A simple generalization of the Hill estimator

Maria de Fátima Brilhante

CEAUL and DM, Universidade dos Açores, e-mail: fbrilhante@uac.pt

M. Ivette Gomes

CEAUL and DEIO, FCUL, Universidade de Lisboa, e-mail: ivette.gomes@fc.ul.pt

Dinis Pestana

CEAUL and DEIO, FCUL, Universidade de Lisboa, e-mail: dinis.pestana@fc.ul.pt

March 27, 2012

Abstract

In this paper we are interested in a simple generalization of the classical Hill estimator of a positive extreme value index (EVI), the primary parameter of extreme events. The Hill estimator can be regarded as the geometric mean, or equivalently the mean of order $p = 0$, of a set of adequate statistics. Instead of such a geometric mean, we shall more generally consider the mean of order $p \geq 0$ of those statistics. Apart from the derivation of the asymptotic behaviour of the new class of EVI-estimators, we shall proceed with an asymptotic comparison, at optimal levels, of the members of such a class, and an adaptive estimation of the tuning parameters under play. Moreover, a large-scale simulation study will be developed as well as an application to simulated and real data.

AMS 2000 subject classification. Primary 62G32; Secondary 65C05.

Keywords and phrases. *Statistics of extremes*; *Semi-parametric estimation*; *Bias estimation*; *Heavy tails*; *Optimal levels*; *Bootstrap methodology.*

1 Estimators under study 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), $X_1, X_2, \ldots, X_n$, with a common distribution function (d.f.) F. Let us denote by $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ the associated ascending order statistics (o.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 normalized, 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

$$EV_\gamma(x) = \begin{cases} \exp(-(1+\gamma x)^{-1/\gamma}),\ 1+\gamma x > 0 & \text{if} \quad \gamma \neq 0 \\ \exp(-\exp(-x)),\ x \in \mathbb{R} & \text{if} \quad \gamma = 0. \end{cases} \tag{1.1}$$

The d.f. F is said to belong to the *max-domain of attraction* of EV_γ, and we write $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$. The parameter γ, in (1.1), is the *extreme value index* (EVI), the primary parameter of extreme events.

Let us denote RV_a the class of regularly varying functions at infinity, with an index of regular variation equal to a, 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*

$$\overline{F}(x) := 1 - F(x), \tag{1.2}$$

and the heavier the right tail, the larger γ is. In this paper we shall essentially 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 $L \in RV_0$, 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 the most diversified areas of application, like computer science, telecommunications, insurance, finance, bibliometrics and biostatistics, among others.

1.1 A brief review of 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 $\overline{F}$, in (1.2), is a *regularly varying function* with a negative index of regular variation equal to $-1/\gamma$, $\gamma > 0$, or equivalently, with $F^{\leftarrow}(x) := \inf\{y : F(y) \geq x\}$ denoting the generalized inverse function of F, the (reciprocal) quantile function

$$U(t) := F^{\leftarrow}(1 - 1/t),\ t \geq 1, \tag{1.3}$$

is of regular variation with index γ, i.e.

$$F \in \mathcal{D}_{\mathcal{M}}^{+} := \mathcal{D}_{\mathcal{M}}\left(EV_\gamma\right)_{\gamma>0} \quad \Longleftrightarrow \quad \overline{F} \in RV_{-1/\gamma} \quad \Longleftrightarrow \quad U \in RV_\gamma \tag{1.4}$$

for all $x > 0$ (Gnedenko, 1943; de Haan, 1970).

The *second-order parameter* ρ (≤ 0) rules the rate of convergence in the first-order condition, in (1.4), and, with U given in (1.3), it is the non-positive parameter appearing in the limiting relation

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho}, \tag{1.5}$$

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).

1.2 The new class of EVI-estimators

For heavy-tailed models in $\mathcal{D}_{\mathcal{M}}^{+}$, the classical EVI-estimators are the Hill estimators (Hill, 1975), which are the averages of the log-excesses, given by

$$V_{ik} := \ln \frac{X_{n-i+1:n}}{X_{n-k:n}}, \quad 1 \leq i \leq k < n. \tag{1.6}$$

We thus have

$$H(k) := \frac{1}{k}\sum_{i=1}^{k} V_{ik}, \quad 1 \leq k < n. \tag{1.7}$$

Note that we can write the distributional identity $X = U(Y)$, with Y a standard Pareto r.v., i.e. a r.v. with d.f. $F_Y(y) = 1 - 1/y,\ y \geq 1$. For the o.s. associated with a strict Pareto sample $(Y_1, Y_2, \ldots, Y_n)$, we have $Y_{n-i+1:n}/Y_{n-k:n} \stackrel{d}{=} Y_{k-i+1:k},\ 1 \leq i \leq k$. Moreover, $kY_{n-k:n}/n \underset{n\to\infty}{\stackrel{p}{\longrightarrow}} 1$, i.e. $Y_{n-k:n} \stackrel{p}{\sim} n/k$. Consequently, and provided that $k \to \infty$, with $k/n \to 0$, as $n \to \infty$,

$$U_{ik} := \frac{X_{n-i+1:n}}{X_{n-k:n}} = \frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})} = \frac{U(Y_{n-k:n}Y_{k-i+1:k})}{U(Y_{n-k:n})} = Y_{k-i+1:k}^{\gamma}(1 + o_p(1)), \tag{1.8}$$

i.e. $U_{ik} \stackrel{p}{\sim} Y_{k-i+1:k}^{\gamma}$. Hence, we have the approximation $\ln U_{ik} \approx \gamma\ \ln Y_{k-i+1:k} = \gamma E_{k-i+1:k}$, $1 \leq i \leq k$, with E denoting a standard exponential r.v. The log-excesses, $V_{ik} = \ln U_{ik}$, $1 \leq i \leq k$, in (1.6), are thus approximately the k top o.s. of a sample of size k from an exponential parent with mean value γ. This justifies the Hill EVI-estimator, in (1.7).

Note now that we can write

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

the logarithm of the *geometric mean* of the statistics $U_{ik} := X_{n-i+1:n}/X_{n-k:n}$, given in (1.8). More generally, we shall now consider as basic statistics for the EVI estimation, the *mean of order* p (MOP) of U_{ik}, i.e. the class of statistics

$$A_p(k) = \begin{cases} \left(\frac{1}{k}\sum_{i=1}^{k} U_{ik}^{p}\right)^{1/p} & \text{if} \quad p > 0 \\ \left(\prod_{i=1}^{k} U_{ik}\right)^{1/k} & \text{if} \quad p = 0. \end{cases} \tag{1.9}$$

From (1.8), we can write $U_{ik}^{p} = Y_{k-i+1:k}^{\gamma p}(1 + o_p(1))$. Since

$$\mathbb{E}(Y^a) = \mathbb{E}(e^{a\ln Y}) = \mathbb{E}(e^{aE}) = \int_0^\infty e^{-(1-a)x}dx = \frac{1}{1-a} \quad \text{if} \quad a < 1, \tag{1.10}$$

$$A_p(k) \underset{n\to\infty}{\stackrel{p}{\longrightarrow}} \left(\frac{1}{1-\gamma p}\right)^{1/p}, \quad \text{i.e.} \quad \frac{1 - A_p^{-p}(k)}{p} = \frac{1 - \exp(-p\ln A_p(k))}{p} \underset{n\to\infty}{\longrightarrow} \gamma,$$

if $p < 1/\gamma$.

