REVSTAT – Statistical Journal
Volume 3, Number 2, November 2005, 113–136

DIRECT REDUCTION OF BIAS OF THE CLASSICAL HILL ESTIMATOR *

Authors: Frederico Caeiro
– Universidade Nova de Lisboa, FCT(DM) and CEA, Portugal
(fac@fct.unl.pt)

M. Ivette Gomes
– Universidade de Lisboa, FCUL (DEIO) and CEAUL, Portugal
(ivette.gomes@fc.ul.pt)

Dinis Pestana
– Universidade de Lisboa, FCUL (DEIO) and CEAUL, Portugal
(dinis.pestana@fc.ul.pt)

Received: January 2005 Revised: June 2005 Accepted: July 2005

Abstract:

- In this paper we are interested in an adequate estimation of the dominant component of the bias of Hill's estimator of a positive tail index γ, in order to remove it from the classical Hill estimator in different asymptotically equivalent ways. If the second order parameters in the bias are computed at an adequate level k_1 of a larger order than that of the level k at which the Hill estimator is computed, there may be no change in the asymptotic variances of these reduced bias tail index estimators, which are kept equal to the asymptotic variance of the Hill estimator, i.e., equal to γ^2. The asymptotic distributional properties of the proposed estimators of γ are derived and the estimators are compared not only asymptotically, but also for finite samples through Monte Carlo techniques.

Key-Words:

- *Statistics of Extremes; semi-parametric estimation; bias estimation; heavy tails.*

AMS Subject Classification:

- 62G32, 65C05.

---

*Research partially supported by FCT / POCTI and POCI / FEDER.

1. INTRODUCTION AND MOTIVATION FOR THE NEW TAIL INDEX ESTIMATORS

In *Statistics of Extremes*, the tail index γ is the basic parameter of extreme events. Such a parameter plays a relevant role in other extreme events' parameters, like high quantiles and return periods of high levels, among others. The tail index is a real-valued parameter and the heavier the tail, the larger the tail index γ is. Heavy-tailed models have revealed to be quite useful in most diversified fields, like computer science, telecommunication networks, insurance and finance. In the field of Extremes, we usually say that a model F is *heavy-tailed* whenever the *tail function*, $\overline{F} := 1-F$, is a regularly varying function with a negative index of regular variation equal to $\{-1/\gamma\}$, $\gamma>0$, or 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. This means that, for every $x > 0$,

$$(1.1) \qquad \lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-1/\gamma} \quad \Longleftrightarrow \quad \lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^{\gamma} \ .$$

We shall here concentrate on these Pareto-type distributions. Note that (1.1) is equivalent to saying that

$$(1.2) \qquad 1-F(x) = x^{-1/\gamma}\, L_F(x) \quad \Longleftrightarrow \quad U(x) = x^{\gamma}\, L_U(x) \ ,$$

with L_F and L_U slowly varying functions, i.e., functions $L_\bullet$ such that $L_\bullet(tx)/L_\bullet(t) \to 1$, as $t\to\infty$, for all $x>0$.

The *second order parameter* ρ (≤ 0), rules the rate of convergence in the first order condition (1.1) (or equivalently, (1.2)), and is the non-positive parameter appearing in the limiting relation

$$(1.3) \qquad \lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \lim_{t\to\infty} \frac{\ln L_U(tx) - \ln L_U(t)}{A(t)} = \frac{x^{\rho}-1}{\rho} \ ,$$

which we assume to hold for all $x > 0$ and where $|A(t)|$ must then be of regular variation with index ρ (Geluk and de Haan, 1987). We shall assume everywhere that $\rho < 0$.

Remark 1.1. For the strict Pareto model, $F(x) = 1 - C\ x^{-1/\gamma}$, $x \geq C^{\gamma}$, and indeed only for this model, the numerator of the fraction in the left hand-side of (1.3) is null, i.e., $\ln U(tx) - \ln U(t) - \gamma \ln x \equiv 0$.

Remark 1.2. For Hall's class of Pareto-type models (Hall, 1982; Hall and Welsh, 1985), with a tail function

$$(1.4) \qquad 1-F(x) \,=\, Cx^{-1/\gamma}\Big(1 + Dx^{\rho/\gamma} + o\big(x^{\rho/\gamma}\big)\Big) \ , \qquad \text{as} \ \ x\to\infty \ ,$$

$C > 0$, $D \in \mathbb{R}_0$, $\rho < 0$, (1.3) holds and we may choose $A(t) = \gamma\rho D C^{\rho} t^{\rho}$.

This is a class of models where (1.2) (or equivalently, (1.1)) holds true, with an asymptotically constant slowly varying function L_F (or equivalently, L_U).

To obtain information on the distributional behaviour of the second order parameters' estimators, we shall further assume that the rate of convergence in (1.3) is ruled by a function $B(t)$ such that $|B(t)|$ is also of regular variation with the same index ρ, i.e., we assume that

$$\lim_{t\to\infty} \frac{\frac{\ln U(tx)-\ln U(t)-\gamma \ln x}{A(t)} - \frac{x^\rho-1}{\rho}}{B(t)} = \frac{x^{2\rho}-1}{2\rho} \tag{1.5}$$

holds for all $x > 0$.

Remark 1.3. Condition (1.5) holds true for models with a tail function

$$1 - F(x) = C x^{-1/\gamma}\Big(1 + D_1 x^{\rho/\gamma} + D_2 x^{2\rho/\gamma} + o\big(x^{2\rho/\gamma}\big)\Big)\,, \quad \text{as } x \to \infty\ . \tag{1.6}$$

For the most common heavy-tailed models, like the Fréchet and the Student's t, condition (1.5) holds true, i.e., these models belong to the class in (1.6).

For intermediate k, i.e., a sequence of integers $k = k_n$, $1 \le k < n$, such that

$$k = k_n \to \infty\ , \qquad k_n = o(n)\,, \quad \text{as } n \to \infty\ , \tag{1.7}$$

we shall consider, as basic statistics, both the log-excesses over the random high level $\{\ln X_{n-k:n}\}$, i.e.,

$$V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n}\ , \qquad 1 \le i \le k < n\ , \tag{1.8}$$

and the scaled log-spacings,

$$U_i := i\,\big\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\big\}\ , \qquad 1 \le i \le k < n\ , \tag{1.9}$$

where $X_{i:n}$ denotes, as usual, the i-th ascending order statistic (o.s.), $1 \le i \le n$, associated to a random sample $(X_1, X_2, ..., X_n)$.

We may write $X_{i:n} \stackrel{d}{=} U(Y_{i:n})$, where $\{Y_i\}$ denotes a sequence of unit Pareto random variables (r.v.'s), i.e., $P(Y \le y) = 1 - 1/y,\ y \ge 1$. Also, for $j > i$, $Y_{j:n}/Y_{i:n} \stackrel{d}{=} Y_{j-i:n-i}$, $\ln Y_{i:n} \stackrel{d}{=} E_{i:n}$, where $\{E_i\}$ denotes a sequence of independent standard exponential r.v.'s, i.e., $P(E \le x) = 1 - \exp(-x),\ x \ge 0$, and for intermediate k, $Y_{n-k:n} \sim n/k \ \to\ \infty$, as $n \to \infty$. Consequently, whenever we are under the first order framework in (1.1), we get

$$V_{ik} \stackrel{d}{=} \ln \frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})} = \ln \frac{U(Y_{n-k:n}\ Y_{k-i+1:k})}{U(Y_{n-k:n})} \sim \gamma\, E_{k-i+1:k}\ , \qquad 1 \le i \le k\ ,$$

i.e., V_{ik}, $1 \le i \le k$, are approximately the k o.s.'s from an exponential sample of size k and mean value γ. Also, since $\ln Y_{1:i} \overset{d}{=} E_{1:i} \overset{d}{=} E_i/i$,

$$U_i \overset{d}{=} i\left(\ln \frac{U(Y_{n-i+1:n})}{U(Y_{n-i:n})}\right) = i\left(\ln \frac{U(Y_{n-i:n}\, Y_{1:i})}{U(Y_{n-i:n})}\right) \sim \gamma\, E_i \ , \qquad 1 \le i \le k \ ,$$

i.e., the U_i's, $1 \le i \le k$, are approximately independent and exponential with mean value γ. Then the Hill estimator of γ (Hill, 1975),

$$H(k) \equiv H_n(k) = \frac{1}{k}\sum_{i=1}^{k} V_{ik} = \frac{1}{k}\sum_{i=1}^{k} U_i \ , \tag{1.10}$$

is consistent for the estimation of γ whenever (1.1) holds and k is intermediate, i.e., (1.7) holds.

