A note on the asymptotic variance at optimal levels of a bias-corrected Hill estimator

M. Ivette Gomes, Dinis Pestana, Frederico Caeiro

Accepted Manuscript

A note on the asymptotic variance at optimal levels of a bias-corrected Hill estimator

M. Ivette Gomes, Dinis Pestana, Frederico Caeiro

PII: S0167-7152(08)00401-X
DOI: 10.1016/j.spl.2008.08.016
Reference: STAPRO 5192

To appear in: *Statistics and Probability Letters*

Received date: 10 December 2007
Revised date: 14 May 2008
Accepted date: 25 August 2008

Please cite this article as: Gomes, M.I., Pestana, D., Caeiro, F., A note on the asymptotic variance at optimal levels of a bias-corrected Hill estimator. *Statistics and Probability Letters* (2008), doi:10.1016/j.spl.2008.08.016

A note on the asymptotic variance at optimal levels of a bias-corrected Hill estimator†

M. Ivette Gomes
CEAUL, DEIO, Faculty of Science
University of Lisbon, Portugal

Dinis Pestana
CEAUL, DEIO, Faculty of Science
University of Lisbon, Portugal

Frederico Caeiro
Department of Mathematics and CMA
New University of Lisbon, Portugal

For heavy tails, with a positive tail index γ, classical tail index estimators, like the Hill estimator, are known to be quite sensitive to the number of top order statistics k used in the estimation, whereas second-order reduced-bias estimators show much less sensitivity to changes in k. In the recent minimum-variance reduced-bias (MVRB) tail index estimators, the estimation of the second order parameters in the bias has been performed at a level k_1 of a larger order than that of the level k at which we compute the tail index estimators. Such a procedure enables us to keep the asymptotic variance of the new estimators equal to the asymptotic variance of the Hill estimator, for all k at which we can guarantee the asymptotic normality of the Hill statistics. These values of k, as well as larger values of k, will also enable us to guarantee the asymptotic normality of the reduced-bias estimators, but, to reach the minimal mean squared error of these MVRB estimators, we need to work with levels k and k_1 of the same order. In this note we derive the way the asymptotic variance varies as a function of q, the finite limiting value of k/k_1, as the sample size n increases to infinity.

Keywords and phrases: Statistics of extremes; reduced-bias semi-parametric estimation; tail index; asymptotic theory.

†Research partially supported by FCT / POCTI, POCI and PPCDT / FEDER.

1 Introduction

In *statistics of extremes*, and whenever we are interested in large values, a model F is said to be *heavy-tailed* if the right *tail function*, $\overline{F} := 1 - F$, is a regularly varying function with a negative index of regular variation equal to $-1/\gamma$, $\gamma > 0$. We then use the notation $\overline{F} \in RV_{-1/\gamma}$, where, for any real a, RV_a stands for the class of *regularly varying* functions at infinity with an *index of regular variation* equal to a, i.e., positive measurable functions g such that $\lim_{t\to\infty} g(tx)/g(t) = x^a$, for all $x > 0$. Equivalently, the quantile function $U(t) = F^{\leftarrow}(1 - 1/t)$, $t \geq 1$, with $F^{\leftarrow}(x) = \inf\{y : F(y) \geq x\}$, is of regular variation with index γ, i.e.,

$$F \text{ is heavy-tailed} \quad \Longleftrightarrow \quad \overline{F} = 1 - F \in RV_{-1/\gamma} \quad \Longleftrightarrow \quad U \in RV_\gamma \tag{1}$$

for some $\gamma > 0$ (Gnedenko, 1943; de Haan, 1970, 1984). Then, we are in the domain of attraction for maxima of an *extreme value* distribution function (d.f.), $EV_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma})$, $x \geq -1/\gamma$, and we write $F \in \mathcal{D}_{\mathcal{M}}(EV_{\gamma>0})$. The parameter γ is the *tail index*, the primary parameter of extreme events.

The *second order parameter*, ρ (≤ 0), rules the rate of convergence in the first order condition in (1), and it is the parameter appearing in

$$\lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho}, \tag{2}$$

which holds for every $x > 0$, and where $|A|$ must be in RV_ρ (Geluk and de Haan, 1987). We shall moreover assume that $\rho < 0$. This condition has been widely accepted as an appropriate condition to specify the tail of a Pareto-type distribution in a semi-parametric way.

In order to obtain information on the order of the asymptotic bias of any second-order reduced-bias tail index estimator, we need further assuming a third order condition, ruling now the rate of convergence in the second order condition in (2), and which guarantees that

$$\lim_{t\to\infty} \Big(\frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} - \frac{x^\rho - 1}{\rho}\Big)/B(t) = \frac{x^{\rho+\rho'} - 1}{\rho + \rho'} \tag{3}$$

for all $x > 0$, and where $|B|$ must be in $RV_{\rho'}$. There appears then a third order parameter $\rho' \leq 0$, which we also assume to be negative. Such a condition has already been used in Gomes *et al.* (2002) and Fraga Alves *et al.* (2003), for the full derivation of the asymptotic behaviour of ρ-estimators and in Gomes *et al.* (2004; 2007; 2008a) and Caeiro *et al.* (2005), for the study of specific reduced-bias tail index estimators. More restrictively, and for some details in the paper, we shall assume that we can choose in (3),

$$A(t) = c\ t^{\rho} =: \gamma\ \beta\ t^{\rho}, \qquad B(t) = c'\ t^{\rho'} =: \beta'\ t^{\rho'}, \quad \beta,\ \beta' \neq 0 \quad \rho,\ \rho' < 0. \tag{4}$$

Condition (3) (and *a fortiori* condition (2)), as well as (4), hold for most common Pareto-type distributions, like the extreme value, the Fréchet, the Generalized Pareto and the Student's t.

For *intermediate* k, i.e., a sequence of integers $k = k_n$, $k \in [1, n)$, such that

$$k = k_n \to \infty \quad \text{and} \quad k_n = o(n), \quad \text{as} \quad n \to \infty, \tag{5}$$

we shall consider, as basic statistics, the *log-excesses* over a random high level, i.e.,

$$V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n}, \quad 1 \le i \le k < n, \tag{6}$$

and the *scaled log-spacings*,

$$W_i := i\,\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\}\,, \quad 1 \le i \le k < n, \tag{7}$$

where $X_{i:n}$ denotes, as usual, the i-th ascending order statistic (o.s.), $1 \le i \le n$, associated to an independent, identically distributed (i.i.d.) random sample $(X_1, X_2, \cdots, X_n)$. It is well known that if (5) holds, and under the first order framework in (1), the log-excesses, V_{ik}, $1 \le i \le k$, in (6), are approximately the k o.s.'s from an exponential sample of size k and mean value γ. Also, under the same conditions, the scaled log-spacings, W_i, $1 \le i \le k$, in (7), are approximately i.i.d. and exponential with mean value γ. Consequently, the Hill estimator of γ (Hill, 1975),