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

$$H_p(k) := \begin{cases} \frac{1}{p}\left(1 - \exp(-p \ln A_p(k))\right) & \text{if} \quad p > 0 \\ \\ \ln A_0(k) = H(k) & \text{if} \quad p = 0, \end{cases} \tag{1.11}$$

dependent now on this *tuning* parameter $p \geq 0$, and with $A_p(k)$ given in (1.9), and with $H_0(k) \equiv H(k)$, given in (1.7).

1.3 Scope of the paper

In this paper, we shall deal in Section 2 with the derivation of the asymptotic behaviour of the new class of MOP EVI-estimators in (1.11). In Section 3 we compare asymptotically, at optimal levels, the above mentioned class of estimators, drawing some concluding remarks. In Section 4, we provide a method for the adaptive choice of the tuning parameters k and p, on the basis of the bootstrap methodology. Section 5 is dedicated to the finite sample properties of the new class of estimators, done through a large-scale simulation study. Finally, in Section 6, we illustrate the behaviour of the new class of MOP EVI-estimators, together with the adaptive choice provided in Section 4, through an application to simulated random samples, as well as to sets of real data in the fields of insurance, finance and environment.

2 Asymptotic behaviour of the class of MOP EVI-estimators

In order to have consistency of the Hill estimator, in (1.7), in all $\mathcal{D}_{\mathcal{M}}^+$, we need to work with *intermediate* values of k, i.e. a sequence of integers $k = k_n$, $1 \leq k < n$, such that

$$k = k_n \to \infty \quad \text{and} \quad k_n = o(n), \text{ as } n \to \infty. \tag{2.1}$$

Under the second-order framework, in (1.5), the asymptotic distributional representation

$$H(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k + \frac{1}{1-\rho}\, A(n/k)(1 + o_p(1)) \tag{2.2}$$

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

$$Z_k = \sqrt{k}\Big(\sum_{i=1}^{k} E_i/k - 1\Big) \tag{2.3}$$

is an asymptotically standard normal r.v.

We now state the main theorem in this paper.

Theorem 1. *Under the validity of the first-order condition, in* (1.4)*, and for intermediate sequences* $k = k_n$*, i.e. if* (2.1) *holds, the class of estimators* $H_p(k)$*, in* (1.11)*, is consistent for the estimation of* γ*, provided that* $p < 1/\gamma$*.*

If we moreover assume the validity of the second-order condition, in (1.5)*, the asymptotic distributional representation*

$$H_p(k) \;\stackrel{d}{=}\; \gamma + \frac{\sigma_p(\gamma)\; Z_k^{(p)}}{\sqrt{k}} + b_p(\gamma|\rho)\; A(n/k) + o_p(A(n/k)) \tag{2.4}$$

holds for $p < 1/(2\gamma)$*, with* $Z_k^{(p)}$ *asymptotically standard normal,*

$$\sigma_p(\gamma) := \frac{\gamma(1-p\gamma)}{\sqrt{1-2p\gamma}} \quad \text{and} \quad b_p(\gamma|\rho) := \frac{1-p\gamma}{1-p\gamma-\rho}. \tag{2.5}$$

Proof. As we have seen before, on the basis of (1.10) and the law of large numbers, the statistics in (1.11) are consistent for the estimation of γ for all $p < 1/\gamma$. With Y denoting again a strict Pareto r.v., and working under the second-order framework in (1.5), we can write

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

Consequently,

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

On the basis of (1.10), and for $a < 1/2$, we have

$$\mathbb{V}ar(Y^a) = \frac{1}{1-2a} - \left(\frac{1}{1-a}\right)^2 = \frac{a^2}{(1-a)^2(1-2a)}.$$

We thus know that for $p < 1/(2\gamma)$,

$$\frac{\sqrt{k}(1-p\gamma)\sqrt{1-2p\gamma}\left(\frac{1}{k}\sum_{i=1}^{k} Y_i^{p\gamma} - \frac{1}{1-p\gamma}\right)}{p\gamma} =: Z_k^{(p)} \tag{2.6}$$

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 $p < 1/\gamma$,

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

Let's go back to the EVI-estimator in (1.11):

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

We can thus further write

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

i.e. (2.4) follows, with $\sigma_p(\gamma)$ and $b_p(\gamma|\rho)$ given in (2.5). ■

Remark 1. *For* $p = 0$, $Z_k^{(0)} \equiv Z_k$, *with* Z_k *and* $Z_k^{(p)}$ *given in* (2.3) *and* (2.6), *respectively, and on the basis of* (2.4), *we get for* $H_0(k) \equiv H(k)$, *in* (1.7), *the particular result in* (2.2), *as derived in de* Haan and Peng (1998).

Remark 2. *Note that that, for any* $\gamma > 0$, *the asymptotic standard deviation* $\sigma_p(\gamma) = \gamma(1-p\gamma)/\sqrt{1-2p\gamma}$, *in* (2.5), *is increasing in* $p \geq 0$. *In* Figure 1, *we present such a standard deviation, as a function of* p, *for* $\gamma = 0.1, 0.5$ *and 1.*

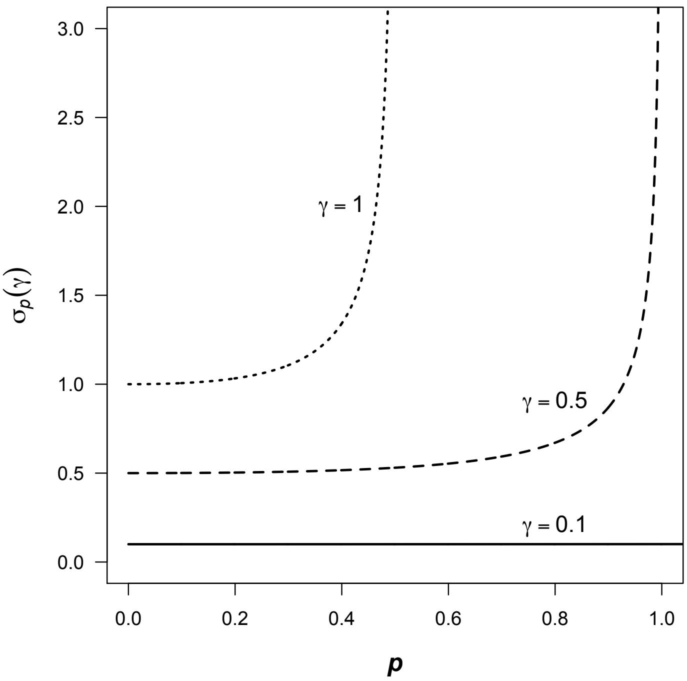

Figure 1: The asymptotic standard deviation $\sigma_p(\gamma)$ for $\gamma = 0.1, 0.5$ and 1, as a function of $p \geq 0$.

Remark 3. *On the other side, also for any* $\gamma > 0$, $\rho < 0$ *and* $p \neq (1-\rho)/\gamma$, *the asymptotic bias* $b_p(\gamma|\rho) = (1-p\gamma)/(1-p\gamma-\rho)$, *also in* (2.5), *is decreasing in* p. *Such a performance is shown in* Figure 2, *where we present graphically such a bias, as a function of* p, *again for values of* $\gamma = 0.1, 0.5$ *and* 1.