Under the second order framework in (1.3) the asymptotic distributional representation

$$H(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + \frac{1}{1-\rho}\, A(n/k)\,\big(1 + o_p(1)\big) \tag{1.11}$$

holds true (de Haan and Peng, 1998), where $Z_k^{(1)} = \sqrt{k}\big(\sum_{i=1}^{k} E_i/k - 1\big)$ is an asymptotically standard normal r.v.

Remark 1.4. If the underlying model is the strict Pareto model in Remark 1.1, $\ln X_{i:n} = \gamma\, E_{i:n} + \gamma \ln C$, and the use of Rényi's representation of exponential order statistics, as a linear combination of independent unit exponential r.v.'s (Rényi, 1953), $E_{i:n} = \sum_{j=1}^{i} E_j/(n-j+1)$, $1 \le i \le n$, leads us to

$$H(k) \overset{d}{=} \frac{\gamma}{k}\sum_{i=1}^{k}\{E_{n-i+1:n} - E_{n-k:n}\} \overset{d}{=} \frac{\gamma}{k}\sum_{i=1}^{k}\sum_{j=i}^{k}\frac{E_j}{j} \overset{d}{=} \frac{\gamma}{k}\sum_{j=1}^{k} E_j \overset{d}{=} \frac{\gamma}{k}\, Ga(k) \ ,$$

where $Ga(k)$ denotes a Gamma r.v. with a shape parameter equal to k, i.e., a r.v. with probability density function (*p.d.f.*) $x^{k-1}\exp(-x)/\Gamma(k)$, $x \ge 0$, with $\Gamma(t)$ denoting the complete gamma function, $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}\,dx$. Then for every k, the Hill estimator in (1.10) is unbiased for the estimation of γ, i.e., $\mathbb{E}\,(H(k)) = \gamma$ for any k, and $\sqrt{k}\,(H(k)-\gamma)\,/\gamma$ is asymptotically standard normal, as $k \to \infty$.

We shall assume that we are in the class of models in (1.6). Consequently, we may choose

$$A(t) = \alpha\, t^\rho =: \gamma\,\beta\, t^\rho, \quad B(t) = \beta'\, t^\rho \ , \qquad \beta,\, \beta' \ne 0, \quad \rho < 0 \ . \tag{1.12}$$

The adequate accommodation of the bias of Hill's estimator has been extensively addressed in recent years by several authors. The idea is to go further into the second order framework in (1.3). Then,

$$V_{ik} \overset{d}{=} \ln \frac{U(Y_{n-k:n}\, Y_{k-i+1:k})}{U(Y_{n-k:n})} \overset{d}{\approx} \gamma\, E_{k-i+1:k} + A(n/k)\,\frac{Y_{k-i+1:k}^{\rho} - 1}{\rho} \ , \quad 1 \le i \le k \ ,$$

and

$$U_i \overset{d}{=} i \ln \frac{U(Y_{n-i:n} Y_{1:i})}{U(Y_{n-i:n})} \overset{d}{\approx} \gamma\left(1 + \frac{A(n/k)}{\gamma} \left(\frac{k}{i}\right)^{\rho}\right) E_i \ , \qquad 1 \leq i \leq k \ .$$

Beirlant *et al.* (1999) and Feuerverger and Hall (1999) work with the scaled log-spacings U_i, $1 \leq i \leq k$, in slightly different but equivalent ways, and consider the joint estimation of the first order parameter γ and the second order parameters at the same level k; in a similar set-up, Gomes and Martins (2002) advance with the "external" estimation of the second order parameter ρ, i.e., the estimation of ρ at a lower level (larger k) than the one used for the tail index estimation, being then able to reduce the asymptotic variance of the proposed tail index estimator, but they pay no special attention to the extra "scale" parameter $\beta \neq 0$ in the A function in (1.12). More recently, Gomes *et al.* (2004b) deal with a joint external estimation of both the "scale" and the "shape" parameters in the A function, being able to reduce the bias without increasing the asymptotic variance, which is kept at the value γ^2, the asymptotic variance of Hill's estimator, for an adequate choice of the level k. Such an estimator, also considered here for comparison with two new proposed estimators, is based on a linear combination of the excesses V_{ik} in (1.8), and is given by

$$(1.13) \qquad WH_{\widehat{\beta},\widehat{\rho}}(k) := \frac{1}{k} \sum_{i=1}^{k} e^{\widehat{\beta}\,(n/k)^{\widehat{\rho}}\,\left((i/k)^{-\widehat{\rho}}-1\right)/\left(\widehat{\rho}\ln(i/k)\right)}\, V_{ik} \ ,$$

for adequate consistent estimators $\widehat{\beta}$ and $\widehat{\rho}$ of the second order parameters β and ρ, respectively, and with WH standing for *Weighted Hill.* In the same spirit, Gomes and Pestana (2004) study, mainly computationally, the estimator

$$(1.14) \qquad \overline{H}_{\widehat{\beta},\widehat{\rho}}(k) := H(k) - \frac{\widehat{\beta}}{1-\widehat{\rho}} \left(\frac{n}{k}\right)^{\widehat{\rho}} H\left(\left(\frac{(1-\widehat{\rho})^2\, n^{-2\widehat{\rho}}}{-2\,\widehat{\rho}\,\widehat{\beta}^2}\right)^{1/(1-2\widehat{\rho})}\right),$$

with H the Hill estimator in (1.10).

We shall here consider the estimator

$$(1.15) \qquad \overline{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k) := H(k)\left(1 - \frac{\widehat{\beta}}{1-\widehat{\rho}} \left(\frac{n}{k}\right)^{\widehat{\rho}}\right),$$

together with the asymptotically equivalent variant,

$$(1.16) \qquad \overline{\overline{\overline{H}}}_{\widehat{\beta},\widehat{\rho}}(k) := H(k) \exp\left(-\frac{\widehat{\beta}}{1-\widehat{\rho}} \left(\frac{n}{k}\right)^{\widehat{\rho}}\right).$$

The dominant component of the bias of Hill's estimator, $A(n/k)/(1-\rho) = \gamma\,\beta\,(n/k)^{\rho}/(1-\rho)$, is thus estimated through $H(k)\,\widehat{\beta}\,(n/k)^{\widehat{\rho}}/(1-\widehat{\rho})$ and directly removed from Hill's classical tail index estimator, through two asymptotically equivalent expressions, provided that k is intermediate, i.e., provided that (1.7) holds true.

Remark 1.5. Note that the estimator $\overline{H}$ in (1.14) has been built in a way similar to the estimator $\widetilde{\overline{H}}$ in (1.15). The difference is that the bias $\gamma\,\beta\,(n/k)^{\rho}/(1-\rho)$ is estimated through $H\big(\widehat{k}_0\big)\,\widehat{\beta}\,(n/k)^{\widehat{\rho}}/(1-\widehat{\rho})$, with $\widehat{k}_0$ an estimate of the "optimal" level for the Hill estimator, in the sense of minimum mean squared error in Hall's class of models.

Remark 1.6. The reason to consider the two asymptotically equivalent estimators in (1.15) and (1.16) — also asymptotically equivalent to the estimator in (1.14) — lies in the fact that we have the clear experience that asymptotically equivalent estimators may exhibit quite different sample paths' properties. In practice, one never knows the peculiarities of the underlying models, and it is thus sensible to work with a set of a few estimators of the primary parameter of rare events, in order to take "the best decision".