$$H(k) \equiv H_n(k) = \frac{1}{k}\sum_{i=1}^{k} V_{ik} = \frac{1}{k}\sum_{i=1}^{k} W_i, \tag{8}$$

is consistent for the estimation of γ under the first order framework and for intermediate k. Note that the Hill estimator in (8) is the maximum likelihood estimator of the tail index γ, under a strict Pareto model, with d.f. $F_P(x) = 1 - x^{-1/\gamma},\ x \ge 1$. Under the second order framework in (2) and for intermediate k, i.e., if (5) holds, the asymptotic distributional representation

$$H_n(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\ Z_k^{(1)} + \frac{A(n/k)}{1-\rho} + o_p(A(n/k)) \tag{9}$$

holds, where $Z_k^{(1)} = \sqrt{k}\big(\sum_{i=1}^{k} E_i/k - 1\big)$, with $\{E_i\}$ i.i.d. standard exponential random variables (r.v.'s), is asymptotically standard normal (de Haan and Peng, 1998). How to choose k in some optimal sense as a function of the sample size n is thus an important problem. Under condition (2) and with $A(t)$ chosen as in (4), the optimal k for the estimation of γ through the Hill estimator in (8), in the sense of minimal asymptotic mean squared error, is given by

$$k_0^H \equiv k_0^H(n;\beta,\rho) = \Big(\frac{(1-\rho)n^{-\rho}}{\beta\sqrt{-2\rho}}\Big)^{2/(1-2\rho)}. \tag{10}$$

We thus need to estimate β and ρ, in order to estimate k_0^H. But the Hill estimator is quite sensitive to k, with a large bias for moderate k. Hence the need for bias reduction techniques.

The most simple minimum-variance reduced-bias (MVRB) estimators in the literature are the bias-corrected Hill estimators in Caeiro *et al.* (2005), with the functional form,

$$\overline{H}_{\hat\beta,\hat\rho}(k) := H(k)\Big(1 - \hat\beta\left(\frac{n}{k}\right)^{\hat\rho}/(1-\hat\rho)\Big), \tag{11}$$

dependent upon the Hill estimator $H(k)$ and $(\hat\beta, \hat\rho)$, adequate consistent estimators of the second order parameters β and ρ, respectively. To achieve consistency of $(\hat\beta, \hat\rho)$ we need to use a number k_1 of top o.s.'s such that $\sqrt{k_1}A(n/k_1) \to \infty$ (see Section 3). If we further have $\hat\rho - \rho = o_p(1/\ln n)$, $\sqrt{k}(\overline{H}_{\hat\beta,\hat\rho}(k) - \gamma)$ is asymptotically normal with mean value equal to zero and variance γ^2 at least for intermediate values k such that $\sqrt{k}A(n/k) \to \lambda$, finite (Caeiro *et al.*, 2005), i.e. if $k = o(k_1)$. Further information on the order of the asymptotic bias of the reduced-bias estimator in (11), for a slightly more restrictive class of models than the one in (3), is provided in Caeiro and Gomes (2008a). For the use of this estimator in quantile estimation, see Gomes and Pestana (2007) and Caeiro and Gomes (2008b).

In Section 2 of this paper, we shall provide details on conditions under which we are able to keep the asymptotic variance of the MVRB-estimators in (11) equal to γ^2, together with their behaviour under a third order framework, working essentially with values $k = o(k_1)$. In Section 3, we shall briefly review the estimation of the second order parameters β and ρ. Next, in Section 4, we provide some information on the asymptotic behaviour of $\sqrt{k}\{\overline{H}_{\hat\beta,\hat\rho}(k) - \gamma\}$, whenever $\sqrt{k}A(n/k) \to \infty$, $\sqrt{k}\ A^2(n/k) \to \lambda_A$ and $\sqrt{k}\ A(n/k)B(n/k) \to \lambda_B$, both finite, λ_A or $\lambda_B \neq 0$, and when we consider the estimators of ρ in Section 3, computed at any "optimal" level $k_1 = k_1^{opt}$, in the sense of a level such that $\sqrt{k_1}\ A^2(n/k_1) \to \lambda_{A_1}$ and $\sqrt{k_1}\ A(n/k_1)B(n/k_1) \to \lambda_{B_1}$, both finite, λ_{A_1} or $\lambda_{B_1} \neq 0$. For the class of models under consideration, i.e., models for which (3) holds, with A and B chosen as in (4) for arbitrary $\rho,\ \rho' < 0$, we thus have $k/k_1 \to q > 0$, whenever $n \to \infty$. Finally, in Section 5, we advance with a small-scale simulation study to illustrate possible adaptive choices of k and k_1.

2 Asymptotic behaviour of the MVRB-estimators under a third order framework

In a trial to keep the asymptotic variance of the reduced-bias estimators in (11) equal to γ^2, we can state the following theorem, a generalization of Theorem 3.1 in Caeiro *et al.* (2005), where $U_n \overset{p}{\sim} V_n$ denotes that U_n/V_n converges in probability towards one, as $n \to \infty$.

Theorem 1. *Under the third order framework in* (3)*, with A and B chosen as in* (4)*, if* (5) *holds, with $Z_k^{(1)}$ the asymptotically standard normal r.v. in* (9) *and $\overline{H}_{\beta,\rho}(k)$ equal to the quantity $\overline{H}_{\hat\beta,\hat\rho}(k)$ given in* (11)*, with $\hat\beta$ and $\hat\rho$ replaced by β and ρ, respectively,*

$$\overline{H}_{\beta,\rho}(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} - A(n/k)\Big(\frac{A(n/k)}{\gamma(1-\rho)^2} - \frac{B(n/k)}{1-\rho-\rho'} + O_p\Big(\frac{1}{\sqrt{k}}\Big)\Big)(1+o_p(1)). \qquad (12)$$

Consequently, and as $n \to \infty$, if

$$\sqrt{k}\ A(n/k) \to \infty, \text{ with } \sqrt{k}\ A^2(n/k) \to \lambda_A \text{ and } \sqrt{k}\ A(n/k)\ B(n/k) \to \lambda_B, \qquad (13)$$

λ_A and λ_B finite, $\sqrt{k}\,\big(\overline{H}_{\beta,\rho}(k) - \gamma\big)$ converges in distribution to a Normal$\big(b_{\overline{H}}, \gamma^2\big)$ r.v., with

$$b_{\overline{H}} \equiv b_{\overline{H}}(\gamma,\rho,\rho') \equiv ABIAS_{\overline{H}} = -\lambda_A/(\gamma(1-\rho)^2) + \lambda_B/(1-\rho-\rho') =: \lambda_A u_{\overline{H}} + \lambda_B v_{\overline{H}}. \qquad (14)$$