These aforementioned results claim for an asymptotic comparison, at optimal levels of the class of EVI-estimators in (1.11), a topic to be dealt with next, in Section 3.

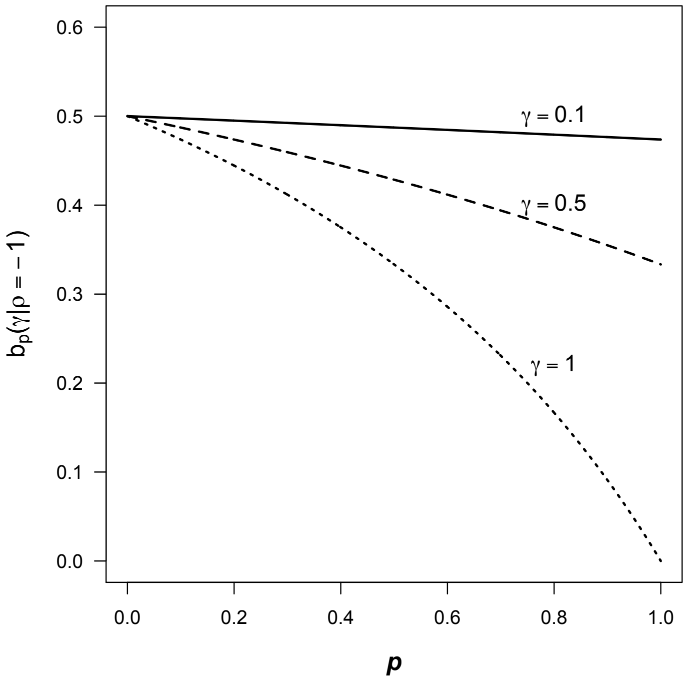

Figure 2: The asymptotic bias ruler, $b_p(\gamma|\rho = -1)$, for $\gamma = 0.1, 0.5$ and 1, as a function of $p \geq 0$.

3 Asymptotic comparison of MOP EVI-estimators at optimal levels

We shall next proceed to the comparison of the estimators under study 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

$$\widehat{\gamma}_n^\bullet(k) = \gamma + \frac{\sigma_\bullet\ Z_k^\bullet}{\sqrt{k}} + b_\bullet\ A(n/k) + o_p(A(n/k)), \tag{3.1}$$

for any intermediate sequence of integers $k = k_n$, and where $Z_k^\bullet$ is asymptotically standard normal. Then, $\sqrt{k}\big(\widehat{\gamma}_n^\bullet(k) - \gamma\big) \overset{d}{\to} N(\lambda_{_A} b_\bullet, \sigma_\bullet^2)$ provided that k is such that $\sqrt{k}\ A(n/k) \to \lambda_{_A}$, finite, as $n \to \infty$. We then write $Bias_\infty\big(\widehat{\gamma}_n^\bullet(k)\big) := b_\bullet\ A(n/k)$ and $Var_\infty\big(\widehat{\gamma}_n^\bullet(k)\big) := \sigma_\bullet^2/k$.

The so-called *asymptotic mean square error* (AMSE) is then given by

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

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

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

where $\widehat{\gamma}_{n0}^\bullet := \widehat{\gamma}_n^\bullet(k_{0|\bullet}(n))$ and $k_{0|\bullet}(n) := \arg\min_k \mathrm{MSE}\,(\widehat{\gamma}_n^\bullet(k))$. Moreover, if we slightly restrict the second-order condition in (1.5), assuming that

$$A(t) = \gamma\beta t^\rho, \quad \rho < 0, \tag{3.2}$$

we can write

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

Let us now turn back to the MOP EVI-estimator $H_p(k)$ in (1.11). We have

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

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

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

In order to measure the performance of H_{p0}, it is then sensible to consider again the following:

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* (3.1) *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*

$$\mathrm{AREFF}_{1|2} \equiv \mathrm{AREFF}_{\widehat{\gamma}_{n0}^{(1)}|\widehat{\gamma}_{n0}^{(2)}} := \sqrt{\frac{\mathrm{LMSE}\big(\widehat{\gamma}_{n0}^{(2)}\big)}{\mathrm{LMSE}\big(\widehat{\gamma}_{n0}^{(1)}\big)}} = \left(\left(\frac{\sigma_2}{\sigma_1}\right)^{-2\rho}\left|\frac{b_2}{b_1}\right|\right)^{\frac{1}{1-2\rho}}. \tag{3.3}$$

Remark 4. *Note that the* AREFF *indicator, in* (3.3)*, has been conceived so that the highest the* AREFF *indicator is, the better is the first estimator.*

For the class of MOP EVI-estimators in (1.11), we get the AREFF-indicator

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

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

$$\mathrm{AREFF}^*_{a|0} = \left(\left(\frac{\sqrt{1-2a}}{1-a} \right)^{-2\rho} \left| \frac{1-a-\rho}{(1-\rho)(1-a)} \right| \right)^{\frac{1}{1-2\rho}}. \quad (3.4)$$

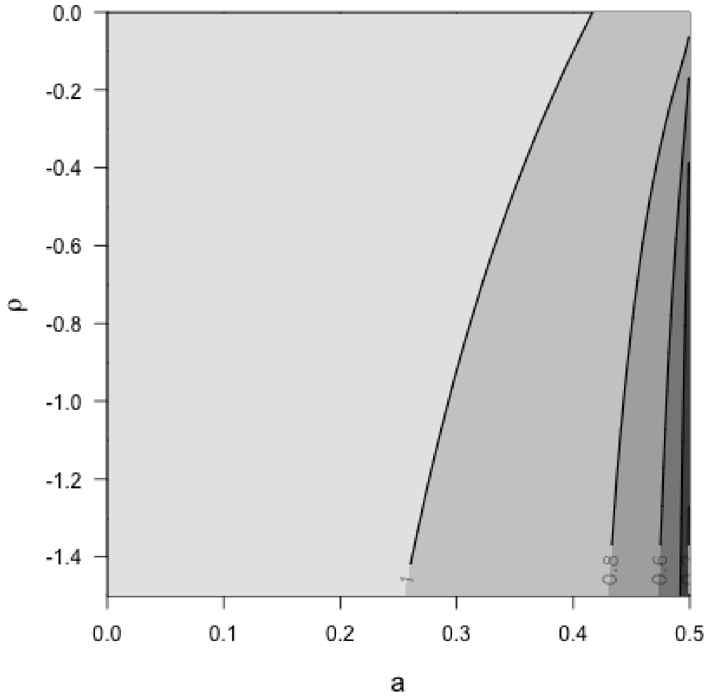

Figure 3: The indicator $ARREF^*_{a|0}$, in (3.4), as a function of (a, ρ).

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 estimators performs better than the Hill estimator.