1.1. A technical motivation

In the lines of Gomes and Martins (2004):

Lemma 1.1. *Under the second order framework in* (1.3), *and for levels* k *such that* (1.7) *holds, the distributional representation*

$$\frac{\alpha}{k}\sum_{i=1}^{k}\Big(\frac{i}{k}\Big)^{\alpha-1}U_i \;\stackrel{d}{=}\; \gamma+\frac{\gamma\,\alpha}{\sqrt{(2\alpha-1)\,k}}\;Z_k^{(\alpha)}+\frac{\alpha\,A(n/k)}{\alpha-\rho}\,\big(1+o_p(1)\big) \tag{1.17}$$

holds true for any $\alpha\geq 1$, *where*

$$Z_k^{(\alpha)} \;=\; \sqrt{(2\alpha-1)\,k}\,\left(\frac{1}{k}\sum_{i=1}^{k}\Big(\frac{i}{k}\Big)^{\alpha-1}E_i-\frac{1}{\alpha}\right) \tag{1.18}$$

are asymptotically standard normal r.v.'s. The asymptotic covariance structure between the r.v.'s in (1.18) *is given by*

$$\mathbb{C}ov_{\infty}\big(Z_k^{(\alpha)},\,Z_k^{(\beta)}\big) \;=\; \frac{\sqrt{(2\alpha-1)\,(2\beta-1)}}{\alpha+\beta-1}\;. \tag{1.19}$$

If we assume that only the tail index parameter γ in unknown, and similarly to the result in Gomes and Pestana (2004), we shall now state and prove a theorem that provides an obvious technical motivation for the estimator in (1.15) (or in (1.16)):

Theorem 1.1. *Under the second order framework in* (1.3)*, further assuming that $A(t)$ may be chosen as in* (1.12)*, and for levels k such that* (1.7) *holds, we get, for $\widetilde{\overline{H}}_{\beta,\rho}(k)$ in* (1.15) (*or for $\overline{\overline{H}}_{\beta,\rho}(k)$ in* (1.16))*, an asymptotic distributional representation of the type*

$$\gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + R_k\,, \qquad \textit{with}\ \ R_k = o_p\big(A(n/k)\big)\ , \tag{1.20}$$

where $Z_k^{(1)}$ is the asymptotically standard normal r.v. in (1.18) *for $\alpha = 1$. Consequently, both $\sqrt{k}\,(\widetilde{\overline{H}}_{\beta,\rho}(k) - \gamma)$ and $\sqrt{k}\,(\overline{\overline{H}}_{\beta,\rho}(k) - \gamma)$ are asymptotically normal with variance equal to γ^2, and with a null mean value not only when $\sqrt{k}\ A(n/k) \longrightarrow 0$, but also when $\sqrt{k}\ A(n/k) \longrightarrow \lambda \neq 0$, finite, as $n \to \infty$.*

Proof: The results related to the estimator in (1.15) come straighforwardly from the fact that if all parameters are known, apart from the tail index γ, we get from (1.11),

$$\begin{aligned}\widetilde{\overline{H}}_{\beta,\rho}(k) &\stackrel{d}{=} \left(\gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + \frac{A(n/k)}{1-\rho}\,(1 + o_p(1))\right) \times \left(1 - \frac{A(n/k)}{\gamma(1-\rho)}\right)\\ &\stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + o_p(A(n/k)\ ,\end{aligned}$$

i.e., (1.20) holds. Since, for intermediate k,

$$\exp\left(-\frac{A(n/k)}{1-\rho}\right) = 1 - \frac{A(n/k)}{1-\rho} + o_p\big(A(n/k)\big)\ ,$$

the same distributional representation in (1.20) holds true for the tail index estimator in (1.16). The remaining of the theorem follows then straightforwardly. □

1.2. A graphical motivation

For the second order parameters' estimators, discussed later on, in section 2, and as a supporting example of the technical motivation given in 1.1, we exhibit in Figure 1, the differences between the sample paths of the estimators $\widetilde{\overline{H}}_{\bullet}(k)$ in (1.15), for a sample of size $n = 10{,}000$ from a Fréchet model, with d.f. $F(x) = \exp(-x^{-1/\gamma})$, $x \geq 0$, with $\gamma = 1$, when we compute $\widehat{\beta}$ and $\widehat{\rho}$ at the same level k used for the estimation of the tail index γ (*left*), when we compute only $\widehat{\beta}$ at that same level k, being $\widehat{\rho}$ computed at a larger k-value, let us say an intermediate level k_1 such that $\sqrt{k_1}\ A(n/k_1) \to \infty$, as $n \to \infty$ (*center*) and when both $\widehat{\rho}$ and $\widehat{\beta}$ are computed at that high level k_1 (*right*). For the notation used in Figure 1, see subsection 4.2. The high stability, around the target value

$\gamma = 1$, of the sample path in Figure 1 (*right*), is for sure related to the result in Lemma 1.1, and not purely coincidential. It thus seems sensible to compare asymptotically these estimation procedures, in order to detect the reasons for the differences in behaviour.

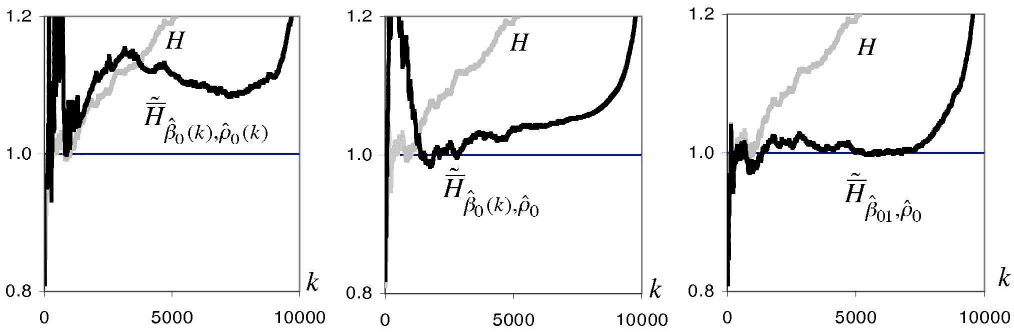


Figure 1: Sample paths of the Hill estimates H in (1.10) and the tail index estimates $\widetilde{\widetilde{H}}$ in (1.15), obtained through the estimation of (β, ρ) at the level $k_1 := \min(n-1,\, 2n^{0.995}/\ln\ln n)$ (*right*) versus the estimation at the same level both for β and ρ (*left*) and only for β (*center*).

1.3. Scope of the paper

When we look at the expression of the estimators from (1.13) till (1.16), we see that one of the topics to deal with is the adequate estimation of (β, ρ) in order to get the tail index estimators $\widetilde{\widetilde{H}}_{\widehat\beta,\widehat\rho}(k)$ and $\overline{\overline{H}}_{\widehat\beta,\widehat\rho}(k)$. In section 2 of this paper, we shall thus briefly review the estimation of the two second order parameters β and ρ. Section 3 is devoted to the derivation of the asymptotic behaviour of the estimator $\widetilde{\widetilde{H}}_{\widehat\beta,\widehat\rho}(k)$ in (1.15) (equivalently, $\overline{\overline{H}}_{\widehat\beta,\widehat\rho}(k)$ in (1.16)), estimating β and ρ at a larger k value than the one used for the tail index estimation. We also do that only with the estimation of ρ, estimating β at the same level k used for the tail index estimation. In section 4, and through the use of simulation techniques, we shall exhibit the performance of the new estimators in (1.15) and (1.16), comparatively to the WH estimator in (1.13), to the classical Hill estimator and to $\widetilde{\widetilde{H}}_{\widehat\beta_{\widehat\rho}(k),\widehat\rho}(k)$, for the β-estimator, $\widehat\beta_{\widehat\rho}(k)$, in Gomes and Martins (2002). We have here considered only an external estimation of the second order parameter ρ. Such a decision is related to the discussion in Gomes and Martins (2002) on the advantages of an external estimation of the second order parameter ρ (or even their misspecification, as in Gomes and Martins (2004)) versus an internal estimation at the same level k. Indeed, the estimation of γ, β and ρ at the same level k leads to very volatile mean values and mean squared error patterns. Finally, in section 5, some overall conclusions are drawn.

