the estimation of the second-order parameters β and ρ.

Algorithm 5.1. Step 1. Given an observed sample $(x_1, \ldots, x_n)$, compute $H_0(k) \equiv H(k)$, in (3), for $k = 1, 2, \ldots, n - 1$.

Step 2. Compute, for the tuning parameters $\tau = 0$ and $\tau = 1$, the observed values of $\hat\rho_\tau(k)$, the most simple class of estimators in Fraga Alves et al. (2003). Such estimators have the functional form

$$\hat\rho_\tau(k) := \min\Big(0, 3(W^{(\tau)}_{k,n} - 1)/(W^{(\tau)}_{k,n} - 3)\Big),$$

dependent on the statistics

$$W^{(0)}_{k,n} := \frac{\ln\left(M^{(1)}_{k,n}\right) - \frac{1}{2}\ln\left(M^{(2)}_{k,n}/2\right)}{\frac{1}{2}\ln\left(M^{(2)}_{k,n}/2\right) - \frac{1}{3}\ln\left(M^{(3)}_{k,n}/6\right)},$$

$$W^{(1)}_{k,n} := \frac{M^{(1)}_{k,n} - \left(M^{(2)}_{k,n}/2\right)^{1/2}}{\left(M^{(2)}_{k,n}/2\right)^{1/2} - \left(M^{(3)}_{k,n}/6\right)^{1/3}},$$

where $M^{(j)}_{k,n}$, $j = 1, 2, 3$, are given in (12).

Step 3. Consider $\big\{\hat\rho_\tau(k)\big\}_{k \in \mathcal{K}}$, with $\mathcal{K} = \big(\lfloor n^{0.995} \rfloor, \lfloor n^{0.999} \rfloor\big)$, compute their median, denoted χ_τ, and compute $I_\tau := \sum_{k \in \mathcal{K}} \big(\hat\rho_\tau(k) - \chi_\tau\big)^2$, $\tau = 0, 1$. Next choose the tuning parameter $\tau^* = 0$ if $I_0 \leq I_1$; otherwise, choose $\tau^* = 1$.

Step 4. Work with $\hat{\rho} \equiv \hat{\rho}_{\tau^*} = \hat{\rho}_{\tau^*}(k_1)$ and $\hat{\beta} \equiv \hat{\beta}_{\tau^*} := \hat{\beta}_{\hat{\rho}_{\tau^*}}(k_1)$, $k_1 = \lfloor n^{0.999} \rfloor$, being $\hat{\beta}_{\hat{\rho}}(k)$ the estimator in Gomes and Martins (2002), given by

$$\hat{\beta}_{\hat{\rho}}(k) := \left(\frac{k}{n}\right)^{\hat{\rho}} \frac{d_k(\hat{\rho})\, D_k(0) - D_k(\hat{\rho})}{d_k(\hat{\rho})\, D_k(\hat{\rho}) - D_k(2\hat{\rho})},$$

dependent on the estimator $\hat{\rho} = \hat{\rho}_{\tau^*}(k_1)$, and where, for any $\alpha \le 0$,

$$d_k(\alpha) := \frac{1}{k}\sum_{i=1}^{k} (i/k)^{-\alpha} \quad \text{and} \quad D_k(\alpha) := \frac{1}{k}\sum_{i=1}^{k} (i/k)^{-\alpha}\, U_i,$$

with $U_i = i\left(\ln X_{n-i+1:n} - \ln X_{n-i:n}\right)$, $1 \le i \le k < n$, the scaled log-spacings.

Step 5. Next, consider sub-sample sizes $n_1 = \lfloor n^b \rfloor$, $b = 0.925(0.001)0.999$, $n_2 = \lfloor n_1^2/n \rfloor + 1$.

Step 6. For l from 1 until $B = 250$ (number of bootstrap iterations), generate independently, from the empirical d.f. $F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} I_{[X_i \le x]}$ associated with the observed sample, the bootstrap samples $(x_1^*, \ldots, x_{n_2}^*)$ and $(x_1^*, \ldots, x_{n_2}^*, x_{n_2+1}^*, \ldots, x_{n_1}^*)$, of sizes n_2 and n_1, respectively.

Step 7. Denoting $T_{k,n}^*$ the bootstrap counterpart of $T_{k,n}$, in (32), obtain, for $1 \le l \le B$, $t_{k,n_1,l}^*$, $1 < k < n_1$, $t_{k,n_2,l}^*$, $1 < k < n_2$, the observed values of the statistic T_{k,n_i}^*, $i = 1, 2$, and compute, for $i = 1, 2$ and $k = 2, \ldots, n_i - 1$,

$$\mathrm{MSE}^*(n_i, k) = \frac{1}{B}\sum_{l=1}^{B}\left(t_{k,n_i,l}^*\right)^2.$$

Step 8. Obtain $\hat{k}_{0|T}^*(n_i) := \arg\min_{1<k<n_i} \mathrm{MSE}^*(n_i, k)$, $i = 1, 2$, and return to Step 6 if $\hat{k}_{0|T}^*(n_2) > \hat{k}_{0|T}^*(n_1)$.

Step 9. Compute $\hat{k}_{0|0}^* \equiv \hat{k}_{0|H_0}^*(n; n_1)$, with $\hat{k}_{0|p}^*$ given in (35).

Step 10. Compute $H_0^* \equiv H_{0,n,n_1|T}^*$, with H_p^* given in (36), and the MSE-estimate

$$\widehat{\mathrm{MSE}}_0^* \equiv \widehat{\mathrm{MSE}}_0^*(n_1) \equiv \widehat{\mathrm{MSE}}(\hat{k}_{0|0}^*|H_0^*) := \frac{(H_0^*)^2}{\hat{k}_{0|0}^*} + \left(\frac{H_0^*\hat{\beta}\left(n/\hat{k}_{0|0}^*\right)^{\hat{\rho}}}{1-\hat{\rho}}\right)^2 =: (\hat{\sigma}_{00}^*)^2 + (\hat{b}_{00}^*)^2.$$

Step 11. For $p = a/(20H_0^*)$, with H_0^* the estimate obtained in Step 10, and $a = 1, 2, \ldots, 9$, compute $H_p(k)$, $k = 1, 2, \ldots, n-1$, and perform the algorithm from Step 5 until Step 8.

Step 12. Compute $\hat{k}_{0|p}^* \equiv \hat{k}_{0|p}^*(n; n_1)$, given in (35).