Let $(\hat\beta,\hat\rho)$ be any consistent estimator of the vector of second order parameters (β,ρ) such that

$$\hat\rho - \rho = o_p(1/\ln n), \text{ as } n \to \infty. \qquad (15)$$

Then, with $a_{\overline{H}} = -1/(1-\rho)$,

$$\overline{H}_{\hat\beta,\hat\rho}(k) - \overline{H}_{\beta,\rho}(k) \overset{p}{\sim} a_{\overline{H}} A(n/k)\Big\{\big(\hat\beta - \beta\big)/\beta + (\hat\rho - \rho)\left[\ln(n/k) - a_{\overline{H}}\right]\Big\}. \qquad (16)$$

Consequently, $\sqrt{k}\{\overline{H}_{\hat\beta,\hat\rho}(k) - \gamma\}$ is asymptotically normal with null mean value and variance $\sigma_0^2 = \gamma^2$, not only when $\sqrt{k}\ A(n/k) \to 0$, but also whenever $\sqrt{k}\ A(n/k) \to \lambda$, finite. This same result still holds for levels k such that $\sqrt{k}\ A(n/k) \to \infty$, provided that $\sqrt{k}\ A^2(n/k) \to 0$, $\sqrt{k}\ A(n/k)B(n/k) \to 0$, and $\hat\beta - \beta$ as well as $(\hat\rho - \rho)\ln n$ are $o_p\big(1/\sqrt{k}\ A(n/k)\big)$.

Proof. Under the conditions in the theorem, directly from (3) and the same line of reasoning of de Haan and Peng (1998), the asymptotic distributional representation

$$H_n(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + \left(\frac{A(n/k)}{1-\rho} + \frac{A(n/k)\ B(n/k)}{1-\rho-\rho'} + O_p\Big(\frac{A(n/k)}{\sqrt{k}}\Big)\right)(1+o_p(1))$$

follows. On the other hand, as $\overline{H}_{\beta,\rho}(k) = H_n(k)(1 - A(n/k)/(\gamma(1-\rho)))$, we get (12), and consequently, the stated asymptotic normality of $\sqrt{k}\big(\overline{H}_{\beta,\rho}(k) - \gamma\big)$.

Next, and directly from the expression of $\overline{H}_{\beta,\rho}(k)$, we get

$$\frac{\partial \overline{H}_{\beta,\rho}}{\partial \beta} \overset{p}{\sim} -\frac{A(n/k)}{\beta(1-\rho)}, \qquad \frac{\partial \overline{H}_{\beta,\rho}}{\partial \rho} \overset{p}{\sim} -\frac{A(n/k)}{1-\rho}\Big(\ln(n/k) + \frac{1}{1-\rho}\Big).$$

The use of Cramer's delta-method, together with the validity of (15), enables us to write

$$\overline{H}_{\hat\beta,\hat\rho}(k) = \overline{H}_{\beta,\rho}(k) - \frac{A(n/k)}{1-\rho}\Big\{\frac{\hat\beta - \beta}{\beta} + (\hat\rho - \rho)\Big(\ln(n/k) + \frac{1}{1-\rho}\Big)\Big\}(1 + o_p(1)).$$

Consequently, (16) follows, as well as the remaining of the theorem, provided that we pay attention to the validity of (12). □

Remark 1. *Note that the optimal level, in the sense of minimal asymptotic mean squared error, for the estimation of the tail index γ through the reduced-bias estimator $\overline{H}_{\beta,\rho}(k)$ (assuming thus that β and ρ are known) is such that* (13) *holds. An important question to answer is how far can this same result be true for the reduced-bias estimator $\overline{H}_{\hat\beta,\hat\rho}(k)$ in* (11).

3 A brief review of the second order parameters' estimators

3.1 The estimation of ρ

We shall base the estimation of ρ on estimators of the type of the ones in Fraga Alves *et al.* (2003). Such a class of estimators has been first parameterised in a tuning parameter $\tau \geq 0$, but more generally we may have τ real (Caeiro and Gomes, 2006). It is defined as,

$$\widehat{\rho}_\tau(k) \equiv \widehat{\rho}_n(k;\tau) := -\left|\frac{3(T_n^{(\tau)}(k) - 1)}{T_n^{(\tau)}(k) - 3}\right|, \quad T_n^{(\tau)}(k) := \frac{\big(M_n^{(1)}(k)\big)^\tau - \big(M_n^{(2)}(k)/2\big)^{\tau/2}}{\big(M_n^{(2)}(k)/2\big)^{\tau/2} - \big(M_n^{(3)}(k)/6\big)^{\tau/3}}, \quad (17)$$

for $\tau \neq 0$ and with the usual continuation for $\tau = 0$, where, with V_{ik} given in (6),

$$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k} V_{ik}^j, \; j \geq 1 \qquad \big[M_n^{(1)} \equiv H, \text{ the Hill estimator in (8)}\big].$$

We shall here summarize a result proved in Fraga Alves *et al.* (2003), making now explicit the random behaviour of the term leading to the asymptotic variance, needed later on, when dealing with the estimation of the three parameters, γ, β and ρ, at levels of the same order.

Theorem 2 (Fraga Alves *et al.*, 2003). *Under the second order framework in* (2), *if k is intermediate, i.e.,* (5) *holds, and if $\sqrt{k}\ A(n/k) \to \infty$, as $n \to \infty$, the statistics $\widehat{\rho}_n(k;\tau)$ in* (17) *converge in probability towards ρ, as $n \to \infty$, for any real τ. If* (3) *holds, with $\rho < 0$, we can further guarantee that there exist constants $(u_{\rho,\tau}, v_{\rho,\rho'}, \sigma_\rho)$ and an asymptotically standard normal r.v. W_k^R, such that*

$$\widehat{\rho}_n(k;\tau) - \rho \ \stackrel{d}{=}\ \frac{\sigma_\rho\ W_k^R}{\sqrt{k}\ A(n/k)} + \big(u_{\rho,\tau} A(n/k) + v_{\rho,\rho'} B(n/k)\big)(1 + o_p(1)). \tag{18}$$

Moreover, with $Z_k^{(\alpha)} = \frac{1}{\sqrt{k}}\ \sum_{i=1}^k E_i^\alpha/\Gamma(\alpha+1) - \sqrt{k}$, for any $\alpha \geq 1$, $\Gamma(t)$ denoting the complete Gamma function, we can write

$$W_k^R = \Big((3-\rho)Z_k^{(1)} - (3-2\rho)Z_k^{(2)} + (1-\rho)Z_k^{(3)}\Big) / \sqrt{2\rho^2 - 2\rho + 1}. \tag{19}$$