2. SECOND ORDER PARAMETER ESTIMATION

2.1. The estimation of ρ

We shall first address the estimation of ρ. We have nowadays two general classes of ρ-estimators, which work well in practice, the ones introduced in Gomes *et al.* (2002) and Fraga Alves *et al.* (2003). We shall consider here particular members of the class of estimators of the second order parameter ρ proposed by Fraga Alves *et al.* (2003). Under adequate general conditions, they are semi-parametric asymptotically normal estimators of ρ, whenever $\rho < 0$. Moreover, for a large diversity of models, they give rise, for a wide range of large k-values, to highly stable sample paths, as functions of k, the number of top o.s.'s used. Such a class of estimators has been parameterised by a tuning parameter $\tau \geq 0$, and may be defined as

$$\widehat{\rho}_\tau(k) \;\equiv\; \widehat{\rho}_n^{(\tau)}(k) \;:=\; -\left|\frac{3\,\bigl(T_n^{(\tau)}(k)-1\bigr)}{T_n^{(\tau)}(k)-3}\right|\,, \tag{2.1}$$

where

$$T_n^{(\tau)}(k) := \begin{cases} \dfrac{\Bigl(M_n^{(1)}(k)\Bigr)^{\tau} - \Bigl(M_n^{(2)}(k)/2\Bigr)^{\tau/2}}{\Bigl(M_n^{(2)}(k)/2\Bigr)^{\tau/2} - \Bigl(M_n^{(3)}(k)/6\Bigr)^{\tau/3}} & \text{if } \tau > 0 \\[3ex] \dfrac{\ln\Bigl(M_n^{(1)}(k)\Bigr) - \frac{1}{2}\ln\Bigl(M_n^{(2)}(k)/2\Bigr)}{\frac{1}{2}\ln\Bigl(M_n^{(2)}(k)/2\Bigr) - \frac{1}{3}\ln\Bigl(M_n^{(3)}(k)/6\Bigr)} & \text{if } \tau = 0\,, \end{cases}$$

with

$$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}\left\{\ln\frac{X_{n-i+1:n}}{X_{n-k:n}}\right\}^j, \quad j \geq 1 \qquad \Bigl[M_n^{(1)} \equiv H \;\text{ in } (1.10)\Bigr]\,.$$

We shall here summarize a particular case of the results proved in Fraga Alves *et al.* (2003), now related to the asymptotic behaviour of the ρ-estimator in (2.1), under the second order framework in (1.3):

Proposition 2.1. *Under the second order framework in* (1.3), *with* $\rho < 0$, *if* (1.7) *holds, and if* $\sqrt{k}\,A(n/k) \to \infty$, *as* $n \to \infty$, *the statistic* $\widehat{\rho}_n^{(\tau)}(k)$ *in* (2.1) *converges in probability towards* ρ, *as* $n \to \infty$, *for any* $\tau \in \mathbb{R}$. *More than this:* $\widehat{\rho}_n^{(\tau)}(k) - \rho = O_p\bigl(1/\bigl(\sqrt{k}\,A(n/k)\bigr)\bigr)$, *provided that, under the third order framework in* (1.5), $\sqrt{k}\;A(n/k)\;B(n/k) \to \lambda_B$, *finite. If* $\sqrt{k}\;A(n/k)\;B(n/k) \to \infty$, *then* $\widehat{\rho}_n^{(\tau)}(k) - \rho = O_p\left(B(n/k)\right) = O_p\left(A(n/k)\right)$.

Remark 2.1. The theoretical and simulated results in Fraga Alves *et al.* (2003), together with the use of these estimators in the Generalized Jackknife statistics of Gomes *et al.* (2000), as done in Gomes and Martins (2002), as well as their use in the estimator in (1.13) (Gomes *et al.*, 2004*b*) and in the estimator in (1.14) (Gomes and Pestana, 2004), led us to advise in practice the consideration of the tuning parameters $\tau = 0$ for the region $\rho \in [-1, 0)$ and $\tau = 1$ for the region $\rho \in (-\infty, -1)$, together with the level $k_0 = \min(n-1, [2n/\ln\ln n])$. As done before, we however advise practitioners not to choose blindly the value of τ. It is sensible to draw a few sample paths of $\widehat{\rho}_\tau(k)$ in (2.1), as functions of k, electing the value of τ which provides higher stability for large k, by means of any stability criterion. For more details, see Gomes and Figueiredo (2003).

Remark 2.2. When we consider the level k_0 suggested in Remark 2.1, together with any of the ρ-estimators in this section, computed at that level k_0, $\{\widehat{\rho}(k_0) - \rho\}$ is of the order of $B(n/k_0) = (\ln\ln n)^\rho$, a very slow rate of convergence towards zero. We shall here work with a level

$$k_1 = \min\Big(n-1,\, \big[2n^{0.995}/\ln\ln n\big]\Big) \,. \tag{2.2}$$

Then $\widehat{\rho} - \rho = O_p((n^{0.005}\ln\ln n)^\rho)$ (provided that $\rho > -49.75$), and consequently, for any intermediate level k, $(\widehat{\rho} - \rho)\ln(n/k) = o_p(1)$, and $\sqrt{k}\,A(n/k)\,(\widehat{\rho} - \rho)\ln(n/k) = o_p(1)$ whenever $\sqrt{k}\,A(n/k) \to \lambda$, finite. This is going to be a fundamental result in the proof of Theorem 3.1, enabling the replacement of ρ by $\widehat{\rho}$, without disturbing the distributional result in Theorem 1.1, provided we estimate β adequately. In all the Monte Carlo simulations, we have considered the level k_1 in (2.2) and the following ρ-estimators in (2.1): $\widehat{\rho}_0 = \widehat{\rho}_0(k_1)$ if $\rho \ge -1$ and $\widehat{\rho}_1 = \widehat{\rho}_1(k_1)$ if $\rho < -1$.

2.2. Estimation of the second order parameter β

We have considered the estimator of β obtained in Gomes and Martins (2002) and based on the scaled log-spacings $U_i = i\,\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\}$ in (1.9), $1 \le i \le k$. Let us denote $\widehat{\rho}$ any of the estimators in (2.1) computed at the level k_1 in (2.2). The β-estimator is given by

$$\widehat{\beta}_{\widehat{\rho}}(k) := \left(\frac{k}{n}\right)^{\widehat{\rho}} \frac{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\right)\left(\frac{1}{k}\sum_{i=1}^{k} U_i\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}} U_i\right)}{\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}}\right)\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\widehat{\rho}} U_i\right) - \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-2\widehat{\rho}} U_i\right)} \,. \tag{2.3}$$

In Gomes and Martins (2002) and later in Gomes *et al.* (2004b), the following result has been proved:

Proposition 2.2. *If the second order condition* (1.3) *holds, with* $A(t) = \gamma\,\beta\,t^\rho$, $\rho < 0$, *if* $k = k_n$ *is a sequence of intermediate positive integers, i.e.* (1.7) *holds, and if* $\sqrt{k}\,A(n/k) \underset{n\to\infty}{\longrightarrow} \infty$, *then* $\widehat{\beta}_{\widehat{\rho}}(k)$ *in* (2.3) *is consistent for the estimation of* β, *whenever* $\widehat{\rho} - \rho = o_p(1/\ln n)$. *Moreover, if* ρ *is known,*

$$\widehat{\beta}_\rho(k) \overset{d}{=} \beta + \frac{\gamma\,\beta\,(1-\rho)\sqrt{1-2\rho}}{\rho\,\sqrt{k}\,A(n/k)}\,W_k^B + R_k^B\,, \qquad \text{with } R_k^B = o_p(1) \tag{2.4}$$

and W_k^B *asymptotically standard normal. More precisely we may write*

$$W_k^B = \frac{(1-\rho)\sqrt{1-2\rho}}{|\rho|}\left(\frac{Z_k^{(1)}}{1-\rho} - \frac{Z_k^{(1-\rho)}}{\sqrt{1-2\rho}}\right), \tag{2.5}$$

with $Z_k^{(\alpha)}$, $\alpha \geq 1$, *given in* (1.18).