Step 13. Compute $H_p^* \equiv H_{p,n,n_1|T}^*$, given in (36), and the MSE-estimate

$$\widehat{\mathrm{MSE}}_p^* \equiv \widehat{\mathrm{MSE}}_p^*(n_1) \equiv \widehat{\mathrm{MSE}}(\hat{k}_{0|p}^*|H_p^*) := \frac{\sigma_p^2(H_p^*)}{\hat{k}_{0|p}^*} + \left(\frac{H_p^*\hat{\beta}(1-pH_p^*)\left(n/\hat{k}_{0|p}^*\right)^{\hat{\rho}}}{1-pH_p^*-\hat{\rho}}\right)^2 =: (\hat{\sigma}_{0p}^*)^2 + (\hat{b}_{0p}^*)^2, \quad (37)$$

where $\sigma_p(\gamma)$ has been defined in (22).

Step 14. Compute the median, χ_p, of $\widehat{\mathrm{MSE}}_p^*(n_1)$ for the values of n_1 in Step 5, and consider $p_{\min}^* := \arg\inf_p \chi_p$.

Step 15. Choose $n_1^* := \arg\min_{n_1} \widehat{\mathrm{MSE}}_{p_{\min}^*}^*(n_1)$, with $\widehat{\mathrm{MSE}}_p^*(n_1)$, obtained in Step 13.

Step 16. Consider the adaptive threshold estimate $\hat{k}_0^{**} := \hat{k}_{0|p_{\min}^*}(n; n_1^*)$ and the final EVI-estimate $H^{**} := H_{p_{\min}^*}^* = H_{p,n,n_1^*|T}^*$.

Remark 6. For any $p \ge 0$, and with $\hat{k}_{0|p}^*$ and $(\sigma_{0p}^*, b_{0p}^*)$ given in (35) and (37), respectively, the r.v. $\left(H_p(\hat{k}_{0|p}^*) - \gamma - b_{0p}^*\right)/\sigma_{0p}^*$ is approximately $\mathcal{N}(0, 1)$. We can then get approximate $100(1-\alpha)\%$ confidence intervals (CIs) for γ, given by

$$\left(H_p(\hat{k}_{0|p}^*) - b_{0p}^* - \xi_{1-\alpha/2}\sigma_{0p}^*,\ H_p(\hat{k}_{0|p}^*) - b_{0p}^* + \xi_{1-\alpha/2}\sigma_{0p}^*\right),$$

where ξ_q denotes the quantile of probability q of a standard normal d.f.

Remark 7. We make the following comments.

(i) If there are negative elements in the sample, the value of n must be replaced by $n_0 := \sum_{i=1}^{n} I_{[X_i>0]}$, the number of positive elements in the sample. The same comment applies to n_1 and n_2.

(ii) As already mentioned in several papers essentially related with bias reduction, in Step 2 of the algorithm we are led in almost all situations to the tuning parameter $\tau = 0$ whenever $-1 \le \rho < 0$ and $\tau = 1$, otherwise. We thus claim again for the relevance of the choice $\tau = 0$, the one considered in the applications in Section 7.

(iii) Regarding second-order parameters' estimation, attention should also be paid to the more recent classes of ρ-estimators proposed in Goegebeur et al. (2008, 2010), Ciuperca and Mercadier (2010) and Caeiro and Gomes (2012), and to the estimators of β in Caeiro and Gomes (2006) and in Gomes et al. (2010).

(iv) In Algorithm 5.1 above, we have also dealt with the choice of the tuning parameter n_1 associated with the bootstrap methodology, but again, the method is only moderately dependent on the choice of the nuisance parameter n_1, in Step 5 of Algorithm 5.1. This enhances the practical value of the method. Moreover, although aware of the need of $n_1 = o(n)$, it seems that, once again, we get good results up till n.

(v) The Monte-Carlo procedure in Steps 6–16 of Algorithm 5.1 can be replicated, if we want to associate standard bootstrap errors to the OSF and to the EVI-estimates. The value of B can also be adequately chosen.

(vi) We would like to stress again that the use of the random sample of size n_2, $(x_1^*, \ldots, x_{n_2}^*)$, and of the extended sample of size n_1, $(x_1^*, \ldots, x_{n_2}^*, x_{n_2+1}^*, \ldots, x_{n_1}^*)$, leads to a higher precision of the result with a smaller B, the number of bootstrap samples generated. Indeed, if we had generated the samples of sizes n_1 and n_2 independently, we would have got a wider confidence interval for the EVI, should we have kept the same value for B. This is quite similar to the use of the simulation technique of "*Common Random Numbers*" in comparison algorithms, when we want to decrease the variance of a final answer to $z = y_1 - y_2$, inducing a positive dependence between y_1 and y_2.

(vii) For a different way to overcome the complex uncertainties associated with threshold choice, see MacDonald et al. (2011).

6. Finite sample properties of the EVI-estimators

We have implemented multi-sample Monte Carlo simulation experiments of size 5000×20 for the class of MOP EVI-estimators, in (8), comparatively with the MVRB EVI-estimators, in (16), for sample sizes $n = 100, 200, 500, 1000, 2000$ and 5000, from the following underlying models:

(1) the *Fréchet* model, with d.f. $F(x) = \exp(-x^{-1/\gamma})$, $x \ge 0$, $\gamma = 0.1, 0.25, 0.5$ and 1;
(2) the *extreme value* model, with d.f. $F(x) = EV_\gamma(x)$, with $EV_\gamma(x)$ given in (1), for the same values of γ;
(3) the *generalised Pareto* model, with d.f. $F(x) = 1 + \ln EV_\gamma(x) = 1 - \left(1 + \gamma x\right)^{-1/\gamma}$, $0 \le x < -1/\gamma$, $EV_\gamma(x)$ given in (1), also for the same γ-values;
(4) the *Student-*t_ν, with $\nu = 1, 2, 4$, i.e. for values of $\gamma = 1, 0.5, 0.25 (\gamma = 1/\nu)$.

For details on multi-sample simulation, see Gomes and Oliveira (2001).

6.1. Mean values and mean square error patterns

For each value of n and for each of the above-mentioned models, we have first simulated the mean values (E) and the root mean square errors (RMSEs) of the estimators $H_p(k)$, in (8), as functions of the number of top order statistics k involved in the estimation and for $p = j/(10\gamma)$, with j assuming values from 0 to 4, with step 1. As a curiosity, we have also considered values of j from 5 until 12, again with step 1. Note that for $j \le 9$ we can guarantee consistency. Values of $j \ge 10$ are totally outside the scope of Theorem 2. Some of those values, based on the first replicate with a size 5000, are pictured in Figs. 5–7, for samples of size $n = 1000$ from EV_γ underlying parents with $\gamma = 0.25$, $\gamma = 0.5$ and $\gamma = 1$, respectively. As mentioned above, we also picture the patterns of E and RMSE for the MVRB EVI-estimators in (16).