4 An adaptive choice of p and k

For the new class of MOP EVI-estimators $H_p(k)$, in (1.11), let us use the notations

$$k_{0|p}(n) := \arg\min_k \mathrm{MSE}(H_p(k)) = k_{A|p}(n)(1 + o(1)), \tag{4.1}$$

with

$$k_{A|p}(n) := \arg\min_k \mathrm{AMSE}\big(H_p(k)\big). \tag{4.2}$$

For any admissible p, i.e. $p < 1/(2\gamma)$, and provided that we first estimate γ adequately, the bootstrap methodology can thus enable us to consistently estimate the optimal sample fraction (OSF), $k_{0|p}(n)/n$, with $k_{0|p}(n)$ defined in (4.1), on the basis of a consistent estimator of $k_{A|p}(n)$, in (4.2), in a way similar to the one used in Draisma *et al.* (1999), Danielson *et al.* (2001) and Gomes and Oliveira (2001), for the classical adaptive Hill EVI estimation, performed through $H(k) \equiv H_0(k)$, in (1.7), and for second-order reduced-bias estimation in Gomes *et al.* (2011b, 2012a). With the usual notation $[x]$ for the integer part of x, we shall here use again the auxiliary statistics

$$T_{k,n} \equiv T(k|H_p) \equiv T_{k,n|p} := H_p([k/2]) - H_p(k), \quad k = 2, \ldots, n-1, \tag{4.3}$$

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 (1.5), we get, for all $p \ge 0$,

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

with $P_k^{(p)}$ asymptotically standard normal, and $(\sigma_p(\gamma), b_p(\gamma|\rho))$ given in (2.5).

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

$$k_{0|p}(n) = k_{0|T}(n) \times (1 - 2^\rho)^{\frac{2}{1-2\rho}}\,(1 + o(1)). \tag{4.4}$$

Given the sample $\underline{X}_n = (X_1, \ldots, X_n)$ from any unknown model F, and the functional in (4.3), $T_{k,n} =: \phi_k(\underline{X}_n)$, $1 < k < n$, 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

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

the empirical d.f. associated with the available sample, $\underline{X}_n$.

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

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

Consequently, for another sample size, $n_2 = [n_1^2/n] + 1$, we have

$$\big(k^*_{0|T}(n_1)\big)^2 / k^*_{0|T}(n_2) = k_{0|T}(n)(1+o(1)), \quad \text{as } n \to \infty. \tag{4.5}$$

On the basis of (4.5), we are now able to first consistently estimate $k_{0|T}$, and next $k_{0|p}$ through (4.4), 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

$$\hat k^*_{0|p} \equiv \hat k^*_{0|p}(n; n_1) := \min\Big(n-1,\ \big[c_{\hat\rho}\ (\hat k^*_{0|T}(n_1))^2 / \hat k^*_{0|T}([n_1^2/n]+1)\big] + 1\Big). \tag{4.6}$$

The adaptive estimate of γ is then given by

$$H^*_p \equiv H^*_{p,n,n_1|T} := H_p(\hat k^*_{0|p}(n; n_1)). \tag{4.7}$$

4.1 An algorithm for an adaptive MOP EVI-estimation

We now proceed with the description of an algorithm for the adaptive estimation of γ, where in **Steps 1**, **2** and **3** we reproduce the algorithm provided in Gomes and Pestana (2007) for the estimation of the second-order parameters β and ρ.

Algorithm 4.1.

Step 1 *Given an observed sample* $(x_1, \ldots, x_n)$, *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*

$$\hat{\rho}_\tau(k) := \min\left(0,\ 3(W_{k,n}^{(\tau)} - 1)/(W_{k,n}^{(\tau)} - 3)\right),$$

dependent on the statistics

$$W_{k,n}^{(0)} := \frac{\ln\left(M_{k,n}^{(1)}\right) - \frac{1}{2}\ln\left(M_{k,n}^{(2)}/2\right)}{\frac{1}{2}\ln\left(M_{k,n}^{(2)}/2\right) - \frac{1}{3}\ln\left(M_{k,n}^{(3)}/6\right)}, \quad W_{k,n}^{(1)} := \frac{M_{k,n}^{(1)} - \left(M_{k,n}^{(2)}/2\right)^{1/2}}{\left(M_{k,n}^{(2)}/2\right)^{1/2} - \left(M_{k,n}^{(3)}/6\right)^{1/3}},$$

where, with V_{ik}, $1 \le i \le k$, *the log-excesses, given in* (1.6),

$$M_{k,n}^{(j)} := \frac{1}{k}\sum_{i=1}^{k} V_{ik}^j, \quad j = 1, 2, 3.$$

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

Step 3 *Work with* $\hat{\rho} \equiv \hat{\rho}_{\tau^*} = \hat{\rho}_{\tau^*}(k_1)$ *and* $\hat{\beta} \equiv \hat{\beta}_{\tau^*} := \hat{\beta}_{\hat{\rho}_{\tau^*}}(k_1)$, *with* $k_1 = [n^{0.999}]$, *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 4 *Compute* $H_0(k) \equiv H(k)$, $k = 1, 2, \ldots, n-1$, *given in* (1.7).

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

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,*

$$(x_1^*, \dots, x_{n_2}^*) \quad and \quad (x_1^*, \dots, x_{n_2}^*, x_{n_2+1}^*, \dots, x_{n_1}^*),$$

bootstrap samples of sizes n_2 and n_1, respectively.

Step 7 *Denoting $T_{k,n}^*$ the bootstrap counterpart of $T_{k,n}$, in (4.3), 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 $k = 2, \dots, n_i - 1$,*

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

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 (4.6).*

Step 10 *Compute $H_0^* \equiv H_{0,n,n_1|T}^*$, with H_p^* given in (4.7), as well as the MSE-estimate*

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

Step 11 *For $p = a/(20H_0^*)$, with H_0^* the estimate obtained in* **Step 10**, *and $a = 1(1)9$, compute $H_p(k)$, $k = 1, 2, \dots, n-1$, 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 (4.6).*

Step 13 *Compute $H_p^* \equiv H_{p,n,n_1|T}^*$, given in (4.7), as well as the MSE-estimate*

$$\begin{aligned}\widehat{MSE}_p^* &\equiv \widehat{MSE}_p^*(n_1) \equiv \widehat{\mathrm{MSE}}(\hat{k}_{0|p}^*|H_p^*) \\ &:= \frac{\sigma_p^2(H_p^*)}{\hat{k}_{0|p}^*} + \left(\frac{H_p^*\hat{\beta}(1-pH_p^*)\left(n/\hat{k}_{0|p}^*\right)^{\hat{\rho}}}{1-pH_p^*-\hat{\rho}}\right)^2 =: (\hat{\sigma}_{0p}^*)^2 + (\hat{b}_{0p}^*)^2, \quad (4.8)\end{aligned}$$

where $\sigma_p(\gamma)$ has been defined in (2.5).

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

Step 15 *Choose* $n^*_1 := \arg\min_{n_1} \widehat{MSE}^*_{p^*_{min}}(n_1)$*, with* $\widehat{MSE}^*_p(n_1)$*, 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 5. *For any* $p \geq 0$*, and with* $\hat{k}^*_{0|p}$ *and* $(\sigma^*_{0p}, b^*_{0p})$ *given in* (4.6) *and* (4.8)*, respectively, the r.v.* $\big(H_p(\hat{k}^*_{0|p}) - \gamma - b^*_{0p}\big)/\sigma^*_{0p}$ *is approximately* Normal(0,1). *We can then get approximate* $100(1-\alpha)\%$ *confidence intervals* (CIs) *for* γ*, given by*

$$\Big(H_p(\hat{k}^*_{0|p}) - b^*_{0p} - \xi_{1-\alpha/2}\sigma^*_{0p},\ H_p(\hat{k}^*_{0|p}) - b^*_{0p} + \xi_{1-\alpha/2}\sigma^*_{0p}\Big).$$

where ξ_p *denotes the quantile of probability* p *of a standard normal d.f.*

Remark 6. *We make the following comments:*

(i) *If there are negative elements in the sample, the value of* n*, in* Algorithm 4.1*, 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 on several papers essentially related with bias reduction, in* **Step 2** *of* Algorithm 4.1 *we are led in almost all situations to the tuning parameter* $\tau = 0$ *whenever* $-1 \leq \rho < 0$ *and* $\tau = 1$*, otherwise. We thus claim again for the relevance of the choice* $\tau = 0$*, the one considered in the applications in* Section 6.