The asymptotic distributional representation (2.4) *holds true as well for* $\widehat{\beta}_{\widehat{\rho}}(k)$, *with* $\widehat{\rho}$ *any of the consistent* ρ*-estimators in* (2.1) *computed at the level* k_1 *in* (2.2). *If* $\sqrt{k}\;A(n/k)\,R_k^B \to \lambda_R^B$, *finite, we may further guarantee the asymptotic normality of* $\widehat{\beta}_{\widehat{\rho}}(k)$. *If we consider* $\widehat{\beta}_{\widehat{\rho}(k)}(k)$, *then*

$$\widehat{\beta}_{\widehat{\rho}(k)}(k) - \beta \overset{p}{\sim} -\beta\,\ln(n/k)\left(\widehat{\rho}(k) - \rho\right)\,. \tag{2.6}$$

Remark 2.3. Note that when we consider the level k_1 in (2.2), the same restrictions for ρ as in Remark 2.2, and $\widehat{\beta} \equiv \widehat{\beta}_{\widehat{\rho}}(k_1)$, with $\widehat{\rho}$ any of the estimator in (2.1), computed also at the same level k_1, we may use (2.6) and derive that $\{\widehat{\beta} - \beta\}$ is of the order of $\ln(n/k_1)B(n/k_1) = O\left(\ln n\,\left(n^{0.005}\ln\ln n\right)^\rho\right)$. This result will also be needed in the proof of Theorem 3.1 and will enable us to keep the distributional result in Theorem 1.1.

Remark 2.4. Note also that if we estimate β through $\widehat{\beta}_{\widehat{\rho}}(k)$, since $\{\widehat{\beta}_{\widehat{\rho}}(k) - \beta\}$ is of the order of $1/\left(\sqrt{k}\;A(n/k)\right)$, we shall no longer be able to guarantee the distributional result in Theorem 1.1 (for details see Remark 3.2).

3. ASYMPTOTIC BEHAVIOUR OF THE ESTIMATORS

Let us assume first that we estimate both β and ρ externally at the level k_1 in (2.2). We may state the following:

Theorem 3.1. *Under the conditions of Theorem 1.1, let us consider the tail index estimators* $\overline{\widetilde{H}}_{\widehat{\beta},\widehat{\rho}}(k)$ *and* $\overline{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k)$ *in* (1.15) *and* (1.16)*, respectively, for any of the estimators* $\widehat{\rho}$ *and* $\widehat{\beta}$ *in* (2.1) *and* (2.3)*, respectively, both computed at the level* k_1 *in* (2.2) *and such that* $\widehat{\rho}-\rho = o_p(1/\ln n)$*. Then,* $\sqrt{k}\left\{\overline{\widetilde{H}}_{\widehat{\beta},\widehat{\rho}}(k)-\gamma\right\}$ *as well as* $\sqrt{k}\left\{\overline{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k)-\gamma\right\}$ *are asymptotically normal with variance equal to* γ^2 *and null mean value, not only when* $\sqrt{k}\ A(n/k) \to 0$*, but also whenever* $\sqrt{k}\ A(n/k)\to\lambda$*, finite, as* $n\to\infty$*.*

Proof: If we estimate consistently β and ρ through the estimators $\widehat{\beta}$ and $\widehat{\rho}$ in the conditions of the theorem, we may use Taylor's expansion series, and write,

$$
\begin{aligned}
\overline{\widetilde{H}}_{\widehat{\beta},\widehat{\rho}}(k) &\stackrel{d}{=} H(k)\times\left(1-\frac{\beta}{1-\rho}\left(\frac{n}{k}\right)^{\rho}-\left(\widehat{\beta}-\beta\right)\frac{1}{1-\rho}\left(\frac{n}{k}\right)^{\rho}\left(1+o_p(1)\right)\right.\\
&\qquad\left.-\frac{\beta}{1-\rho}\left(\widehat{\rho}-\rho\right)\left(\frac{n}{k}\right)^{\rho}\left(\frac{1}{1-\rho}+\ln(n/k)\right)\left(1+o_p(1)\right)\right)\\
&\stackrel{d}{=} \overline{\widetilde{H}}_{\beta,\rho}(k)-\frac{A(n/k)}{1-\rho}\left(\frac{\widehat{\beta}-\beta}{\beta}+\left(\widehat{\rho}-\rho\right)\ \ln(n/k)\right)\left(1+o_p(1)\right)\ .
\end{aligned}
$$

Since $\widehat{\beta}$ and $\widehat{\rho}$ are consistent for the estimation of β and ρ, respectively, and $(\widehat{\rho}-\rho)\ln(n/k) = o_p(1)$ (see Remark 2.2), the summands related to $(\widehat{\beta}-\beta)$ and $(\widehat{\rho}-\rho)$ are both $o_p(A(n/k))$, and the result in the theorem, related to the $\overline{\widetilde{H}}$-estimator, follows immediately, provided that $\sqrt{k}\ A(n/k)\to\lambda$, finite. The reasoning is exactly the same for the $\overline{\overline{H}}$-estimator. □

Remark 3.1. Note however that the levels k such that $\sqrt{k}\ A(n/k)\to\lambda$, finite, are sub-optimal for this type of estimators.

If we consider γ and β estimated at the same level, we are going to have an increase in the variance of our final tail index estimator $\overline{\widetilde{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}(k)$ (or equivalently, $\overline{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}(k)$). Similarly to Corollary 2.1 of Theorem 2.1 in Gomes and Martins (2002), there in connection with a ML-tail index estimator, as well as in Theorem 3.2 in Gomes *et al.* (2004b), in connection with the tail index estimator in (1.13), we may also get:

Theorem 3.2. *If the second order condition* (1.3) *holds, if* $k=k_n$ *is a sequence of intermediate integers, i.e.,* (1.7) *holds, and if* $\sqrt{k}\ A(n/k)\underset{n\to\infty}{\longrightarrow}\lambda$*, finite, non necessarily null, then*

$$
\sqrt{k}\left(\overline{\widetilde{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}(k)-\gamma\right)\ \underset{n\to\infty}{\stackrel{d}{\longrightarrow}}\ Normal\left(0,\ \sigma^2_{H_2}:=\gamma^2\left(\frac{1-\rho}{\rho}\right)^2\right), \tag{3.1}
$$

i.e., the asymptotic variance of $\widetilde{\overline{H}}_{\widehat\beta_{\widehat\rho}(k),\,\widehat\rho}(k)$ *increases of a factor* $((1-\rho)/\rho)^2$, *greater than one, for every* $\rho \le 0$. *The same result holds obviously true for* $\overline{\overline{H}}_{\widehat\beta_{\widehat\rho}(k),\,\widehat\rho}(k)$.

Proof: If we consider

$$\widetilde{\overline{H}}_{\widehat\beta_{\widehat\rho}(k),\,\widehat\rho}(k) \;:=\; H(k)\left(1 - \frac{\widehat\beta_{\widehat\rho}(k)}{1-\widehat\rho}\left(\frac{n}{k}\right)^{\widehat\rho}\right) ,$$

we now get

$$\widetilde{\overline{H}}_{\widehat\beta_{\widehat\rho}(k),\,\widehat\rho}(k) \;=\; \widetilde{\overline{H}}_{\beta,\,\rho}(k) - \frac{A(n/k)}{1-\rho}\left(\frac{\widehat\beta_{\widehat\rho}(k)-\beta}{\beta} + (\widehat\rho-\rho)\ln(n/k)\right)\big(1+o_p(1)\big) .$$

Since $\big(\widehat\beta_{\widehat\rho}(k)-\beta\big)/\beta$ is now of the order of $1/\big(\sqrt{k}\,A(n/k)\big)$, the term of the order of $1/\sqrt{k}$ is going to be, from (1.20), (2.4) and (2.5),

$$\frac{\gamma}{\sqrt{k}}\left(Z_k^{(1)} + \frac{(1-\rho)\,(1-2\rho)}{\rho^2}\left(\frac{Z_k^{(1)}}{1-\rho} - \frac{Z_k^{(1-\rho)}}{\sqrt{1-2\rho}}\right)\right) ,$$

which may be written as

$$\frac{\gamma}{\sqrt{k}}\left(\left(\frac{1-\rho}{\rho}\right)^2 Z_k^{(1)} - \frac{(1-\rho)\,\sqrt{1-2\rho}}{\rho^2}\; Z_k^{(1-\rho)}\right) ,$$