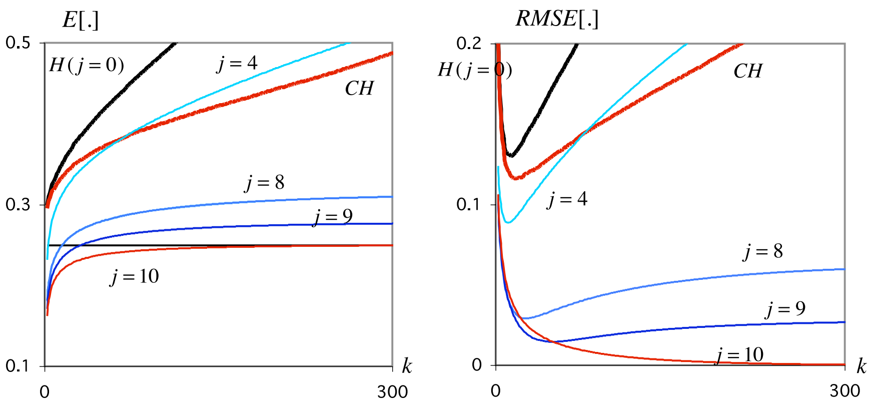

Fig. 5. Mean values (*left*) and RMSE (*right*) of the EVI-estimators under study for an EV_γ d.f. with $\gamma = 0.25$.

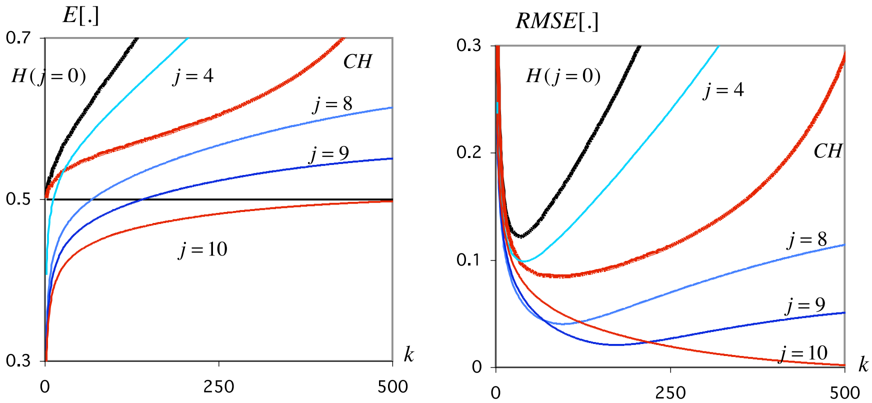

Fig. 6. Mean values (*left*) and RMSE (*right*) of the EVI-estimators under study for an EV_γ d.f. with $\gamma = 0.5$.

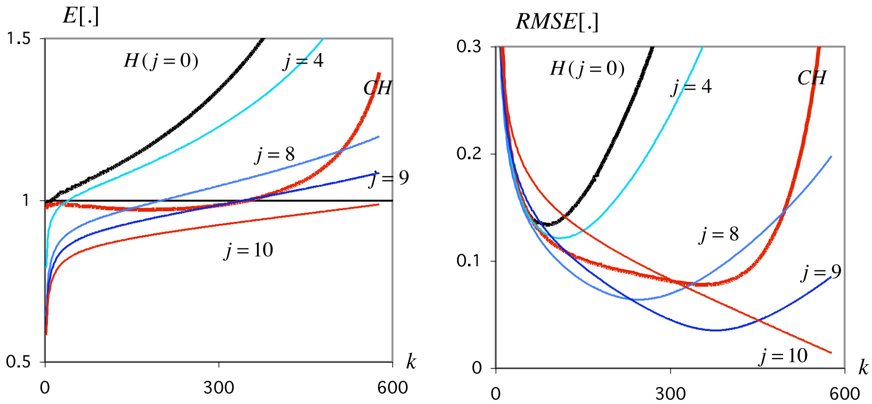

Fig. 7. Mean values (*left*) and RMSE (*right*) of the EVI-estimators under study for an EV_γ d.f. with $\gamma = 1$.

Similar patterns have been obtained for all other simulated models. We can always find an optimal value for p, clear from these pictures in what concerns RMSE, but also valid for mean values at optimal levels, in the sense of minimal RMSE, as we shall see next, in Section 6.1.1. Astonishingly, $p = 1/\gamma$ $(j = 10)$ provides very interesting results, even for large values of γ.

6.1.1. Mean values of the EVI-estimators at optimal levels

Table 1 is again related with the EV_γ model. We there present, for $n = 200, 500, 1000, 2000$ and 5000, the simulated mean values at optimal levels (levels where RMSEs are minima as functions of k) of the EVI-estimators CH, in (16) and $H_p(k)$, in (8), for $p = j/(10\gamma)$, and the three regions, $j = 0, 2, 4$, $j = 5, 7, 9$ and $j = 10$. Information on 95% confidence intervals, computed on the basis of the 20 replicates with 5000 runs each, is also provided. Among the estimators considered, and distinguishing the three regions of j-values, the one providing the smallest squared bias is underlined, and written in **bold** whenever it outperforms the behaviour achieved in the previous region.

Remark 8. We may draw the following specific comments.

- As intuitively expected, H_{p0} are decreasing in p until $p_{\min}$, approaching the true value of γ, not only for the EV_γ model, but for all simulated models.
- The above mentioned remark means that, regarding bias, and if we restrict ourselves to the region of p-values where we can guarantee asymptotic normality, we can safely take $p = 4/(10\gamma)$. However, we have to pay attention to variance, which increases with p.
- The MOP EVI-estimators outperform the MVRB EVI-estimators, unless both γ and n are large.

6.1.2. Mean square errors and relative efficiency indicators at optimal levels

We have computed the Hill estimator, in (8) whenever $p = 0$, at the simulated value of $k_{0|0} := \arg\min_k \mathrm{MSE}\big(H_0(k)\big)$, the simulated optimal k in the sense of minimal RMSE, not relevant in practice, but providing an indication of the best

Table 1
Simulated mean values, at optimal levels, and associated 95% confidence intervals, of $CH(k)$ and $H_p(k)$, with $p = j/(10\gamma)$, $j = 0, 2, 4, 5, 7, 9, 10$, for EV underlying parents.