Consequently, if (13) *holds, $\sqrt{k}\ A(n/k)\,(\widehat{\rho}_n(k;\tau) - \rho)$ is asymptotically normal with a mean value $\lambda_{_A} u_{\rho,\tau} + \lambda_{_B} v_{\rho,\rho'}$ and variance*

$$\sigma_\rho^2 \equiv \sigma_\rho^2(\gamma) = \big(\gamma(1-\rho)^3/\rho\big)^2 \left(2\rho^2 - 2\rho + 1\right). \tag{20}$$

Let us next assume that $k_1 = k_1^{opt}$ is "optimal" for the estimation of ρ, in the sense of a value k_1 that enables us to guarantee the asymptotic normality of the ρ-estimator with a non-null asymptotic bias. That level k_1 is then such that $\sqrt{k_1}\ A(n/k_1)B(n/k_1) \ \to\ \lambda_{_{B_1}}$, finite, and $\sqrt{k_1}\ A^2(n/k_1) \ \to\ \lambda_{_{A_1}}$, also finite, with at least one of them non-null, let us say $\lambda_{_{B_1}}$. We then get $k_1 = O\big(n^{-2(\rho+\rho')/(1-2(\rho+\rho'))}\big)$. Denoting $\hat\rho = \hat\rho_n(k_1;\tau)$ for any τ, i.e., any of the ρ-estimators in this section computed at such a level k_1, $\{\hat\rho - \rho\}$ is, in probability, of the order of $1/\left(\sqrt{k_1}A(n/k_1)\right) = O\big(n^{\rho'/(1-2(\rho+\rho'))}\big) = o(1/\ln n)$, i.e., (15) holds, and this is the crucial condition on $\hat\rho$, needed in Theorem 1, if we want to keep equal to γ^2 the asymptotic variance of the reduced-bias estimator in (11). At the current state-of-the-art, such a k_1 has only a theoretical interest. From a practical point of view additional research is needed, in order to have adaptive ways of selecting this optimal threshold $k_1 = k_1^{opt}$.

3.2 Estimation of β based on the scaled log-spacings

We have here considered the β-estimator obtained in Gomes and Martins (2002) and based on the scaled log-spacings W_i, $1 \leq i \leq k$, in (7). On the basis of any consistent estimator $\hat\rho$ of the second order parameter ρ, we shall consider the β-estimator, $\hat{\beta}(k;\hat{\rho})$, where, for any $\rho<0$,

$$\hat{\beta}(k;\rho) := \frac{\left(\frac{k}{n}\right)^{\rho}\left\{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho}\right)\left(\frac{1}{k}\sum_{i=1}^{k}W_i\right)-\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho}W_i\right)\right\}}{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho}\right)\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\rho}W_i\right)-\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-2\rho}W_i\right)}. \quad (21)$$

Gomes and Martins (2002) kept up to the second order framework and studied the behaviour of $\hat{\beta}(k;\rho)$ in (21). From Theorem 3 in Gomes *et al.* (2008a), we can guarantee that under the second order framework in (2), with $\rho<0$ and $\widehat{\rho}_\tau(k)$ in (17), the rate of convergence of $\hat{\beta}(k;\widehat{\rho}_\tau(k))$ is of the order of $\{\ln(n/k)/(\sqrt{k}\ A(n/k))\}$. Under the assumption that $\ln(n/k)=o(\sqrt{k}A(n/k))$, $\hat{\beta}(k;\widehat{\rho}_\tau(k))$ is thus consistent for the estimation of β, and if (15) holds for $\widehat{\rho}_\tau(k)$,

$$\hat{\beta}(k;\widehat{\rho}_\tau(k)-\beta \overset{p}{\sim} -\beta\ln(n/k)\left(\widehat{\rho}_\tau(k))-\rho\right). \quad (22)$$

Here, we go into the third order framework in (3), and assume, just as in Gomes *et al.* (2008a), that β and ρ are going to be estimated at the same level k. On the basis of Theorem 2, we can state, without the need of a proof:

Theorem 3. *Under the third order framework in* (3)*, if apart from* $\sqrt{k}A(n/k)/\ln(n/k)\to\infty$*, we assume* (13)*, then, with* σ_ρ^2 *given explicitly in* (20)*,* $\sqrt{k}\ A(n/k)\big(\beta-\hat{\beta}(k;\widehat{\rho}_\tau(k))\big)/(\beta\ln(n/k))$ *is asymptotically* $Normal\big(\lambda_{A}u_{\rho,\tau}+\lambda_{B}v_{\rho,\rho'},\ \sigma_\rho^2\big)$*, with* $(u_{\rho,\tau},v_{\rho,\rho'},\sigma_\rho)$ *implicitly given in* (18).

4 Further results on the asymptotic behaviour of the MVRB tail index estimators

Let us define, for any real τ,

$$\widetilde{H}_\tau(k;k_1) := \overline{H}_{\hat{\beta}(k_1;\widehat{\rho}_\tau(k_1)),\widehat{\rho}_\tau(k_1)}(k), \quad (23)$$

with $\overline{H}_{\hat{\beta},\hat{\rho}}(k)$, $\hat{\rho}_\tau(k)$ and $\hat{\beta}(k,\rho)$ given in (11), (17) and (21), respectively. On the basis of Theorems 1, 2 and 3, we state the following result.

Theorem 4. *For the class of models in* (3)*, with A and B chosen as in* (4)*, let us consider a value* $k_1=k_1^{opt}$ *"optimal" for the estimation of* ρ*, in the sense that* $\sqrt{k_1}A(n/k_1)\to\infty$*,* $\sqrt{k_1}A^2(n/k_1)\to\lambda_{A_1}$*, finite, and* $\sqrt{k_1}A(n/k_1)B(n/k_1)\to\lambda_{B_1}$*, also finite, with at least one of them non-null, and the tail index estimator* $\widetilde{H}_\tau(k;k_1)$ *given in* (23).

1. *If $k = o(k_1)$, $\sqrt{k}\big(\widetilde{H}_\tau(k;k_1)-\gamma\big)$ is asymptotically normal with null mean value and variance equal to γ^2.*