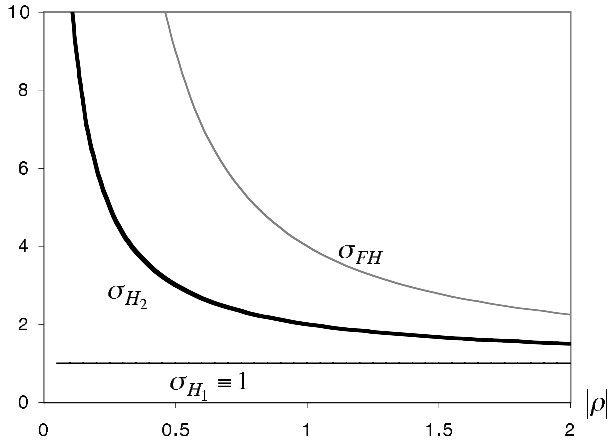
with $Z_k^{(\alpha)}$ the asymptotically standard normal r.v. in (1.18). Taking into account the fact that from (1.19), the asymptotic covariance between $Z_k^{(1)}$ and $Z_k^{(1-\rho)}$ is given by $\sqrt{1-2\rho}\,/\,(1-\rho)$, together with the fact that $\sqrt{k}\;A(n/k)\,(\widehat\rho-\rho)\ln(n/k) \to 0$ (see Remark 2.2), (3.1) follows. □

Remark 3.2. If we compare Theorems 3.1 and 3.2 we see that the estimation of the two parameters γ and β at the same level k induces an increase in the asymptotic variance of the final γ-estimator of a factor given by $((1-\rho)/\rho)^2$, greater than 1 for all $\rho \le 0$. As may be seen in Gomes and Martins (2002) the asymptotic variance of the estimator in Feuerverger and Hall (1999) (where the three parameters are computed at the same level k) is given by

$$\sigma^2_{FH} \;:=\; \gamma^2\left(\frac{1-\rho}{\rho}\right)^4 .$$

In Figure 2 we provide both a picture and some values of $\sigma_{H_1}/\gamma \equiv 1$, σ_{H_2}/γ and σ_{FH}/γ, as functions of $|\rho|$.

It is obvious from Figure 2 that, whenever possible, it seems convenient to estimate both β and ρ "externally", at a k-value higher than the one used for the estimation of the tail index γ.



$\lvert\rho\rvert$	σ_{H_1}	σ_{H_2}	σ_{FH}
0.1	1.00	11.00	121.00
0.2	1.00	6.00	36.00
0.3	1.00	4.33	18.78
0.4	1.00	3.50	12.25
0.5	1.00	3.00	9.00
1.0	1.00	2.00	4.00
1.5	1.00	1.67	2.78
2.0	1.00	1.50	2.25
2.5	1.00	1.40	1.96
3.0	1.00	1.33	1.78

Figure 2: "Rulers" of the asymptotic standard deviations, σ_{H_1} and σ_{H_2} of the estimators under study, together with σ_{FH}, for $\gamma = 1$.

Remark 3.3. More generally, to obtain information on the asymptotic bias of $\widetilde{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k)$, $\widetilde{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}(k)$ and $\widetilde{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}(k)}(k),\widehat{\rho}(k)}(k)$ — or equivalently, of $\overline{\overline{H}}_{\widehat{\beta},\widehat{\rho}}(k)$, $\overline{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}(k)$ and $\overline{\overline{H}}_{\widehat{\beta}_{\widehat{\rho}(k)}(k),\widehat{\rho}(k)}(k)$ — we need to go further into a third order framework, specifying, like has been done in (1.5), the rate of convergence in the second order condition in (1.3). This is however beyond the scope of this paper.

4. FINITE SAMPLE BEHAVIOUR OF THE ESTIMATORS

4.1. Underlying models

In this section we shall consider the following models in the class (1.6):

- the *Fréchet* model, with distribution function (d.f.) $F(x) = \exp(-x^{-1/\gamma})$, $x \geq 0$, $\gamma > 0$, for which $\rho = -1$;
- the *Generalized Pareto* (*GP*) model, with d.f. $F(x) = 1 - (1+\gamma x)^{-1/\gamma}$, $x \geq 0$, $\gamma > 0$, for which $\rho = -\gamma$;
- the *Burr* model, with d.f. $F(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho}$, $x \geq 0$, $\gamma > 0$, $\rho < 0$;
- the *Student*'s t_ν-model with ν degrees of freedom, with a probability density function (*p.d.f.*)
$$f_{t_\nu}(t) = \frac{\Gamma\big((\nu+1)/2\big)}{\sqrt{\pi\nu}\;\Gamma(\nu/2)} \left[1 + \frac{t^2}{\nu}\right]^{-(\nu+1)/2}, \qquad t \in \mathbb{R} \quad (\nu > 0)~,$$
for which $\gamma = 1/\nu$ and $\rho = -2/\nu$.

4.2. The simulation design

We have here implemented multi-sample simulation experiments of size $50{,}000 = 5{,}000(runs) \times 10(replicates)$, in order to obtain, for the above mentioned models, the distributional behaviour of the new estimators $\widetilde{\overline{H}}_{\widehat{\beta},\widehat{\rho}}$ and $\overline{\overline{H}}_{\widehat{\beta},\widehat{\rho}}$ in (1.15) and (1.16), respectively, based on the estimation of β at the level k_1 in (2.2), the same level we have used for the estimation of ρ, again not chosen in an optimal way. For details on multi-sample simulation, see Gomes and Oliveira (2001). We use the notation $\widehat{\beta}_{j1} = \beta_{\widehat{\rho}_j}(k_1)$, $j = 0, 1$, with $\widehat{\rho}_j$, $j = 0, 1$ and $\beta_{\widehat{\rho}}(k)$ given in (2.1) and (2.3), respectively. Similarly to what has been done in Gomes *et al.* (2004b) for the WH-estimator in (1.13), these estimators of ρ and β, for $j = 0, 1$, have been incorporated in the estimators under study, leading to $\widetilde{\overline{H}}_{\widehat{\beta}_{01},\widehat{\rho}_0}(k)\,/\,\overline{\overline{H}}_{\widehat{\beta}_{01},\widehat{\rho}_0}(k)$ and to $\widetilde{\overline{H}}_{\widehat{\beta}_{11},\widehat{\rho}_1}(k)\,/\,\overline{\overline{H}}_{\widehat{\beta}_{11},\widehat{\rho}_1}(k)$, respectively. The simulations show that the tail index estimators $\widetilde{\overline{H}}_{\widehat{\beta}_{j1},\widehat{\rho}_j}(k)$ and $\overline{\overline{H}}_{\widehat{\beta}_{j1},\widehat{\rho}_j}(k)$, j equal to either 0 or 1, according as $|\rho| \le 1$ or $|\rho| > 1$ seem to work reasonably well, as illustrated in the sequel. In the simulation we have also included the Hill estimator in (1.10), the "Weighted Hill" estimator in (1.13) and $\widetilde{\overline{H}}_{\beta_{\widehat{\rho}}(k),\widehat{\rho}}$. The estimator $\overline{H}$ in (1.14) exhibits a behaviour quite similar to that of $\overline{\overline{H}}$ in (1.16), as may be seen from the results in Gomes and Pestana (2004), and was not pictured, for sake of simplicity.