EV_γ **parent**, $\gamma = 0.25$					
n	200	500	1000	2000	5000
CH	0.372 ± 0.0021	0.353 ± 0.0014	0.342 ± 0.0017	0.330 ± 0.0008	0.317 ± 0.0008
$j = 0$	0.392 ± 0.0026	0.365 ± 0.0019	0.348 ± 0.0012	0.335 ± 0.0013	0.321 ± 0.0010
$j = 2$	0.350 ± 0.0026	0.338 ± 0.0011	0.330 ± 0.0013	0.321 ± 0.0008	0.312 ± 0.0009
$j = 4$	<u>0.309</u> ± 0.0006	<u>0.303</u> ± 0.0013	<u>0.301</u> ± 0.0013	<u>0.300</u> ± 0.0011	<u>0.294</u> ± 0.0008
$j = 5$	0.295 ± 0.0023	0.293 ± 0.0012	0.289 ± 0.0011	0.288 ± 0.0007	0.285 ± 0.0007
$j = 7$	0.278 ± 0.0007	0.274 ± 0.0004	0.272 ± 0.0005	0.270 ± 0.0005	0.268 ± 0.0003
$j = 9$	**<u>0.261</u>** ± 0.0002	**<u>0.259</u>** ± 0.0002	**<u>0.258</u>** ± 0.0002	**<u>0.257</u>** ± 0.0001	**<u>0.256</u>** ± 0.0001
$j = 10$	**<u>0.250</u>** ± 0.0000	**<u>0.250</u>** ± 0.0000	**<u>0.250</u>** ± 0.0000	**<u>0.250</u>** ± 0.0000	**<u>0.250</u>** ± 0.0000
EV_γ **parent**, $\gamma = 0.5$					
CH	0.573 ± 0.0016	0.564 ± 0.0014	0.558 ± 0.0012	0.550 ± 0.0009	0.541 ± 0.0006
$j = 0$	0.624 ± 0.0033	0.596 ± 0.0011	0.579 ± 0.0016	0.565 ± 0.0010	0.551 ± 0.0010
$j = 2$	0.602 ± 0.0019	0.582 ± 0.0016	0.570 ± 0.0013	0.559 ± 0.0012	0.547 ± 0.0009
$j = 4$	<u>0.570</u> ± 0.0016	<u>0.562</u> ± 0.0015	<u>0.558</u> ± 0.0012	<u>0.550</u> ± 0.0008	<u>0.541</u> ± 0.0007
$j = 5$	0.556 ± 0.0024	0.551 ± 0.0012	0.546 ± 0.0011	0.541 ± 0.0007	0.536 ± 0.0008
$j = 7$	0.532 ± 0.0015	0.526 ± 0.0005	0.524 ± 0.0006	0.523 ± 0.0007	0.520 ± 0.0004
$j = 9$	**<u>0.512</u>** ± 0.0004	**<u>0.510</u>** ± 0.0001	**<u>0.509</u>** ± 0.0001	**<u>0.507</u>** ± 0.0002	**<u>0.506</u>** ± 0.0001
$j = 10$	**<u>0.497</u>** ± 0.0003	**<u>0.499</u>** ± 0.0000	**<u>0.500</u>** ± 0.0000	**<u>0.500</u>** ± 0.0000	**<u>0.500</u>** ± 0.0000
EV_γ **parent**, $\gamma = 1$					
CH	<u>0.975</u> ± 0.0046	<u>1.003</u> ± 0.0024	<u>1.004</u> ± 0.0013	<u>1.003</u> ± 0.0007	<u>1.001</u> ± 0.0004
$j = 0$	1.124 ± 0.0032	1.091 ± 0.0030	1.073 ± 0.0020	1.058 ± 0.0014	1.042 ± 0.0009
$j = 2$	1.109 ± 0.0032	1.083 ± 0.0018	1.067 ± 0.0014	1.054 ± 0.0013	1.040 ± 0.0011
$j = 4$	1.085 ± 0.0019	1.070 ± 0.0017	1.060 ± 0.0015	1.050 ± 0.0010	1.039 ± 0.0010
$j = 5$	1.070 ± 0.0024	1.060 ± 0.0015	1.053 ± 0.0013	1.046 ± 0.0009	1.037 ± 0.0006
$j = 7$	1.041 ± 0.0011	1.035 ± 0.0009	1.031 ± 0.0009	1.027 ± 0.0005	1.023 ± 0.0004
$j = 9$	**<u>1.016</u>** ± 0.0003	<u>1.013</u> ± 0.0004	<u>1.012</u> ± 0.0001	<u>1.010</u> ± 0.0003	<u>1.009</u> ± 0.0000
$j = 10$	0.959 ± 0.0019	0.979 ± 0.0010	0.987 ± 0.0005	0.992 ± 0.0004	0.995 ± 0.0002

possible performance of the Hill estimator. Such an estimator is denoted by H_{00}. We have also compute H_{p0}, the estimator H_p computed at the simulated value of $k_{0|p} := \arg\min_k \text{MSE}\big(H_p(k)\big)$. The simulated indicators are

$$\text{REFF}_{p|0} := \frac{\text{RMSE}\,(H_{00})}{\text{RMSE}\,\big(H_{p0}\big)} = \sqrt{\frac{\text{MSE}\,(H_{00})}{\text{MSE}\,\big(H_{p0}\big)}}. \tag{38}$$

A similar indicator has also been computed for the *CH* EVI-estimator, and as mentioned in Remark 5, the higher these indicators are, the better the associated EVI-estimators perform, comparatively to H_{00}.

Again as an illustration of the results obtained, we present Table 2. In the first row, we provide the RMSE of H_{00} so that we can easily recover the RMSE of all other estimators H_{p0}. The following rows provide the REFF indicators of $CH|H$ and $\text{REFF}_{p|0}$ in (38), for the different MOP EVI-estimators under study. A similar mark (<u>underlined</u> and **bold**) is used for the highest REFF indicator. Also, if the highest value in the first two regions is not achieved by some of the MOP EVI-estimators, we place in *italic* the highest REFF among those MOP EVI-estimators. Confidence intervals are not provided for REFF-indicators larger than 20, achieved when $j = 10$.

Remark 9. We now provide a few comments related with the REFF-indicators.

- Just as for mean values at optimal levels, and again if we restrict ourselves to the region of p-values where we can guarantee asymptotic normality, the best results were obtained for $p = 4/(10\gamma)$ for all simulated models but the Fréchet (independently of γ) and models with $\gamma = 1$.
- For Fréchet underlying parents, the REFF-indicator $\text{REFF}_{p|0}$, provided in Table 3 for $p = j/\gamma, j = 1(1)10$, does not depend on γ.
- Regarding RMSE, and at optimal levels, the consistent MOP EVI-estimators can always beat the MVRB EVI estimators, also computed at optimal levels.

Conjecture 1. *We are just led to conjecture that the choice $p = 1/\gamma$ can be an adequate one, but we have no theoretical support for such a choice, because for such a p-value we can guarantee neither asymptotic normality nor consistency. We have however at the moment a computational validation of the result. Note also that the bias leading term is null for $p = 1/\gamma$, as can be seen from (22). We have thus also considered in the case-studies the MOP EVI-estimate, $\widetilde{H}$, associated with $\widetilde{p} = 1/H_{00}$, and computed at $k = \lfloor n^{0.99} \rfloor$.*

Table 2
Simulated RMSEs of H (first row) and REFF-indicators of $CH(k)$ and $H_p(k)$, $p = j/(10\gamma), j = 2, 4, 5, 7, 9, 10$, for EV underlying parents, together with 95% confidence intervals.