2. *Let us next assume that $\sqrt{k}\ A(n/k) \to \infty$ and, with $\lambda_{_A}$ and $\lambda_{_B}$ finite, $\sqrt{k}\ A^2(n/k) \to \lambda_{_A}$ and $\sqrt{k}\ A(n/k)B(n/k) \to \lambda_{_B}$. More specifically, let k and k_1 be sequences of intermediate integers such that $k/k_1 \to q > 0$, finite, Then,*

$$\sqrt{k}\big(\widetilde{H}_\tau(k;k_1) - \gamma\big) \underset{n\to\infty}{\overset{d}{\rightarrow}} \ Normal\big(b^*,\ \sigma^2(q)\big),$$

where, with $(u_{_{\overline{H}}}, v_{_{\overline{H}}})$ defined in (14), $(u_{\rho,\tau}, v_{\rho,\rho'})$ the constants in (18) and $a_{_{\overline{H}}} = -1/(1-\rho)$,

$$b^* = \lambda_{_A}\left(u_{_{\overline{H}}} + u_{\rho,\tau}\ q^\rho a_{_{\overline{H}}}\left(\ln q - a_{_{\overline{H}}}\right)\right) + \lambda_{_B}\left(v_{_{\overline{H}}} + v_{\rho,\rho'}\ q^{\rho'} a_{_{\overline{H}}}\left(\ln q - a_{_{\overline{H}}}\right)\right),\ \textit{and}$$

$$\sigma^2(q) = \sigma^2(q;\gamma,\rho) = \gamma^2\big(1 + q^{1-2\rho}\big(\ln q + 1/(1-\rho)\big)^2(1-\rho)^4(2\rho^2 - 2\rho + 1)/\rho^2\big), \quad (24)$$

i.e., we get the same rate of convergence, of the order of $1/\sqrt{k}$, for $\widetilde{H}_\tau(k;k_1)$, but with a non-null bias and an asymptotic variance dependent upon q. Such an asymptotic variance is equal to γ^2 for $q = 0$ and $q = q_0 = \exp(-1/(1-\rho))$, and increases with q, after q_0.

Proof. From equations (16) and (22),

$$\hat{R}_n(k;k_1) := \sqrt{k}\left\{\widetilde{H}_\tau(k;k_1) - \overline{H}_{\beta,\rho}(k)\right\} \overset{p}{\sim}\ a_{_{\overline{H}}}\left(\widehat{\rho} - \rho\right)\sqrt{k}A\left(n/k\right)\left(\ln\left(k/k_1\right) - a_{_{\overline{H}}}\right).$$

If $\sqrt{k_1}\ A^2(n/k_1) \to \lambda_{_{A_1}}$ and $\sqrt{k_1}\ A(n/k_1)B(n/k_1) \to \lambda_{_{B_1}}$, both finite, then $\widehat{\rho} - \rho = O_p\left(1/\left(\sqrt{k_1}\ A(n/k_1)\right)\right)$ and

$$\hat{R}_n(k;k_1) \overset{p}{\sim} \sqrt{k}A(n/k)\left(\hat{\rho} - \rho\right)\frac{\ln(k/k_1)}{1-\rho} = O_p\Big(\Big(\frac{k}{k_1}\Big)^{\frac{1}{2}-\rho}\ln\Big(\frac{k}{k_1}\Big)\Big).$$

1. If $k = o(k_1)$, $\hat{R}_n(k;k_1)$ converges in probability towards zero, as $n \to \infty$, and the stated asymptotic normality follows.

2. Let us next think that $k/k_1 \to q > 0$. Since $\sqrt{k_1}\ A(n/k_1) \sim \sqrt{k}\ A(n/k)\ q^{\rho-\frac{1}{2}}$, $A(n/k_1) \sim q^\rho A(n/k)$ and $B(n/k_1) \sim q^{\rho'}B(n/k)$, we can write

$$\begin{aligned}\sqrt{k}\left(\widetilde{H}_\tau(k;k_1) - \gamma\right) &= \gamma\ Z_k^{(1)} + \sqrt{k}A(n/k)\left(u_{_{\overline{H}}}A(n/k) + v_{_{\overline{H}}}\ B(n/k)\right)(1 + o_p(1)) \\ &+ \left(\sigma_\rho W_{k_1}^R + \sqrt{k}\ A(n/k)q^{\rho-\frac{1}{2}}\left(u_{\rho,\tau}A(n/k)\ q^\rho + v_{\rho,\rho'}B(n/k)\ q^{\rho'}\right)(1 + o_p(1))\right) \\ &\qquad \times q^{\frac{1}{2}-\rho}\ a_{_{\overline{H}}}\left(\ln q - a_{_{\overline{H}}} + o(1)\right)(1 + o(1)),\end{aligned}$$

with $a_{\overline{H}} = -1/(1-\rho)$, $(u_{\overline{H}}, v_{\overline{H}})$ defined in (14) and $(u_{\rho,\tau}, v_{\rho,\rho'}, \sigma_\rho, W_k^R)$ given in (18), W_k^R and σ_ρ made explicit in (19) and (20), respectively. Then

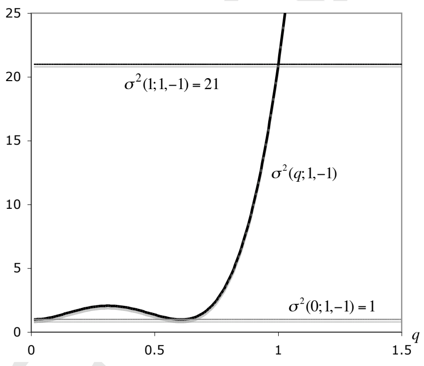
$$\begin{aligned}\sqrt{k}\left(\widetilde{H}_\tau(k;k_1)-\gamma\right) &= \gamma\ Z_k^{(1)} + \sigma_\rho\ q^{\frac{1}{2}-\rho}\ a_{\overline{H}}\left(\ln q - a_{\overline{H}}\right)\ W_{k_1}^R \\ &+ \sqrt{k}\ A^2(n/k)\left(u_{\overline{H}} + u_{\rho,\tau}\ q^\rho a_{\overline{H}}\left(\ln q - a_{\overline{H}}\right)\right)(1+o_p(1)) \\ &+ \sqrt{k}\ A(n/k)B(n/k)\left(v_{\overline{H}} + v_{\rho,\rho'}\ q^{\rho'} a_{\overline{H}}\left(\ln q - a_{\overline{H}}\right)\right)(1+o_p(1)).\end{aligned}$$

As $Cov\left(Z_k^{(1)},\ W_{k_1}^R\right) = 0$, the variance of $\left\{\gamma\ Z_k^{(1)} + \sigma_\rho\ q^{\frac{1}{2}-\rho}\ a_{\overline{H}}\left(\ln q - a_{\overline{H}}\right)\ W_{k_1}^R\right\}$ is the value $\sigma^2(q;\ \gamma,\rho)$ in (24), and the remaining of the proof follows straightforwardly.

□

Remark 2. *Note that if we choose $k_1 = k_1^{opt}$, optimal for the estimation of ρ, the values k in Theorem 1, such that $\sqrt{k}A(n/k) \to \infty$, $\sqrt{k}A^2(n/k) \to 0$ and $\sqrt{k}A(n/k)B(n/k) \to 0$, satisfy the condition $k = o(k_1)$ in item* 1. *of* Theorem 4.