(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) *and in* Ciuperca and Mercadier (2010)*, as well as to the estimators of* β *in* Caeiro and Gomes (2006) *and in* Gomes *et al.* (2010).

(iv) *In the algorithm 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 4.1. *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 the* **Steps 6–16** *of* Algorithm 4.1 *can be replicated* r_1 *times if we want to associate standard 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 us to increase the precision of the result with a smaller* B, *the number of bootstrap samples generated. Indeed, if we had generated the sample of size* n_1 *independently of the sample of size* n_2, *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.

5 Finite sample properties of the new class of EVI-estimators

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

(1) the *Fréchet* model, with d.f. $F(x) = \exp(-x^{-1/\gamma}),\ x \geq 0$, for $\gamma = 0.1, 0.25$ and 0.5;

(2) the *extreme value* model, with d.f. $F(x) = EV_\gamma(x)$, with $EV_\gamma(x)$ given in (1.1), also for $\gamma = 0.1, 0.25$ and 0.5;

(3) the *generalized Pareto* model, with d.f. $F(x) = 1 + \ln EV_\gamma(x) = 1 - \big(1 + \gamma x\big)^{-1/\gamma}$, $0 \leq x < -1/\gamma$, $EV_\gamma(x)$ given in (1.1), also for $\gamma = 0.1, 0.25$ and 0.5;

(4) the *Student*-t_ν, with $\nu = 2, 4$, i.e. for values of $\gamma = 0.5, 0.25$ ($\gamma = 1/\nu$).

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

5.1 Mean values and mean square error patterns of the MOP EVI-estimators, as functions of k

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 (1.11), as functions of the number of top order statistics k involved in the estimation and for $p = j/(10\gamma)$, $j = 0, 1, 2, 3, 4$. Some of those values, based on the first replicate with a size 5000, are pictured in Figures 4 and 5, for samples of size $n = 1000$ from *extreme value* underlying parents with $\gamma = 0.1$ and $\gamma = 0.5$, respectively.

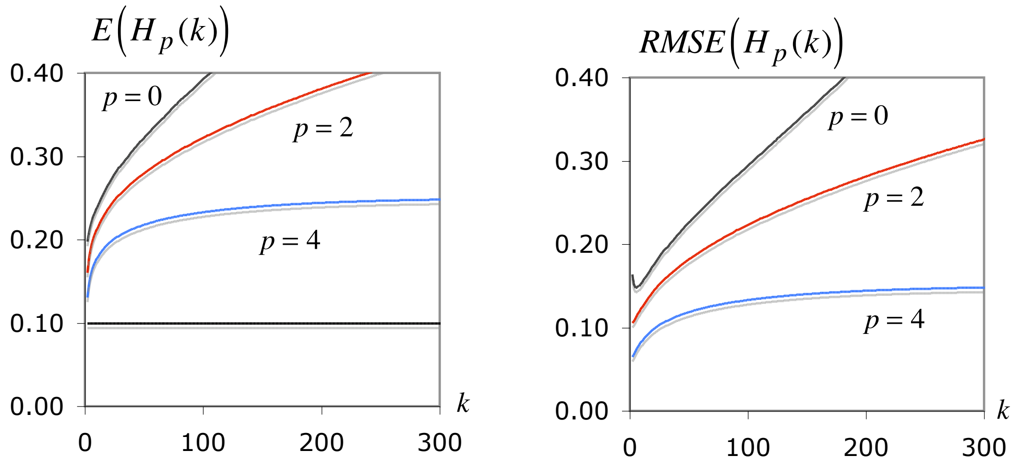

Figure 4: Mean values (*left*) and root mean square errors (*right*) of $H_p(k)$ for an *extreme value* d.f. with $\gamma = 0.1$.

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 RMSEs, but also valid for mean values at optimal levels, in the sense of minimal RMSE, as we shall see next, in Section 5.1.1.

5.1.1 Mean values of the MOP EVI-estimators at optimal levels

Table 1 is again related with the *extreme value* model. We shall there present, for $n = 100,\ 200,\ 500,\ 1000,\ 2000$ and 5000, the simulated mean values at optimal levels (levels

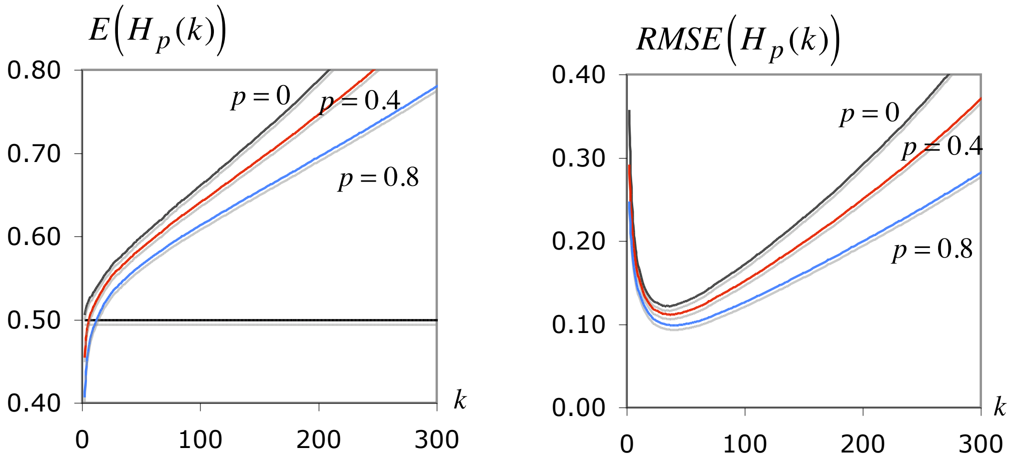

Figure 5: Mean values (*left*) and root mean square errors (*right*) of $H_p(k)$ for an *extreme value* d.f. with $\gamma = 0.5$.

where RMSEs are minima as functions of k) of the EVI-estimators $H_p(k)$, in (1.11), for $p = j/(10\gamma)$, $j = 0, 1, 2, 3, 4$. We shall thus consider the estimators $H_{p0} = H_p(k_{0|p}(n))$, $k_{0|p}(n) := \arg\min_k MSE(H_p(k))$. Information on 95% confidence intervals, computed on the basis of the 20 replicates with 5000 runs each, is also provided. Among the estimators considered, the one providing the smallest squared bias is underlined and in **bold**.

Remark 7. *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 extreme value model, but for all simulated models.*
- *The above mentioned remark means that, regarding bias, we can safely take* $p = 4/(10\gamma)$*. However, we have to pay attention to variance, which increases with* p*.*

5.1.2 Mean square errors and relative efficiency indicators at optimal levels

We shall compute the Hill estimator, in (1.7), as well as in (1.11) whenever $p = 0$, at the simulated value of $k_{0|0} := \arg\min_k MSE\big(H_0(k)\big)$, the simulated optimal k in the sense of minimum RMSE, not relevant in practice, but providing an indication of the best possible performance of Hill's estimator. Such an estimator will be denoted H_{00}. For the

Table 1: Simulated mean values, at optimal levels, of the estimators $H_p(k)$, $p = j/(10\gamma)$, $j = 0, 1, 2, 3, 4$, for *extreme value* underlying parents, together with 95% confidence intervals.