EV_γ **parent**, $\gamma = 0.25$					
n	200	500	1000	2000	5000
$RMSE_H$	0.216 ± 0.0185	0.174 ± 0.0141	0.151 ± 0.0136	0.133 ± 0.0127	0.113 ± 0.0108
CH	1.238 ± 0.0056	1.171 ± 0.0042	1.130 ± 0.0021	1.101 ± 0.0021	1.072 ± 0.0020
$j = 2$	1.245 ± 0.0043	1.180 ± 0.0024	1.147 ± 0.0029	1.121 ± 0.0021	1.099 ± 0.0016
$j = 4$	1.744 ± 0.0067	1.565 ± 0.0055	1.463 ± 0.0066	1.379 ± 0.0052	1.300 ± 0.0040
$j = 5$	2.154 ± 0.0078	1.905 ± 0.0070	1.757 ± 0.0077	1.633 ± 0.0064	1.509 ± 0.0052
$j = 7$	3.745 ± 0.0147	3.264 ± 0.0138	2.975 ± 0.0127	2.723 ± 0.0104	2.459 ± 0.0075
$j = 9$	**11.358** ± 0.0444	**9.895** ± 0.0578	**8.989** ± 0.0493	**8.189** ± 0.0308	**7.331** ± 0.0242
$j = 10$	**4698.72**	**45624.5**	**226148**	**1144373**	**7957240**
EV_γ **parent**, $\gamma = 0.5$					
$RMSE_H$	0.200 ± 0.0199	0.157 ± 0.0153	0.133 ± 0.0132	0.113 ± 0.0100	0.092 ± 0.0081
CH	**1.501** ± 0.0097	**1.476** ± 0.0059	**1.452** ± 0.0057	**1.417** ± 0.0057	**1.359** ± 0.0052
$j = 2$	1.143 ± 0.0023	1.104 ± 0.0020	1.085 ± 0.0022	1.071 ± 0.0020	1.059 ± 0.0021
$j = 4$	1.409 ± 0.0050	1.287 ± 0.0049	1.221 ± 0.0054	1.169 ± 0.0052	1.123 ± 0.0048
$j = 5$	1.650 ± 0.0059	1.468 ± 0.0066	1.362 ± 0.0071	1.275 ± 0.0067	1.189 ± 0.0062
$j = 7$	2.663 ± 0.0115	2.290 ± 0.0110	2.060 ± 0.0104	1.857 ± 0.0097	1.643 ± 0.0113
$j = 9$	**7.616** ± 0.0408	**6.479** ± 0.0360	**5.732** ± 0.0267	**5.071** ± 0.0218	**4.330** ± 0.0278
$j = 10$	**50.756**	**132.335**	**258.474**	**516.237**	**1188.7**
EV_γ **parent**, $\gamma = 1$					
$RMSE_H$	0.202 ± 0.0229	0.151 ± 0.0140	0.122 ± 0.0109	0.100 ± 0.0090	0.077 ± 0.0068
CH	1.182 ± 0.0230	**1.410** ± 0.0212	1.679 ± 0.0192	2.005 ± 0.0192	2.500 ± 0.0218
$j = 2$	1.088 ± 0.0032	1.064 ± 0.0018	1.052 ± 0.0025	1.042 ± 0.0019	1.036 ± 0.0015
$j = 4$	1.225 ± 0.0068	1.134 ± 0.0040	1.084 ± 0.0062	1.045 ± 0.0055	1.014 ± 0.0046
$j = 5$	1.362 ± 0.0074	1.214 ± 0.0051	1.128 ± 0.0072	1.055 ± 0.0072	0.987 ± 0.0061
$j = 7$	2.016 ± 0.0100	1.688 ± 0.0083	1.484 ± 0.0088	1.306 ± 0.0087	1.119 ± 0.0081
$j = 9$	**5.250** ± 0.0314	**4.334** ± 0.0178	**3.696** ± 0.0170	**3.127** ± 0.0165	**2.514** ± 0.0181
$j = 10$	5.086 ± 0.2246	**7.110** ± 0.3126	**8.915** ± 0.2890	**11.316** ± 0.4723	**14.851** ± 0.4968

Table 3
Simulated RMSEs of H/γ (first row) and REFF-indicators of $CH(k)$ and $H_p(k)$ (independent of γ), for $p = j/(10\gamma), j = 1(1)10$, for *Fréchet* parents, together with 95% confidence intervals.

Fréchet parent, γ					
n	200	500	1000	2000	5000
$RMSE_H$	0.163 ± 0.1520	0.117 ± 0.1432	0.091 ± 0.1345	0.071 ± 0.1977	0.052 ± 0.1764
CH	1.237 ± 0.1591	1.337 ± 0.0080	1.460 ± 0.0123	1.574 ± 0.0123	1.795 ± 0.0097
$j = 1$	1.031 ± 0.0011	1.026 ± 0.0010	1.023 ± 0.0009	1.020 ± 0.0010	1.019 ± 0.0010
$j = 2$	1.059 ± 0.0028	1.046 ± 0.0020	1.039 ± 0.0020	*1.032* ± 0.0024	*1.030* ± 0.0019
$j = 3$	1.084 ± 0.0057	1.055 ± 0.0032	*1.041* ± 0.0037	1.028 ± 0.0040	1.022 ± 0.0033
$j = 4$	*1.120* ± 0.0081	*1.060* ± 0.0047	1.027 ± 0.0055	0.999 ± 0.0069	0.982 ± 0.0053
$j = 5$	1.195 ± 0.0092	1.086 ± 0.0061	1.021 ± 0.0071	0.964 ± 0.0090	0.918 ± 0.0072
$j = 6$	1.347 ± 0.0100	1.173 ± 0.0071	1.066 ± 0.0078	0.970 ± 0.0091	0.878 ± 0.0074
$j = 7$	1.645 ± 0.0110	1.383 ± 0.0076	1.219 ± 0.0087	1.071 ± 0.0088	0.924 ± 0.0073
$j = 8$	2.277 ± 0.0130	1.864 ± 0.0097	1.602 ± 0.0107	1.370 ± 0.0099	1.137 ± 0.0090
$j = 9$	**4.196** ± 0.0237	**3.363** ± 0.0151	**2.833** ± 0.0169	**2.370** ± 0.0142	**1.901** ± 0.0140
$j = 10$	**5.383** ± 0.0389	**5.320** ± 0.0231	**5.140** ± 0.0260	**4.875** ± 0.0305	**4.466** ± 0.0248

7. Case-studies

We now consider an application of Algorithm 5.1 to:

(1) two randomly simulated samples, with size $n = 500$, from a Fréchet parent with $\gamma = 0.25$, denoted by FRE_1 and FRE_2;
(2) two randomly simulated samples, with size $n = 1000$, from a Student t_ν parent with $\nu = 4$ ($\gamma = 1/\nu = 0.25$), denoted by STU_1 and STU_2;
(3) the data analysed in Drees (2003) and later on in Araújo Santos et al. (2006) and Gomes et al. (in press-b), the daily log-returns of NASDAQ index from 1997 to 2000, which corresponds to a sample size given by $n = 1037$;
(4) a sample, with size $n = 371$, of automobile claim amounts exceeding 1,200,000 Euro over the period 1988–2001, gathered from several European insurance companies co-operating with the same re-insurer (Secura Belgian Re), already

studied in Beirlant et al. (2004, 2008), Vandewalle and Beirlant (2006) and Gomes et al. (2011b), as an example to excess-of-loss reinsurance rating and heavy-tailed distributions in car insurance, and denoted by SECURA;

(5) a sample, of size $n = 2627$, denoted by FIRES, already considered in Gomes et al. (in press-a) and associated with the number of hectares, exceeding 100 ha, burnt during wildfires recorded in Portugal during 14 years (1990–2003).

Remark 10. Apart from the bootstrap adaptive EVI-estimates in Algorithm 5.1 and the bootstrap CIs, in Remark 6, we consider, for the Hill estimator, the most common estimate of $k_{0|0} \equiv k_{0|H}(n) := \arg\min_k \mathrm{MSE}(H_0(k))$ (Hall, 1982), with $k_{0|\bullet}$ given in (25), i.e.,

$$\hat{k}_{0|0} = \min\left(n-1, \left\lfloor \left((1-\hat{\rho})^2 n^{-2\hat{\rho}}/\left(-2\,\hat{\rho}\,\hat{\beta}^2\right)\right)^{1/(1-2\hat{\rho})} \right\rfloor + 1\right). \tag{39}$$

Next, with $b_{k,n,\beta,\rho} = 1 + \beta(n/k)^\rho/(1-\rho)$, we consider the approximate $100(1-\alpha)\%$ CI for γ,

$$\left(\frac{H(k)}{b_{k,n,\hat{\beta},\hat{\rho}} + \xi_{1-\alpha/2}/\sqrt{k}},\ \frac{H(k)}{b_{k,n,\hat{\beta},\hat{\rho}} - \xi_{1-\alpha/2}/\sqrt{k}}\right).$$

For all data sets under analysis, the sample paths of the ρ-estimates associated with $\tau = 0$ and $\tau = 1$ led to the choice of the estimates associated with $\tau = 0$, on the basis of any stability criterion for large k, including the one in Step 2 of the algorithm. In Tables 4–6, we present a summary of the data analysis performed. In Table 4, apart from an indication of the sample size n, the number n_0 of positive elements in the sample, and the estimates $(\hat{\beta}_0, \hat{\rho}_0)$ of the vector of second-order parameters (β, ρ), in (11), obtained in Step 4, we provide the sub-sample size choice n_1^*, in Step 14, the value a^* and associated $p^*_{\min}$, in Step 15, and the bootstrap threshold estimates, $\widehat{k}_{0|0}$, $\hat{k}^*_{0|0}$ and $\hat{k}^{**}_0$, in (39), Steps 9 and 16, respectively. We further present $\widetilde{p} = 1/H_{00} := 1/H(\widehat{k}_{0|0})$.

Table 4
Values of n, n_0, estimates $(\hat{\beta}_0, \hat{\rho}_0)$, n_1^*, a^* and $p^*_{\min}$, adaptive estimates of the threshold k $(\hat{k}_{0|0}, \hat{k}^*_{0|0}, \hat{k}^{**}_0)$, and $\widetilde{p} = 1/H_{00}$, for the data sets under analysis.

Data	n	n_0	$(\hat{\beta}_0, \hat{\rho}_0)$	n_1^*	a^*	$p^*_{\min}$	$\hat{k}_{0\|0}$	$\hat{k}^*_{0\|0}$	$\hat{k}^{**}_0$	$\widetilde{p}$
FRE_1	500	500	(0.89, −1.02)	409	5	1.493	88	101	101	3.670
FRE_2	500	500	(0.87, −1.50)	366	6	0.544	137	201	176	3.670
STU_1	1000	496	(1.02, −0.72)	483	3	0.435	52	37	40	2.748
STU_2	1000	489	(1.03, −0.67)	479	2	0.528	47	10	14	3.192
SECURA	371	371	(0.81, −0.74)	266	6	1.010	54	53	61	3.423
NASDAQ	1036	570	(1.02, −0.73)	453	3	0.397	58	48	48	2.468
FIRES	2627	2627	(0.48, −0.39)	1917	2	0.136	120	57	71	1.401

In Table 5, we provide the adaptive Hill-estimates, $H_{00} := H(\hat{k}_{0|0})$, and the bootstrap adaptive EVI-estimates $H_0^* := H_{0,n,n_1^*|T}$ and $H^{**} \equiv H_{p^*_{\min}} := H_{p^*_{\min},n,n_1^*|T}$, obtained through Algorithm 5.1. Close to those estimates, and between parenthesis, we place the associated approximate 99% CIs. We further present the ad-hoc estimates $\widetilde{H} = H_{\tilde{p}}(\lfloor n^{0.99} \rfloor)$.

Table 5
Adaptive EVI-estimates and associated 99% CIs obtained through Hill estimators at estimated optimal level (H_{00}), bootstrap adaptive estimates $(H_0^*, H^*_{p^*_{\min}})$ in Algorithm 5.1 and $\widetilde{H} = H_{\tilde{p}}(\lfloor n^{0.99} \rfloor)$, for the data sets under analysis.

Data	$H_{00} := H(\hat{k}_{0\|0})$	$H_0^* := H_{0,n,n_1^*\|T}$	$H^{**} := H_{p^*_{\min},n,n_1^*\|T}$	$\widetilde{H}$
FRE_1	0.272 (0.2020, 0.3406)	0.268 (0.1762, 0.3135)	0.266 (0.1777, 0.3158)	0.252
FRE_2	0.272 (0.2145, 0.3283)	0.276 (0.2012, 0.3015)	0.265 (0.1915, 0.3062)	0.247
STU_1	0.364 (0.2547, 0.5095)	0.359 (0.1870, 0.4913)	0.285 (0.1266, 0.4065)	0.364
STU_2	0.313 (0.2153, 0.4453)	0.242 (0.0381, 0.4323)	0.267 (0.1311, 0.3900)	0.313
SECURA	0.292 (0.1998, 0.3837)	0.297 (0.1591, 0.3692)	0.282 (0.1565, 0.3508)	0.274
NASDAQ	0.405 (0.2876, 0.5531)	0.378 (0.2139, 0.4947)	0.372 (0.2101, 0.4909)	0.405
FIRES	0.714 (0.5332, 0.8220)	0.738 (0.4288, 0.9321)	0.725 (0.4428, 0.8889)	0.682

In Table 6, we provide information on the bootstrap estimates of bias and RMSE of the adaptive estimates obtained in the implementation of Algorithm 5.1, the sizes of the CIs in Table 5, respectively denoted by s_H, s_0^* and $s_{p^*_{\min}}$, and the corrected-bias bootstrap adaptive estimates, $\overline{H}_0^* = H_0^* - b^*_{0,0}$ and $\overline{H}^{**} = H^*_{p^*_{\min}} - b^*_{0,p^*_{\min}}$. The smallest bias and RMSE estimates, the smallest size and the EVI-estimate close to the true value of γ (known only for the four initial samples) are written in **bold**. The second smallest size is written in *italic*.