The pattern of $\sigma^2(q;\ \gamma,\rho)$ in (24), as a function of q, is of the same type for all $(\gamma,\ \rho)$, and is pictured in Figure 1, *left*: this variance converges towards $\sigma_0^2 = \gamma^2$, as $q \to 0$, next increases till a value slightly larger than γ^2, then decreases again till γ^2 at $q_0 = \exp\left(-1/(1-\rho)\right)$, $e^{-1} < q_0 < 1$, and finally increases fast, taking the value $\sigma_1^2 = \gamma^2\left(1 + ((1-\rho)/\rho)^2 - 2(1\rho)^3/\rho\right)$, for $q = 1$. As the variance of the estimator $\widetilde{H}_\tau(k;k_1)$ in (23), with $k/k_1 \to q > 0$, is well approximated by $\sigma^2(q;\gamma,\rho)/k$, we provide in Figure 1, *right*, the pattern of $\sigma^2(q;\gamma,\rho)/q$, which provides an indication on the behaviour of the variance of our estimator as a function of q.



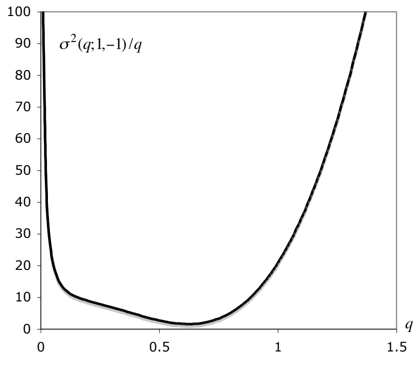


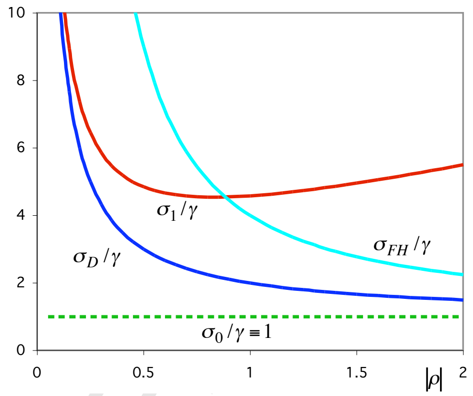
Figure 1: Pattern of $\sigma^2(q;\gamma,\rho)$ (*left*) and $\sigma^2(q;\gamma,\rho)/q$ (*right*), as a function of q, for $\gamma = 1$ and $\rho = -1$.

Remark 3. *If we compare Theorems* 1 *and* 4*, we see that the estimation of γ, β and ρ at the same level k induces an increase in the asymptotic variance of the final γ-estimator, unless we choose*

$q = q_0 = \exp(-1/(1-\rho))$ *and* $k = q \times k_1$*, with* k_1 *optimal for the estimation of* ρ*. The use of* $q = 1$ *in* (24) *leads us to the asymptotic variance* $\sigma_1^2 = \sigma^2(1) = \gamma^2\{1+((1-\rho)/\rho)^2 - 2(1-\rho)^3/\rho\}$*, greater than* $\sigma_0^2 = \gamma^2$*, the value associated with* $q = 0$ *in* (24)*. As noticed before in* Gomes and Martins (2002)*, the asymptotic variance of the estimator in* Feuerverger and Hall (1999) (*where also the three parameters are computed at the same level* k) *is given by* $\sigma_{FH}^2 := \gamma^2\left((1-\rho)/\rho\right)^4$*, the asymptotic variance also achieved by* Peng and Qi (2004)*, for an approximate second-order reduced-bias maximum likelihood tail index estimator. Moreover, it is also known that if we estimate* ρ *at an adequate large level* k_1*, but estimate both* γ *and* β *at the same level* k*, we already induce an extra increase in the asymptotic variance of the final* γ*-estimator, which is then equal to* $\sigma_D^2 = \gamma^2((1-\rho)/\rho)^2$*, the minimal asymptotic variance of any "asymptotically unbiased" estimator in Drees' class of functionals* (Drees, 1998)*. We have*

$$\sigma_0 < \sigma_D < \sigma_1 < \sigma_{FH} \quad \textit{if} \quad |\rho| < 0.8832 \quad \textit{and} \quad \sigma_0 < \sigma_D < \sigma_{FH} < \sigma_1 \quad \textit{if} \quad |\rho| > 0.8832.$$

In Figure 1 *we provide both a picture and some values of* $\sigma_0/\gamma \equiv 1$, σ_D/γ, σ_1/γ *and* σ_{FH}/γ*, as functions of* $|\rho|$.



$\|\rho\|$	σ_0/γ	σ_D/γ	σ_1/γ	σ_{FH}/γ
0.1	1.00	11.00	12.19	121.00
0.2	1.00	6.00	7.37	36.00
0.3	1.00	4.33	5.87	18.78
0.4	1.00	3.50	5.19	12.25
0.5	1.00	3.00	4.85	9.00
1.0	1.00	2.00	4.58	4.00
1.5	1.00	1.67	4.96	2.78
2.0	1.00	1.50	5.50	2.25
2.5	1.00	1.40	6.10	1.96
3.0	1.00	1.33	6.74	1.78

Figure 2: Asymptotic standard deviations, σ_0, σ_D and σ_1, together with σ_{FH}, for $\gamma = 1$

It is obvious from Figures 1 and 2 that, whenever possible, it seems convenient to estimate both β and ρ externally, at a k_1-value higher than the value k used for the estimation of the tail index γ, if we want to work with a tail index estimator potentially better than the Hill estimator for all k. Ideally, the value k_1 should be optimal for the estimation of the second order parameter ρ. The optimal rate k/k_1 depends obviously on ρ, but it is not a long way from 0.5 for the most common models. Indeed, for $\rho = -0.25, -0.5, -1, -1.5, -2$ we get $q_{min} := \arg\min_q \sigma^2(q;\gamma,\rho)/q = 0.46, 0.53, 0.62, 0.68, 0.72$, respectively, a value quite close to $q_0 = \exp(-1/(1-\rho)) = 0.45, 0.51, 0.61, 0.67, 0.72$, respectively.

5 A small-scale simulation study

In practice, and at the current state-of-the-art, Theorem 1 and the first part of Theorem 4 are more relevant than the second part of Theorem 4, in the sense that it is easier to work with $k = o(k_1)$, with k_1 not necessarily "optimal" but enabling the validity of condition (15), basing the estimation of γ on a value k that can be merely the estimated optimal k_0^H in (10), or a slightly larger value of the same order. For non-optimal, but practical ways of choosing such a k, see Gomes *et al.* (2008b). But the second part of Theorem 4 is potentially more powerful. We merely need to have adequate ways of estimating the optimal level $k_1 = k_1^{opt}$ for the estimation of ρ through the estimator in (17), again in the sense of minimal mean squared error. Provided that an estimate $\hat{k}_1^{opt}$ of k_1^{opt} is available, as well as an estimate $\hat{\rho}$ of ρ, we should then compute the tail index estimator $\overline{H}$ in (11) at a level like $\hat{k} = \hat{k}_1^{opt} \times \exp(-1/(1-\hat{\rho}))$ or even $\hat{k} = \hat{k}_1^{opt}/2$. For the estimation of the second order parameters (β, ρ), we advise the following **Algorithm** (similar to the one in Gomes and Pestana, 2007) :

S1. Given a sample $(X_1, X_2, \cdots, X_n)$, plot, for $\tau = 0$ and $\tau = 1$, the estimates $\hat{\rho}_\tau(k)$ in (17).