EV_γ **parent,** $\gamma = 0.1$						
n	100	200	500	1000	2000	5000
$p = 0$	0.334 ± 0.0009	0.284 ± 0.0007	0.243 ± 0.0005	0.223 ± 0.0015	0.209 ± 0.0014	0.195 ± 0.0011
$p = 1$	0.299 ± 0.0007	0.260 ± 0.0006	0.226 ± 0.0005	0.208 ± 0.0005	0.194 ± 0.0009	0.181 ± 0.0013
$p = 2$	0.263 ± 0.0005	0.234 ± 0.0005	0.208 ± 0.0004	0.193 ± 0.0004	0.180 ± 0.0004	0.167 ± 0.0004
$p = 3$	0.228 ± 0.0003	0.208 ± 0.0003	0.188 ± 0.0003	0.176 ± 0.0004	0.166 ± 0.0003	0.156 ± 0.0003
$p = 4$	**0.197** ± 0.0002	**0.183** ± 0.0002	**0.169** ± 0.0002	**0.160** ± 0.0003	**0.152** ± 0.0002	**0.144** ± 0.0003
EV_γ **parent,** $\gamma = 0.25$						
$p = 0$	0.427 ± 0.0012	0.391 ± 0.0026	0.365 ± 0.0019	0.348 ± 0.0012	0.335 ± 0.0013	0.321 ± 0.0010
$p = 0.4$	0.406 ± 0.0011	0.372 ± 0.0029	0.352 ± 0.0016	0.340 ± 0.0012	0.328 ± 0.0011	0.317 ± 0.0009
$p = 0.8$	0.384 ± 0.0009	0.350 ± 0.0026	0.338 ± 0.0011	0.330 ± 0.0013	0.321 ± 0.0008	0.312 ± 0.0009
$p = 1.2$	0.360 ± 0.0008	0.328 ± 0.0007	0.322 ± 0.0019	0.315 ± 0.0017	0.312 ± 0.0008	0.304 ± 0.0010
$p = 1.6$	**0.336** ± 0.0007	**0.309** ± 0.0006	**0.303** ± 0.0013	**0.301** ± 0.0013	**0.300** ± 0.0011	**0.294** ± 0.0008
EV_γ **parent,** $\gamma = 0.5$						
$p = 0$	0.654 ± 0.0032	0.624 ± 0.0033	0.596 ± 0.0011	0.579 ± 0.0016	0.565 ± 0.0010	0.551 ± 0.0010
$p = 0.2$	0.637 ± 0.0030	0.613 ± 0.0031	0.589 ± 0.0014	0.574 ± 0.0017	0.561 ± 0.0011	0.549 ± 0.0007
$p = 0.4$	0.618 ± 0.0030	0.602 ± 0.0019	0.582 ± 0.0016	0.570 ± 0.0013	0.559 ± 0.0012	0.547 ± 0.0009
$p = 0.6$	0.597 ± 0.0031	0.588 ± 0.0021	0.573 ± 0.0013	0.564 ± 0.0015	0.557 ± 0.0010	0.546 ± 0.0006
$p = 0.8$	**0.577** ± 0.0022	**0.570** ± 0.0016	**0.562** ± 0.0015	**0.556** ± 0.0014	**0.550** ± 0.0008	**0.541** ± 0.0007

EVI-estimators $H_p(k)$, we shall compute H_{p0}, the estimator H_p computed at its simulated optimal level, again in the sense of minimum MSE, i.e. at the simulated value of $k_{0|p} := \arg\min_k MSE\big(H_p(k)\big)$. The simulated indicators are

$$REFF_{p|0} := \frac{RMSE\,(H_{00})}{RMSE\,(H_{p0})} = \sqrt{\frac{MSE\,(H_{00})}{MSE\,(H_{p0})}}. \tag{5.1}$$

Remark 8. *An indicator higher than one means a better performance than the Hill estimator. Consequently, the higher these indicators are, the better the H_{p0}-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 RMSEs of all other estimators

H_{p0}. The following rows provide the $REFF$ indicators, $REFF_{p|0}$ in (5.1), for the different EVI-estimators under study. Again, the estimator providing the highest $REFF$ indicator (minimum RMSE at optimal level) is underlined and in **bold**.

Table 2: Simulated root mean square errors of H (first row) and $REFF$-indicators of $H_p(k)$, $p = j/(10\gamma)$, $j = 1, 2, 3, 4$, for *extreme value* underlying parents, together with 95% confidence intervals.

EV_γ **parent,** $\gamma = 0.1$						
n	100	200	500	1000	2000	5000
$RMSE_H$	0.268 ± 0.0229	0.216 ± 0.0185	0.174 ± 0.0141	0.151 ± 0.0136	0.133 ± 0.0127	0.113 ± 0.0108
$p = 1$	1.195 ± 0.0012	1.166 ± 0.0008	1.143 ± 0.0011	1.131 ± 0.0015	1.114 ± 0.0024	1.093 ± 0.0023
$p = 2$	1.491 ± 0.0024	1.416 ± 0.0015	1.356 ± 0.0020	1.325 ± 0.0022	1.292 ± 0.0034	1.251 ± 0.0040
$p = 3$	1.934 ± 0.0040	1.792 ± 0.0028	1.677 ± 0.0030	1.618 ± 0.0031	1.562 ± 0.0044	1.495 ± 0.0050
$p = 4$	**2.581** ± 0.0062	**2.343** ± 0.0047	**2.151** ± 0.0044	**2.052** ± 0.0045	**1.963** ± 0.0059	**1.856** ± 0.0062
EV_γ **parent,** $\gamma = 0.25$						
$RMSE_H$	0.246 ± 0.0253	0.200 ± 0.0199	0.157 ± 0.0153	0.133 ± 0.0132	0.113 ± 0.0100	0.092 ± 0.0081
$p = 0.4$	1.124 ± 0.0011	1.102 ± 0.0019	1.076 ± 0.0012	1.063 ± 0.0012	1.053 ± 0.0008	1.044 ± 0.0007
$p = 0.8$	1.292 ± 0.0021	1.245 ± 0.0043	1.180 ± 0.0024	1.147 ± 0.0029	1.121 ± 0.0021	1.099 ± 0.0016
$p = 1.2$	1.532 ± 0.0033	1.451 ± 0.0056	1.334 ± 0.0041	1.271 ± 0.0048	1.220 ± 0.0036	1.176 ± 0.0028
$p = 1.6$	**1.879** ± 0.0048	**1.744** ± 0.0067	**1.565** ± 0.0055	**1.463** ± 0.0066	**1.379** ± 0.0052	**1.300** ± 0.0040
EV_γ **parent,** $\gamma = 0.5$						
$RMSE_H$	0.256 ± 0.0263	0.202 ± 0.0229	0.151 ± 0.0140	0.122 ± 0.0109	0.100 ± 0.0090	0.077 ± 0.0068
$p = 0.2$	1.081 ± 0.0010	1.064 ± 0.0010	1.048 ± 0.0009	1.040 ± 0.0009	1.035 ± 0.0010	1.030 ± 0.0008
$p = 0.4$	1.185 ± 0.0021	1.142 ± 0.0023	1.104 ± 0.0020	1.085 ± 0.0022	1.071 ± 0.0020	1.059 ± 0.0021
$p = 0.6$	1.326 ± 0.0032	1.250 ± 0.0037	1.177 ± 0.0034	1.139 ± 0.0037	1.111 ± 0.0035	1.088 ± 0.0034
$p = 0.8$	**1.530** ± 0.0044	**1.408** ± 0.0050	**1.287** ± 0.0049	**1.221** ± 0.0054	**1.169** ± 0.0052	**1.123** ± 0.0048

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

- *Just as for mean values at optimal levels, the best results were obtained for* $p = 4/(10\gamma)$ *for all simulated models but the* Fréchet.
- *For* Fréchet *underlying parents, the REFF-indicator* $REFF_{p|0}$*, provided in Table 3 for* $p = j/\gamma$*, does not depend of* γ.