Finally, and to illustrate the robustness of the method to changes of n_1, we picture Fig. 8, where we represent the bootstrap RMSE estimates associated with $a = 6$, i.e. $p = a/20H_0^*$, for the samples FRE_2 and SECURA.

Table 6
Adaptive bootstrap estimates of bias and RMSE of the adaptive estimates obtained through Algorithm 5.1, sizes of the CIs in Table 5, and corrected-bias bootstrap adaptive EVI-estimates, for the different data sets under analysis.

Data	$b^*_{0,0}$	$b^*_{0,p^*_{min}}$	$\widehat{RMSE}^*_0$	$\widehat{RMSE}^*_{p^*_{min}}$	s_H	s^*_0	$s_{p^*_{min}}$	$\overline{H}^*_0$	$\overline{H}^{**}$
FRE_1	0.023	**0.019**	0.035	**0.033**	0.139	**0.137**	*0.138*	0.245	**0.247**
FRE_2	0.025	**0.016**	0.031	**0.027**	0.114	**0.100**	*0.115*	0.251	**0.249**
STU_1	0.020	**0.019**	0.062	**0.058**	**0.255**	0.304	*0.280*	0.339	**0.267**
STU_2	0.007	**0.006**	0.077	**0.051**	**0.230**	0.394	*0.259*	0.235	**0.261**
SECURA	0.033	**0.029**	0.052	**0.047**	**0.184**	0.210	*0.194*	0.264	0.254
NASDAQ	0.023	**0.021**	0.059	**0.059**	**0.266**	0.281	*0.281*	0.354	0.351
FIRES	**0.057**	0.059	0.113	**0.105**	**0.289**	0.503	*0.446*	0.680	0.666

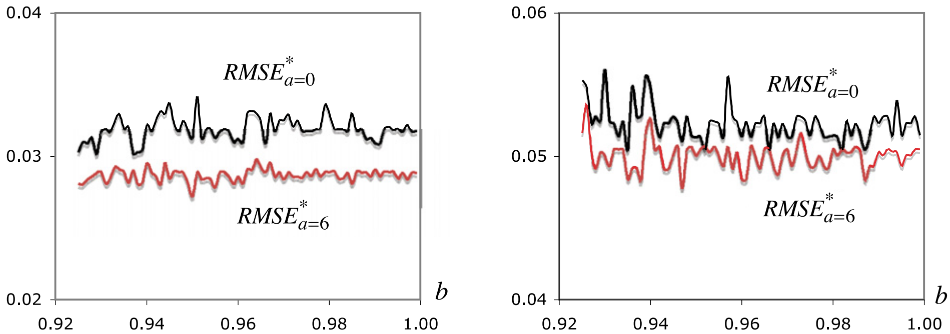

Fig. 8. Bootstrap RMSE estimates for FRE_2 (*left*) and SECURA (*right*), as function of b ($n_1 = n^b$).

7.1. Concluding remarks

- For the four simulated samples, we know the true value of γ, the value 0.25, and we see that such a value belongs to all 99% CIs, but the one associated with the Hill estimate and the STU_1 sample. It is again clear that Hill's estimation leads to a strong over-estimation of the EVI and the adaptive MOP provides a more adequate EVI-estimation.
- The size of the MOP-CI is always the second largest, but the smallest RMSE is always the one associated with the MOP EVI-estimators. A similar comment applies to the smallest BIAS, excluding the FIRES sample. These are obviously arguments in favour of the new methodology.
- These case studies claim for a Monte-Carlo derivation of the properties of the adaptive MOP EVI-estimate provided by Algorithm 5.1. Also, a robust version of these MOP EVI-estimators, similar to the one in Beran and Schell (2012), can be of high practical interest. These are however topics out of the scope of this article.
- Results obtained for other simulated samples, not presented here, clearly indicate an over-estimation of the most common adaptive Hill estimate and an overall best performance of this data-driven MOP method of estimation of the EVI.
- The ad-hoc choice $\widetilde{H}$ works only when H_{00} does not provide a clear over-estimation of the EVI, as happens with the *Fréchet* samples.

Acknowledgements

The research was partially supported by National Funds through **FCT**—Fundação para a Ciência e a Tecnologia, project PEst-OE/MAT/UI0006/2011, and PTDC/FEDER, EXTREMA. We would also like to thank the reviewers for valuable comments on the first version of this paper.

References

Araújo Santos, P., Fraga Alves, M.I., Gomes, M.I., 2006. Peaks over random threshold methodology for tail index and quantile estimation. REVSTAT 4 (3), 227–247.
Beirlant, J., Caeiro, F., Gomes, M.I., 2012. An overview and open researh topics in statistics of univariate extremes. REVSTAT 10 (1), 1–31.
Beirlant, J., Dierckx, G., Guillou, A., 2005. Estimation of the extreme-value index and generalized quantile plots. Bernoulli 11 (6), 949–970.
Beirlant, J., Figueiredo, F., Gomes, M.I., Vandewalle, B., 2008. Improved reduced-bias tail index and quantile estimators. J. Statist. Plann. Inference 138 (6), 1851–1870.
Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. Statistics of Extremes, Theory and Applications. Wiley.
Beirlant, J., Vynckier, P., Teugels, J., 1996. Excess functions and estimation of the extreme-value index. Bernoulli 2, 293–318.
Beran, J., Schell, D., 2012. On robust tail index estimation. Comput. Statist. Data Anal. 56 (11), 3430–3443.
Bingham, N., Goldie, C.M., Teugels, J.L., 1987. Regular Variation. Cambridge University Press, Cambridge.
Brilhante, M.F., Gomes, M.I., Pestana, D.D., 2012. The MOP EVI-estimator revisited, Preprint (submitted for publication).
Caeiro, C., Gomes, M.I., 2012. A semi-parametric estimator of a shape second order parameter. Notas e Comunicações CEAUL 07/2012.