S2. Consider $\{\hat{\rho}_\tau(k)\}_{k\in\mathcal{K}}$, for $k \in \mathcal{K} = \left(\left[n^{0.995}\right], \left[n^{0.999}\right]\right)$, and compute their median, denoted ρ_τ. Choose the *tuning parameter* $\tau_0 := \arg\min_\tau \sum_{k\in\mathcal{K}} \left(\hat{\rho}_\tau(k) - \rho_\tau\right)^2$, and work with $\hat{\rho}_{\tau_0} := \hat{\rho}_{\tau_0}(k_1^h)$, with the superscript h in k_1^h standing for "heuristic", and

$$k_1^h = \left[n^{0.995}\right]. \tag{25}$$

S3. Consider the β-estimator $\hat{\beta}_{\tau_0} = \widehat{\beta}(k_1^h; \hat{\rho}_{\tau_0})$, with $\widehat{\beta}(k;\rho)$ given in (21).

The choice of the level k_1^h in (25) is not crucial for the estimation of (β, ρ). We merely need to consider any reasonably large value of the order of $n^{1-\epsilon}$ for small ϵ, due to the high stability of $\hat{\rho}_{\tau_0}(k)$ around the target ρ for large values k and for a large class of models. With a slight restriction in the class of models where asymptotic normality holds, we are then able to guarantee the validity of (15). The choice of the level k_1^h in (25) can however be crucial if we decide replacing the optimal $k_1 = k_1^{opt}$ in Theorem 4 by $k_1 = k_1^h$ in (25), unless $|\rho|$ is close to 1. As we do not have yet techniques for the estimation of k_1^{opt}, optimal for the estimation of ρ, but we know that for a large variety of heavy-tailed models $\rho = \rho'$ and consequently, such a k_1^{opt} is of the order of $n^{-4\rho/(1-4\rho)}$, we have decided not to pay attention to the scale factor in $k_1^{opt} = C_{\beta,\beta',\rho} n^{-4\rho/(1-4\rho)}$, and to consider the following extra step for a heuristic γ-estimation:

S4. Compute $\widehat{\gamma} = \overline{H}_{\hat{\beta},\hat{\rho}}(\hat{k})$, with $\hat{k} = n^{-4\hat{\rho}_{\tau_0}/(1-4\hat{\rho}_{\tau_0})} \times \exp(-1/(1-\hat{\rho}_{\tau_0}))$.

As soon as we have techniques for an estimation of $C_{\beta,\beta',\rho}$, an interesting topic beyond the scope of this paper, such an estimate should be included as a scale in the value $\hat{k}$ in Step **S4**.

Remark 4. *This algorithm leads in almost all situations to $\tau = 0$ whenever $|\rho| \le 1$ and $\tau = 1$, otherwise. Such an educated guess usually provides slightly better results than a "noisy" estimation of τ, it has been recommended in practice and will be used in the simulations of this Section.*

We have performed Monte Carlo simulation experiments for the following parents with $\rho = \rho'$: the *Fréchet* model, with d.f. $F(x) = \exp(-x^{-1/\gamma})$, $x \ge 0$, $\gamma > 0$, for which $\rho = \rho' = -1$, $\beta = 1/2$, $\beta' = 5/6$, and the *Generalized Pareto* (*GP*) model, with d.f. $F(x) = 1 - (1+\gamma x)^{-1/\gamma}$, $x \ge 0$, $\gamma > 0$, for which $\rho = \rho' = -\gamma$ and $\beta = \beta' = 1$. To understand whether the choice in **S4.** has some robustness to models with $\rho \ne \rho'$, we have also simulated an *extreme value* d.f., with $\gamma = 0.75$. For this model we get $\rho = -\gamma = -0.75$ and $\rho' = \gamma - 1 = -0.25$. For all these distributions Theorem 4 holds, provided that we choose $k_1 = O\big(n^{-2(\rho+\rho')/(1-2(\rho+\rho'))}\big)$.

For each value of $n = 200$, 500, 1000, 2000, 5000 and 10000, for each model and for the estimate $(\hat{\beta}_\tau, \hat{\rho}_\tau)$ of (β, ρ), suggested before in this Section, we have simulated the distributional behaviour of $\widehat{H}_{0\tau} := H_n(\hat{k}_{0\tau}^H)$, $\widehat{\overline{H}}_{0\tau|H} := \overline{H}_{\hat{\beta}_\tau,\hat{\rho}_\tau}(\hat{k}_{0\tau}^H)$ and $\widehat{\overline{H}}_{0\tau|1} := \overline{H}_{\hat{\beta}_\tau,\hat{\rho}_\tau}(\hat{k})$, with $\hat{k}$ provided in Step **S4.** of the algorithm, H, k_0^H and $\overline{H}$ given in (8), (10) and (11), respectively. In Table 1 we present relative efficiency ($REFF$) and bias-reduction indicators (BRI) of $\widehat{\overline{H}}_{0\tau|H}$ (first line) and $\widehat{\overline{H}}_{0\tau|1}$ (second line) relatively to $\widehat{H}_{0\tau}$, the Hill estimator computed at its estimated optimal level. We have thus simulated $REFF_{\widehat{\overline{H}}_{0\tau|H}|\widehat{H}_{0\tau}}$, $REFF_{\widehat{\overline{H}}_{0\tau|1}|\widehat{H}_{0\tau}}$, $BRI_{\widehat{\overline{H}}_{0\tau|H}|\widehat{H}_{0\tau}}$ and $BRI_{\widehat{\overline{H}}_{0\tau|1}|\widehat{H}_{0\tau}}$, where for any two estimators $\hat{\gamma}_1$ and $\hat{\gamma}_2$ of γ, we define

$$REFF_{\hat{\gamma}_1|\hat{\gamma}_2} := \sqrt{MSE_s(\hat{\gamma}_2)/MSE_s(\hat{\gamma}_1)}, \quad BRI_{\hat{\gamma}_1|\hat{\gamma}_2} := \left|E_s(\hat{\gamma}_2 - \gamma)/E_s(\hat{\gamma}_1 - \gamma)\right|,$$

so that the higher than one these indicators are, the better the first estimator performs comparatively with the second one. We also present in the third line of each entry of Table 1 the indicator associated with the behaviour of $\overline{H}$ relatively to H, when both γ-estimators are computed at their simulated optimal levels. These results were based on multi-sample simulations of size 5000×10. For details on multi-sample simulation, see Gomes and Oliveira (2001).