Table 3: Simulated RMSEs of H (first row) and REFF-indicators of $H_p(k)$ (independent on γ), for $p = j/(10\gamma)$, $j = 1, 2, 3, 4$, for *Fréchet* parents, together with 95% confidence intervals.

Fréchet parent, γ						
n	100	200	500	1000	2000	5000
$RMSE_H$	0.021 ± 0.0021	0.016 ± 0.0017	0.012 ± 0.0010	0.009 ± 0.0008	0.007 ± 0.0007	0.005 ± 0.0005
$j = 1$	1.038 ± 0.0012	1.031 ± 0.0011	1.026 ± 0.0010	1.023 ± 0.0009	1.020 ± 0.0010	1.019 ± 0.0010
$j = 2$	1.077 ± 0.0027	1.059 ± 0.0028	1.045 ± 0.0020	1.039 ± 0.0020	$\underline{\mathbf{1.032}} \pm 0.0024$	$\underline{\mathbf{1.030}} \pm 0.0019$
$j = 3$	1.120 ± 0.0046	1.084 ± 0.0057	1.055 ± 0.0032	$\underline{\mathbf{1.041}} \pm 0.0034$	1.028 ± 0.0040	1.022 ± 0.0034
$j = 4$	$\underline{\mathbf{1.185}} \pm 0.0055$	$\underline{\mathbf{1.120}} \pm 0.0081$	$\underline{\mathbf{1.060}} \pm 0.0047$	1.027 ± 0.0055	0.999 ± 0.0069	0.981 ± 0.0053

6 Case-studies

We shall now consider an application of Algorithm 4.1 to

(1) two randomly simulated samples, with size $n = 500$, from a Fréchet parent with $\gamma = 0.25$, denoted FRE_1 and FRE_2;

(2) two randomly simulated samples, with size $n = 1000$, from a Student t_ν parent with $\nu = 4 \quad (\gamma = 1/\nu = 0.25)$, denoted 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.* (2012b), the daily log-returns of NASDAQ index from 1997 to 2000, which corresponds to a sample size $n = 1037$;

(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 SECURA;

(5) a sample, of size $n = 2627$, denoted FIRES, already considered in Gomes *et al.* (2012a) 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 4.1, and the bootstrap CIs, in* Remark 5, *we shall consider, for the Hill estimator, the most common estimate of* $k_{0|H}(n) := \arg\min_k MSE(H_n(k))$ (Hall, 1982), *given by*

$$\hat{k}_{0|H}(n) = \left[\left((1-\hat{\rho})^2 n^{-2\hat{\rho}} / \left(-2\ \hat{\rho}\ \hat{\beta}^2 \right) \right)^{1/(1-2\hat{\rho})} \right] \tag{6.1}$$

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

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

For all data sets under analysis, the sample paths of the ρ-estimates associated with $\tau = 0$ and $\tau = 1$ lead us to choose the estimates associated with $\tau = 0$, on the basis of any stability criterion for large k, including the one in **Step 2** of Algorithm 4.1. In Table 4, Table 5 and Table 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 (3.2), we provide the sub-sample size choice n_1^*, in **Step 14**, the value a^* and associated p^*_{min}, in **Step 15**, the bootstrap threshold estimates, $\hat{k}^*_{0|0}$ and $\hat{k}^{**}_0$, obtained in **Step 9** and **Step 16**, respectively, and the estimates $\hat{k}_{0|H}$, in (6.1).

Data	n	n_0	n_1^*	a^*	p^*_{min}	$(\hat{\beta}_0, \hat{\rho}_0)$	$\hat{k}^*_{0\vert 0}$	$\hat{k}^{**}_0$	$\hat{k}_{0\vert H}$
FRE_1	500	500	409	5	1.493	(0.89, −1.02)	101	101	87
FRE_2	500	500	366	6	0.544	(0.87, −1.50)	201	176	136
STU_1	1000	496	483	3	0.435	(1.02, −0.72)	37	40	51
STU_2	1000	489	479	2	0.528	(1.03, −0.67)	10	14	46
SECURA	371	371	266	6	1.010	(0.81, −0.74)	53	61	53
NASDAQ	1036	570	453	3	0.397	(1.02, −0.73)	48	48	57
FIRES	2627	2627	1917	2	0.136	(0.48, −0.39)	57	71	119

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

In Table 5, we provide the adaptive Hill-estimates, $H := H(\hat{k}_{0|H})$, as well as 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 4.1. Close to those estimates, and between parenthesis, we place the associated approximate 99% CIs.

Data	$H := H(\hat{k}_0^H)$	$H_0^* := H_{0,n,n_1^*,\|T}$	$H^{**} \equiv H_{p_{min}^*} := H_{p_{min}^*,n,n_1^*\|T}$
FRE_1	0.270 (0.2004, 0.3391)	0.268 (0.1762, 0.3135)	0.266 (0.1777, 0.3158)
FRE_2	0.283 (0.2229, 0.3417)	0.276 (0.2012, 0.3015)	0.265 (0.1915, 0.3062)
STU_1	0.358 (0.2501, 0.5044)	0.359 (0.1870, 0.4913)	0.285 (0.1266, 0.4065)
STU_2	0.311 (0.2131, 0.4449)	0.242 (0.0381, 0.4323)	0.267 (0.1311, 0.3900)
SECURA	0.297 (0.2028, 0.3924)	0.297 (0.1591, 0.3692)	0.282 (0.1565, 0.3508)
NASDAQ	0.400 (0.2831, 0.5482)	0.378 (0.2139, 0.4947)	0.372 (0.2101, 0.4909)
FIRES	0.719 (0.5369, 0.8292)	0.738 (0.4288, 0.9321)	0.725 (0.4428, 0.8889)

Table 5: Adaptive EVI-estimates and associated 99% CIs obtained through Hill estimators at estimated optimal level (H), and bootstrap adaptive estimates ($H_0^*, H_{p_{min}^*}^*$) in Algorithm 4.1, for the different data sets under analysis.

In Table 6, we provide information on the bootstrap estimates of bias and RMSE of the adaptive estimates obtained by Algorithm 4.1, the sizes of the CIs in Table 5, respectively denoted 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 Figure 6, where we represent the boostrap RMSE estimates associated with $a = 6$, for the samples FRE_2 and SECURA.