We have simulated four different indicators. Let us denote generically $\widetilde{H}_n$ any of the estimators in (1.13), (1.15) and (1.16), and let

$$k_{0s}^H(n) \;:=\; \arg\min_k MSE_s\big[H_n(k)\big]$$

be the simulated optimal k (in the sense of minimum simulated mean squared error) for the Hill estimator $H_n(k) \equiv H(k)$ in (1.10). The two first indicators are related to the behaviour of the new estimators at Hill's optimal simulated level, i.e.,

$$REFF_n^{\widetilde{H}|H_0} \;:=\; \sqrt{\frac{MSE_s\Big[H_n\big(k_{0s}^H(n)\big)\Big]}{MSE_s\Big[\widetilde{H}_n\big(k_{0s}^H(n)\big)\Big]}} \tag{4.1}$$

and

$$BRI_n^{\widetilde{H}|H_0} \;:=\; \left|\frac{E_s\Big[H_n\big(k_{0s}^H(n)\big)-\gamma\Big]}{E_s\Big[\widetilde{H}_n\big(k_{0s}^H(n)\big)-\gamma\Big]}\right| . \tag{4.2}$$

The two additional indicators are related to the comparison of mean squared errors and bias of the new estimators with those of the Hill's estimators, when all the estimators are considered at their optimal levels. Denoting

$$H_{n0} := H_n\big(k_0^H(n)\big) \qquad \text{and} \qquad \widetilde{H}_{n0} := \widetilde{H}_n\big(k_0^{\widetilde{H}}(n)\big)\ ,$$

with the obvious meaning for $k_0^\bullet(n)$, the two extra simulated indicators are

$$(4.3) \qquad REFF_{n0}^{\widetilde{H}|H} := \sqrt{\frac{MSE_s\left[H_{n0}\right]}{MSE_s\left[\widetilde{H}_{n0}\right]}}$$

and

$$(4.4) \qquad BRI_{n0}^{\widetilde{H}|H} := \left|\frac{E_s\left[H_{n0}-\gamma\right]}{E_s\left[\widetilde{H}_{n0}-\gamma\right]}\right| .$$

Remark 4.1. Note that an indicator higher than one means a better performance than the Hill estimator. Consequently, the higher these indicators are, the better the new estimators perform, comparatively to the Hill estimator.

Remark 4.2. Note also that whereas we have appropriate techniques to deal with the estimation of the optimal level for Hill's estimator, in the sense of minimum mean squared error, we do not have yet equivalent techniques for the reduced bias' estimators. Consequently, the indicators in (4.3) and (4.4) are not useful in practice, but they give us an indication of the potentialities of this type of estimators.

4.3. Mean values and mean squared error patterns

In Figures from 3 till 9, and on the basis of the first replicate, with 5000 runs, we picture for the different underlying models considered, and a sample of size $n = 1000$, the mean values ($E[\bullet]$) and the mean squared errors ($MSE[\bullet]$) of the Hill estimator H, together with $\widetilde{\overline{H}}_{\widehat{\beta}_{j1},\widehat{\rho}_j}$, $\overline{\overline{H}}_{\widehat{\beta}_{j1},\widehat{\rho}_j}$, $WH_{\widehat{\beta}_{j1},\widehat{\rho}_j}$ and $\widetilde{\overline{H}}_{\widehat{\beta}_j(k),\widehat{\rho}_j}$, $j = 0$ or $j = 1$, according as $|\rho| \le 1$ or $|\rho| > 1$. For comparison, we also picture the analogue behaviour of the r.v. $\widetilde{\overline{H}}_{\beta,\rho}$ for the models where there is a big discrepancy between the behaviour of the estimators and that of the r.v.'s. Such a discrepancy suggests that some improvement in the estimation of second order parameters β and ρ is still welcome.

Remark 4.3. For a Burr model and for any of the estimators considered, $BIAS/\gamma$ and MSE/γ^2 are independent of γ, for every ρ. And we have seen no reason to picture the mean value and mean squared error patterns of the estimators for a GP underlying model because for all the estimators considered,

$$\mathbb{E}\,|\,GP(\gamma) \;=\; \gamma \times \mathbb{E}\,|\,BURR(\gamma=1,\ \rho=-\gamma)$$

and

$$MSE\,|\,GP(\gamma) \;=\; \gamma^2 \times MSE\,|\,BURR(\gamma,\rho=-\gamma) .$$

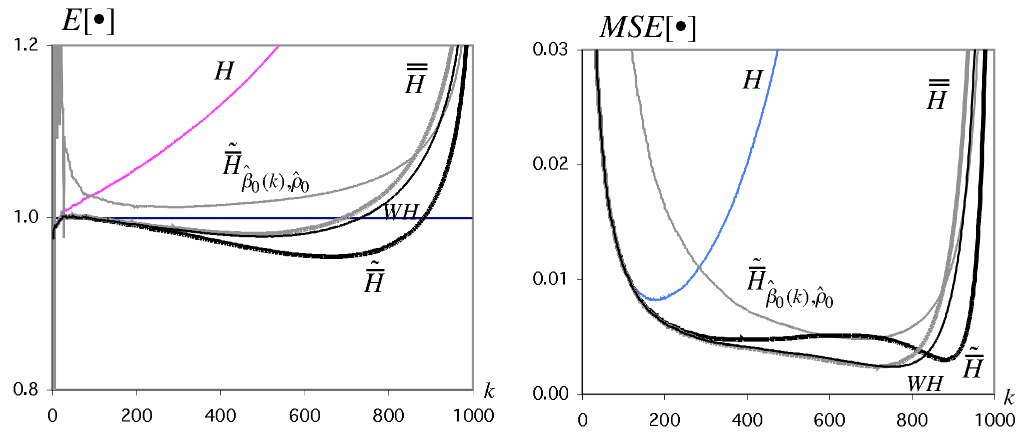


Figure 3: Underlying *Fréchet* parent with $\gamma = 1$ ($\rho = -1$).

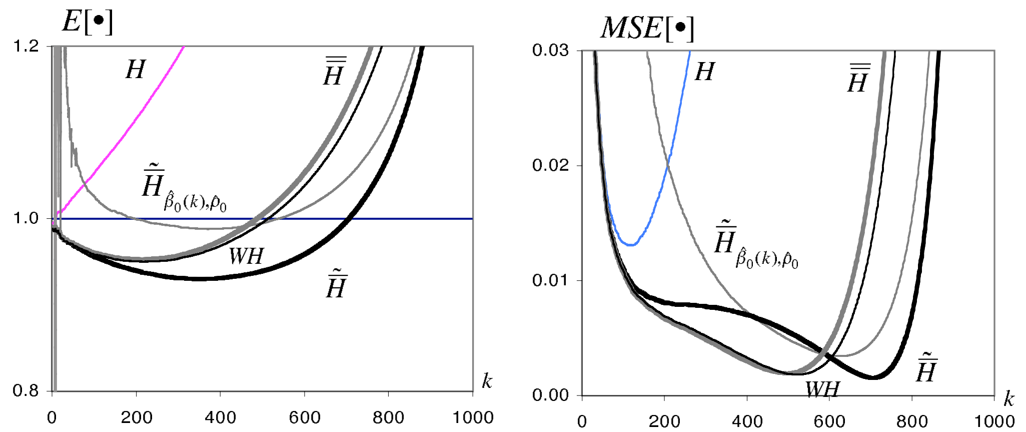


Figure 4: Underlying *Burr* parent with $\gamma = 1$ and $\rho = -1$.

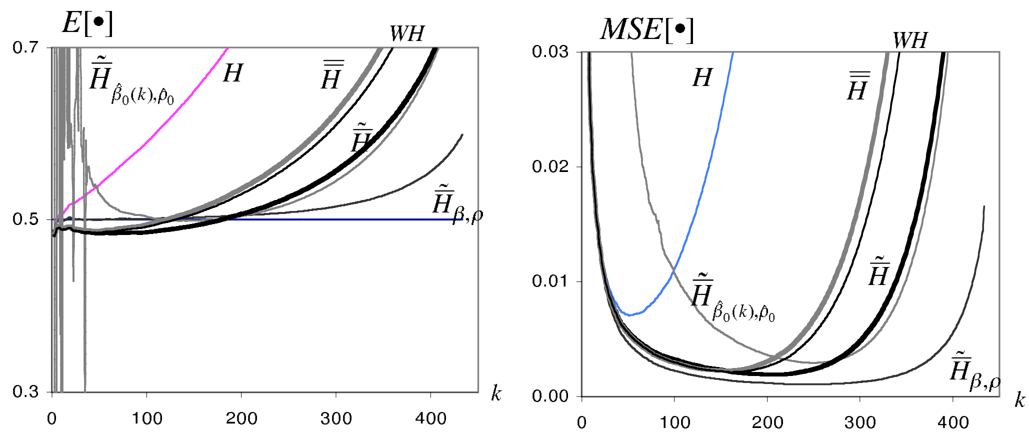


Figure 5: Underlying *Student* parent with $\nu = 2$ degrees of freedom ($\gamma = 0.5$ and $\rho = -1$).