Caeiro, C., Gomes, M.I., 2006. A new class of estimators of a scale second order parameter. Extremes 9, 193–211.
Caeiro, F., Gomes, M.I., Pestana, D.D., 2005. Direct reduction of bias of the classical Hill estimator. REVSTAT 3 (2), 111–136.
Ciuperca, G., Mercadier, C., 2010. Semi-parametric estimation for heavy tailed distributions. Extremes 13 (1), 55–87.
Davison, A., 1984. Modeling excesses over high threshold with an application. In: Tiago de Oliveira, J. (Ed.), Statistical Extremes and Applications. D. Reidel, pp. 461–482.
de Haan, L., 1984. Slow variation and characterization of domains of attraction. In: Tiago de Oliveira, J. (Ed.), Statistical Extremes and Applications. D. Reidel, Dordrecht, pp. 31–48.
de Haan, L., Ferreira, A., 2006. Extreme Value Theory: An Introduction. Springer Science + Business Media, LLC, New York.
de Haan, L., Peng, L., 1998. Comparison of extreme value index estimators. Stat. Neerl. 52, 60–70.
Dekkers, A.L.M., de Haan, L., 1993. Optimal choice of sample fraction in extreme-value estimation. J. Multivariate Anal. 47, 173–195.
Dekkers, A., Einmahl, J., de Haan, L., 1989. A moment estimator for the index of an extreme-value distribution. Ann. Statist. 17, 1833–1855.
Drees, H., 2003. Extreme quantile estimation for dependent data, with applications to finance. Bernoulli 9 (4), 617–657.
Drees, H., Ferreira, A., de Haan, L., 2004. On maximum likelihood estimation of the extreme value index. Ann. Appl. Probab. 14, 1179–1201.
Fraga Alves, M.I., Gomes, M.I., de Haan, L., 2003. A new class of semi-parametric estimators of the second order parameter. Port. Math. 60 (1), 193–213.
Fraga Alves, M.I., Gomes, M.I., de Haan, L., Neves, C., 2009. The mixed moment estimator and location invariant alternatives. Extremes 12, 149–185.
Geluk, J., de Haan, L., 1987. Regular Variation, Extensions and Tauberian Theorems, CWI Tract 40. Center for Mathematics and Computer Science, Amsterdam, Netherlands.
Gnedenko, B.V., 1943. Sur la distribution limite du terme maximum d'une série aléatoire. Ann. of Math. 44 (6), 423–453.
Goegebeur, Y., Beirlant, J., de Wet, T., 2008. Linking Pareto-tail kernel goodness-of-fit statistics with tail index at optimal threshold and second order estimation. REVSTAT 6 (1), 51–69.
Goegebeur, Y., Beirlant, J., de Wet, T., 2010. Kernel estimators for the second order parameter in extreme value statistics. J. Statist. Plann. Inference 140 (9), 2632–2654.
Gomes, M.I., de Haan, L., Henriques-Rodrigues, L., 2008. Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. J. Roy. Statist. Soc. B 70 (1), 31–52.
Gomes, M.I., Figueiredo, F., Neves, M.M., 2012. Adaptive estimation of heavy right tails: the bootstrap methodology in action, Extremes. http://dx.doi.org/10.1007/s10687-011-0146-6 (in press-a).
Gomes, M.I., Henriques-Rodrigues, L., 2010. Comparison at optimal levels of classical tail index estimators: a challenge for reduced-bias estimation? Discuss. Math. Probab. Stat. 30 (1), 35–51.
Gomes, M.I., Henriques-Rodrigues, L., Fraga Alves, M.I., Manjunath, B.G., 2012. Adaptive PORT-MVRB estimation: an empirical eomparison of two heuristic algorithms. J. Stat. Comput. Simul. http://dx.doi.org/10.1080/00949655.2011.652113 (in press-b).
Gomes, M.I., Henriques-Rodrigues, L., Pereira, H., Pestana, D., 2010. Tail index and second order parameters' semi-parametric estimation based on the log-excesses. J. Stat. Comput. Simul. 80 (6), 653–666.
Gomes, M.I., Martins, M.J., 2002. Asymptotically unbiased estimators of the extreme value index based on external estimation of the second order parameter. Extremes 5 (1), 5–31.
Gomes, M.I., Martins, M.J., 2001. Generalizations of the Hill estimator—asymptotic versus finite sample behaviour. J. Statist. Plann. Inference 93, 161–180.
Gomes, M.I., Martins, M.J., Neves, M.M., 2011a. Revisiting the Role of the Generalized JackKnife Methodology in the Field of Extremes. Notas e Comunicações CEAUL 20/2011.
Gomes, M.I., Mendonça, S., Pestana, D.D., 2011b. Adaptive reduced-bias tail index and VaR estimation via the bootstrap methodology. Comm. Statist. Theory Methods 40 (16), 2946–2968.
Gomes, M.I., Miranda, C., Pereira, H., 2005. Revisiting the role of the Jackknife methodology in the estimation of a positive extreme value index. Comm. Statist. Theory Methods 34, 1–20.
Gomes, M.I., Miranda, C., Viseu, C., 2007. Reduced-bias tail index estimation and the Jackknife methodology. Stat. Neerl. 61 (2), 243–270.
Gomes, M.I., Neves, C., 2008. Asymptotic comparison of the mixed moment and classical extreme value index estimators. Statist. Probab. Lett. 78 (6), 643–653.
Gomes, M.I., Oliveira, O., 2001. The bootstrap methodology in Statistics of Extremes: choice of the optimal sample fraction. Extremes 4 (4), 331–358.
Gomes, M.I., Pestana, D.D., 2007a. A simple second order reduced-bias extreme value index estimator. J. Statist. Comput. Simul. 77 (6), 487–504.
Gomes, M.I., Pestana, D.D., 2007b. A sturdy reduced-bias extreme quantile (VaR) estimator. J. Amer. Statist. Assoc. 102 (477), 280–292.
Hall, P., 1982. On some simple estimates of an exponent of regular variation. J. Roy. Statist. Soc. B 44, 37–42.
Hall, P., Welsh, A.W., 1985. Adaptive estimates of parameters of regular variation. Ann. Statist. 13, 331–341.
Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. Ann. Statist. 3, 1163–1174.
MacDonald, A., Scarrott, C.J., Lee, D., Darlow, B., Reale, M., Russell, G., 2011. A flexible extreme value mixture model. Comput. Statist. Data Anal. 55 (6), 2137–2157.
Scarrott, C.J., MacDonald, A., 2012. A review of extreme value threshold estimation and uncertainty quantification. REVSTAT 10 (1), 33–60.
Smith, R.L., 1987. Estimating tails of probability distributions. Ann. Statist. 15 (3), 1174–1207.
Vandewalle, B., Beirlant, J., 2006. On univariate extreme value statistics and the estimation of reinsurance premiums. Insurance Math. Econom. 38 (3), 444–459.