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	0.0230	**0.0190**	0.0352	**0.0328**	0.1387	**0.1373**	*0.1381*	0.245	**0.247**
FRE_2	0.0245	**0.0157**	0.0313	**0.0272**	0.1188	**0.1003**	*0.1147*	0.251	**0.249**
STU_1	0.0201	**0.0187**	0.0624	**0.0575**	**0.2543**	0.3043	*0.2799*	0.339	**0.267**
STU_2	0.0068	**0.0063**	0.0768	**0.0506**	**0.2318**	0.3942	*0.2589*	0.235	**0.261**
SECURA	0.0328	**0.0287**	0.0523	**0.0474**	**0.1896**	0.2101	*0.1943*	0.264	0.254
NASDAQ	0.0233	**0.0214**	0.0593	**0.0586**	**0.2651**	0.2808	*0.2808*	0.354	0.351
FIRES	**0.0572**	0.0594	0.1132	**0.1050**	**0.2923**	0.5033	*0.4461*	0.680	0.666

Table 6: Adaptive bootstrap estimates of bias and RMSE of the adaptive estimates obtained through Algorithm 4.1, sizes of the CIs in Table 5, and corrected-bias bootstrap adaptive EVI-estimates, for the different data sets under analysis.

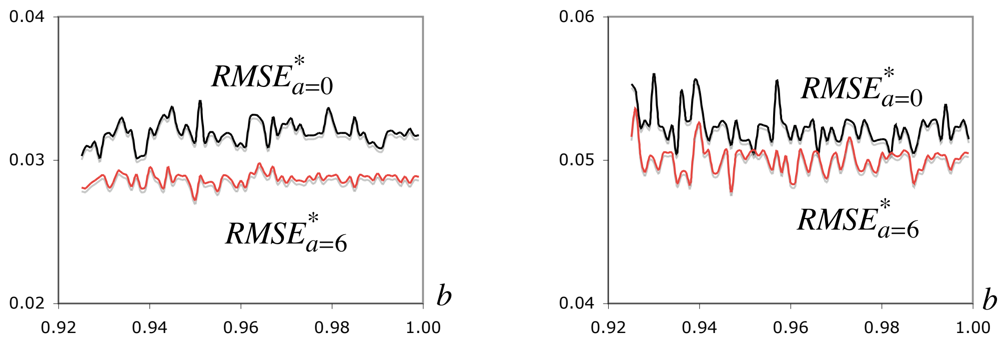

Figure 6: Bootstrap RMSE estimates for FRE_2 (*left*) and SECURA (*right*), as function of b ($n_1 = n^b$).

6.1 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 4.1. This is however a topic 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.

Acknowledgements. Research partially supported by National Funds through **FCT** – Fundação para a Ciência e a Tecnologia, project PEst-OE/MAT/UI0006/2011, and PTDC/FEDER, EXTREMA.

References

[1] Araújo Santos, P., Fraga Alves, M.I. and Gomes, M.I. (2006). Peaks over random threshold methodology for tail index and quantile estimation. *Revstat* **4**:3, 227–247.

[2] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes. Theory and Applications.* Wiley.

[3] Beirlant, J., Figueiredo, F., Gomes, M.I. and Vandewalle, B. (2008). Improved reduced-bias tail index and quantile estimators. *J. Statist. Plann. and Inference* **138**:6, 1851-1870.

[4] Bingham, N., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation.* Cambridge Univ. Press, Cambridge.

[5] Caeiro, C. and Gomes, M.I. (2006). A new class of estimators of a scale second order parameter. *Extremes*, **9**, 193–211.

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

[7] Danielsson, J., de Haan, L., Peng, L. and de Vries, C.G. (2001). Using a bootstrap method to choose the sample fraction in the tail index estimation. *J. Multivariate Analysis* **76**, 226-248.

[8] Dekkers, A.L.M. and de Haan, L. (1993). Optimal choice of sample fraction in extreme-value estimation. *J. of Multivariate Analysis* **47**, 173-195.

[9] Draisma, G., de Haan, L., Peng, L. and Pereira, T. (1999). A bootstrap-based method to achieve optimality in estimating the extreme value index. *Extremes*, **2** (4): 367-404.

[10] Drees, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli* **9**:4, 617–657.

[11] Fraga Alves, M.I., Gomes, M.I. and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica* **60**:1, 193-213.

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

[13] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics* **44**:6, 423–453.

[14] Goegebeur, Y., Beirlant, J. and 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, 5169.

[15] Goegebeur, Y., Beirlant, J. and de Wet, T. (2010). Kernel estimators for the second order parameter in extreme value statistics. *J. Statistical Planning and Inference* **140**:9, 2632-2654.

[16] Gomes, M.I. and Henriques-Rodrigues, L. (2010). Comparison at optimal levels of classical tail index estimators: a challenge for reduced-bias estimation? *Discussiones Mathematica: Probability and Statistics* **30**:1, 35–51.

[17] Gomes, M.I. and Martins, M.J. (2001). Generalizations of the Hill estimator — asymptotic versus finite sample behaviour. *J. Statistical Planning and Inference* **93**, 161-180.

[18] Gomes, M.I. and 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.

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

[20] Gomes, M.I. and Oliveira, O. (2001). The bootstrap methodology in Statistics of Extremes: choice of the optimal sample fraction. *Extremes*, **4**:4, 331–358.

[21] Gomes, M.I. and Pestana, D. (2007). A simple second order reduced-bias extreme value index estimator. *J. Statistical Computation and Simulation* **77**:6, 487-504.

[22] Gomes, M.I., Miranda, C. and Pereira, H. (2005). Revisiting the role of the Jackknife methodology in the estimation of a positive extreme value index. *Communications in Statistics — Theory and Methods* **34**, 1-20.

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

[24] Gomes, M.I., Henriques-Rodrigues, L., Pereira, H. and Pestana, D. (2010). Tail index and second order parameters' semi-parametric estimation based on the log-excesses, *J. Statist. Comput. and Simul.* **80**, 6, 653–666.

[25] Gomes, M.I., Martins, M.J. and Neves, M.M. (2011a). *Revisiting the Role of the Generalized JackKnife Methodology in the Field of Extremes.* Notas e Comunicações CEAUL 20/2012.

[26] Gomes, M.I., Mendonça, S. and Pestana, D.D. (2011b). Adaptive reduced-bias tail index and VaR estimation via the bootstrap methodology, *Comm. in Statistics – Theory and Methods* **40**(16), 2946-2968.

[27] Gomes, M.I., Figueiredo, F. and Neves, M.M. (2012a). Adaptive estimation of heavy right tails: the bootstrap methodology in action, *Extremes*, in press. DOI: 10.1007/s10687-011-0146-6

[28] Gomes, M.I., Henriques-Rodrigues, L., Fraga Alves, M.I. and Manjunath, B.G. (2012b). Adaptive PORT-MVRB Estimation: an Empirical Comparison of two Heuristic Algorithms. *J. Statist. Comput. and Simul.* DOI:10.1080/00949655.2011.652113

[29] Haan, L. de (1970). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*. Mathematical Centre Tract 32, Amsterdam.

[30] Haan, L. de and Peng, L. (1998). Comparison of extreme value index estimators. *Statistica Neerlandica* **52**, 60-70.

[31] Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. Royal Statistical Society* **B44**, 37-42.

[32] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Annals Statistics* **3**, 1163-1174.

[33] Vandewalle, B. and Beirlant, J. (2006). On univariate extreme value statistics and the estimation of reinsurance premiums. *Insurance Mathematics and Economics* **38**:3, 444-459.