Table 1: $REFF$ and BRI indicators

$REFF$ indicators						BRI indicators					
n						n					
200	500	1000	2000	5000	10000	200	500	1000	2000	5000	10000
Fréchet parent, $\gamma = 1$ $(\rho = \rho' = -\mathbf{1})$											
1.026	1.055	1.082	1.109	1.152	1.188	1.845	4.151	10.168	77.432	12.422	6.355
0.974	1.046	1.101	1.160	1.252	1.332	1.789	3.925	10.322	727.308	11.106	5.674
1.200	1.190	1.226	1.299	1.453	1.594	75.998	29.310	5.296	5.742	11.611	16.333
GP parent, $\gamma = 0.5$ $(\rho = \rho' = -\mathbf{0.5})$											
1.639	1.582	1.543	1.509	1.471	1.444	2.099	1.939	1.836	1.750	1.655	1.593
1.636	1.575	1.518	1.452	1.349	1.267	2.200	1.882	1.694	1.542	1.368	1.256
1.377	1.343	1.310	1.277	1.250	1.219	1.289	1.231	1.200	1.251	1.194	1.145
GP parent, $\gamma = 1$ $(\rho = \rho' = -\mathbf{1})$											
1.201	1.156	1.134	1.114	1.099	1.087	2.651	2.305	2.126	1.955	1.839	1.736
1.162	1.184	1.210	1.235	1.272	1.289	2.669	2.217	1.950	1.700	1.455	1.280
1.881	2.066	2.237	2.428	2.743	3.012	8.584	22.085	15.020	217.431	58.593	343.201
GP parent, $\gamma = 2$ $(\rho = \rho' = -\mathbf{2})$											
1.172	1.063	1.046	1.060	1.188	1.166	3.634	2.597	2.359	2.342	2.327	2.207
1.169	1.792	2.128	2.087	1.191	1.170	4.141	2.879	2.329	1.924	1.520	1.320
1.165	1.142	1.133	1.133	1.124	1.131	1.459	1.319	1.321	1.244	1.084	1.118
EV_γ **parent,** $\gamma = 0.75$ $(\rho = -\mathbf{0.75},\ \rho' = -\mathbf{0.25})$											
1.198	1.267	1.304	1.321	1.326	1.330	5.737	138.758	116.318	768.958	36.751	26.594
1.067	1.220	1.322	1.416	1.535	1.625	4.049	20.948	98.951	43.883	20.256	16.500
1.459	1.611	1.788	1.975	2.232	2.389	7.142	3.331	2.858	3.234	3.370	3.466

Remark 5. *Note that only a small percentage of values is smaller than one. They are associated with a sample size* $n = 200$*, the* $REFF$ *measure and a Fréchet parent. Those values are written in italic.*

Finally, and on the basis of the first replicate, with size 5000, we present as a function of the sample fraction k/n the patterns of simulated mean values and mean squared errors of H and

$\overline{H}_0(k; k_1^h)$ in (8) and (23), respectively, for $n = 500, 1000, 5000$ and a *Generalized Pareto* parent, with $\gamma = 0.5$.

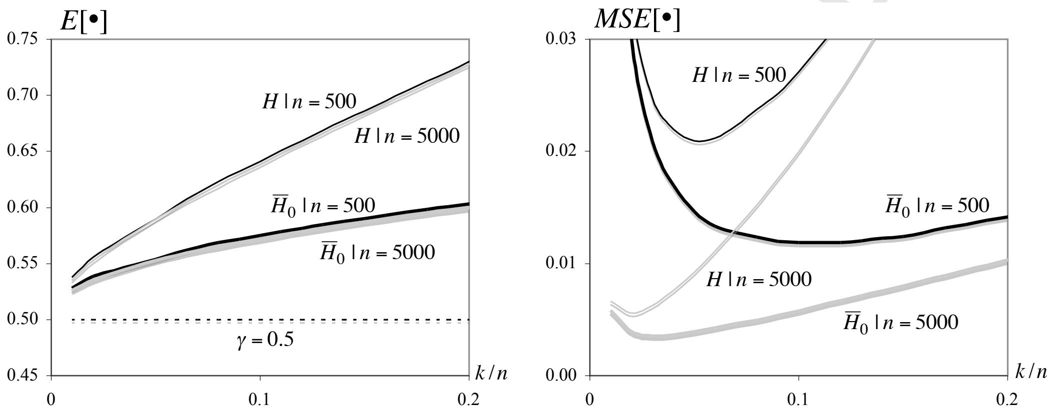


Figure 3: Mean values and mean squared errors of H and $\overline{H}_0(k; k_1^h)$, as a function of the sample fraction k/n used in the estimation, for a few values of n and a *Generalized Pareto* parent, with $\gamma = 0.5$

Figure 3 draws again our attention to the potentiality of these MVRB estimators: they overpass the Hill estimator for all k, both in terms of bias and mean squared errors.

References

[1] Caeiro, F. and Gomes, M.I. (2006). A new class of estimators of a "scale" second order parameter. *Extremes* **9**, 193-211, 2007.

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

[3] Caeiro, F. and Gomes, M.I. (2008b). Semi-parametric second-order reduced-bias high quantile estimation. *Test*, in press, DOI: 10.1007/s11749-008-0108-8, 2008.

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

[5] Drees, H. (1998). A general class of estimators of the extreme value index. *J. Statist. Planning and Inference* **98**, 95-112.

[6] Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.* **27**, 760-781.

[7] 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**, 193-214.

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

[9] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**, 423-453.

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

[11] Gomes, M.I., de Haan, L. and Henriques Rodrigues, L. (2008a). Tail index estimation for heavy-tailed models: accommodation of bias in the weighted log-excesses. *J. Royal Statistical Society* **B70**, Issue 1, 31-52.

[12] Gomes, M.I., de Haan, L. and Peng, L. (2002). Semi-parametric estimation of the second order parameter — asymptotic and finite sample behaviour. *Extremes* **5**:4, 387-414.

[13] Gomes, M.I., Henriques Rodrigues, L., Vandewalle, B. and Viseu, C. (2008b). A heuristic adaptive choice of the threshold for bias-corrected Hill estimators. *J. Statist. Comput. and Simulation* **78**:2, 133-150.

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

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

[16] 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.

[17] Gomes, M. I. and Pestana, D. (2007). A sturdy reduced-bias extreme quantile (VaR) estimator. *J. American Statistical Association*, Vol. **102**, No. 477, 280-292.

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

[19] de Haan, L. (1984). Slow variation and characterization of domains of attraction. In Tiago de Oliveira, ed., *Statistical Extremes and Applications*, 31-48, D. Reidel, Dordrecht, Holland.

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

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

[22] Peng, L. and Qi, Y. (2004). Estimating the first and second order parameters of a heavy tailed distribution. *Australian and New Zealand J. of Statist.* **46**, 305-312.