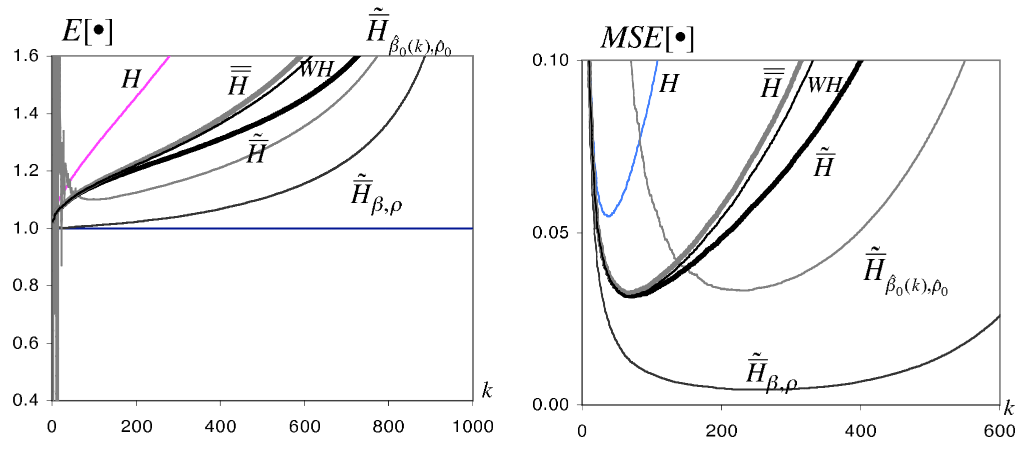


Figure 6: Underlying *Burr* parent with $\gamma = 1$ and $\rho = -0.5$.

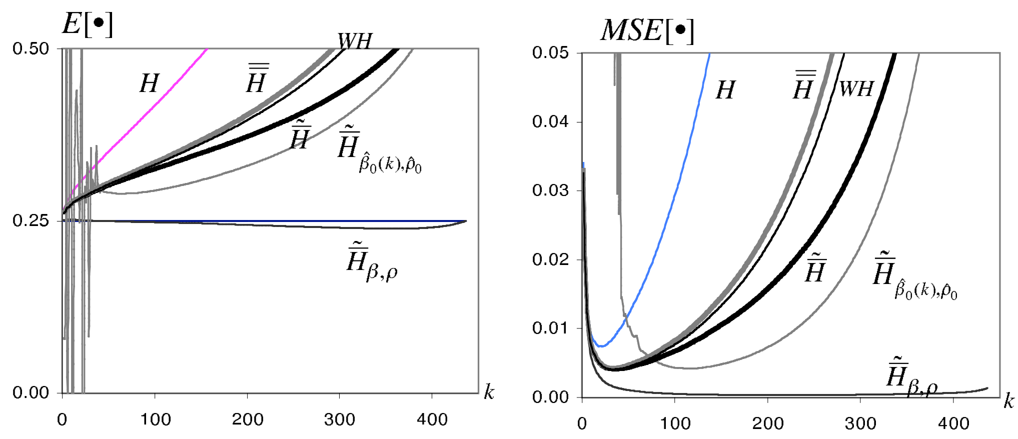


Figure 7: Underlying *Student* parent with $\nu = 4$ degrees of freedom ($\gamma = 0.25$ and $\rho = -0.5$).

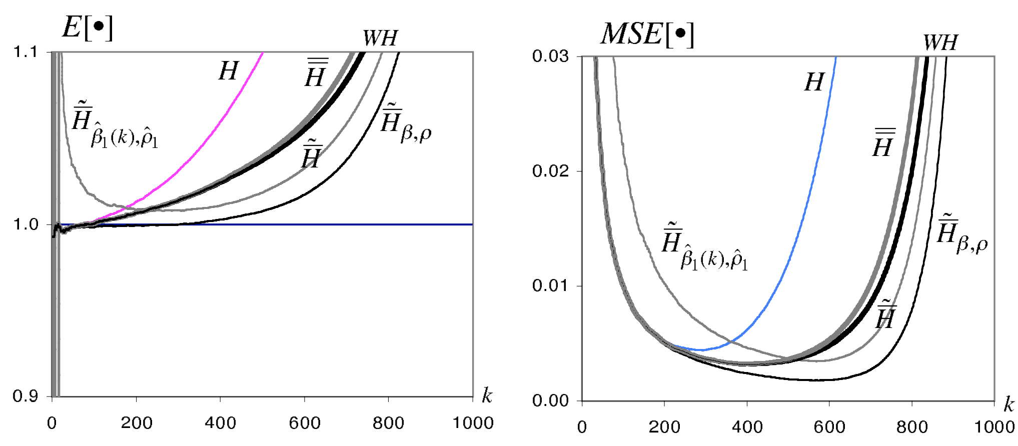


Figure 8: Underlying *Burr* parent with $\gamma = 1$ and $\rho = -2$.

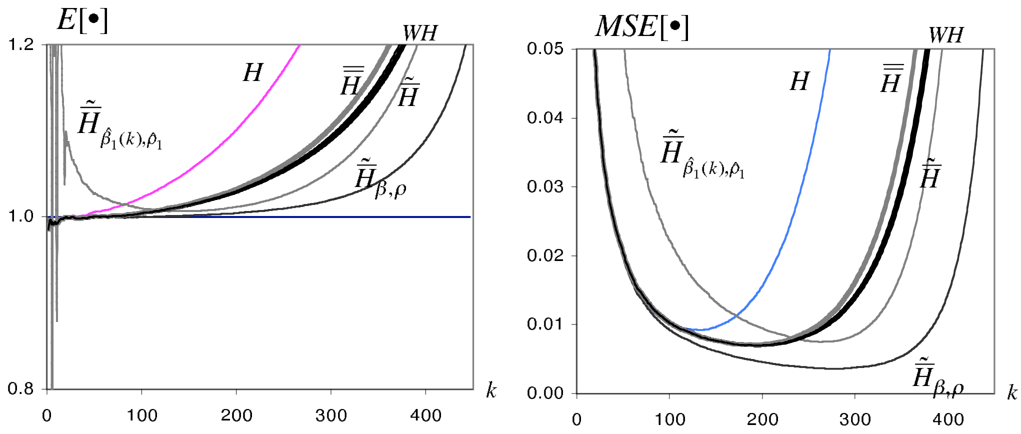


Figure 9: Underlying *Student* parent with with $\nu = 1$ degrees of freedom ($\gamma = 1$ and $\rho = -2$).

Remark 4.4. We may further draw the following specific comments:

- For a Fréchet model (with $\rho = -1$) (Figure 3) the bias of $\overline{\overline{H}}$ is the smallest one, being the one of $\tilde{\tilde{H}}$ the largest one, for small to moderate values of k. All the three reduced bias' statistics overestimate the true value of γ for very small as well as very large values of k, whereas, for moderate values of k, they all underestimate γ.

- For an underlying Burr parent with $\rho = -1$ (Figure 4), the three reduced bias' statistics are negatively biased for small values of k. Again, as for a Fréchet underlying parent, the $\overline{\overline{H}}$-statistic exhibits then the smallest bias and $\tilde{\tilde{H}}$ the largest one. The $\tilde{\tilde{H}}$ statistic is the best one regarding *MSE* at the optimal level, but the WH-statistic is the one with the smallest mean squared error for not too large values of k, followed by $\overline{\overline{H}}$. Quite similar results may be drawn for a Student model with $\rho = -1$ (Figure 5), but for this model, the mean squared error of $\tilde{\tilde{H}}$ is smaller than that of WH, which on its turn is smaller than that of $\overline{\overline{H}}$, for all values of k.

- For values of $\rho > -1$ (Figures 6 and 7), the three reduced bias' statistics are positively biased for all k. The $\tilde{\tilde{H}}$-statistic is better than the WH-statistic, which on its turn behaves slightly better than the $\overline{\overline{H}}$-statistic, both regarding bias and mean squared error.

- For $\rho < -1$ (Figures 8 and 9), we need to use $\hat{\rho}_1$ (instead of $\hat{\rho}_0$). In all the simulated cases the $\tilde{\tilde{H}}$ and the WH-statistics are the best ones and exhibit quite similar properties, but they are not a long way from the $\overline{\overline{H}}$-statistic.

Remark 4.5. For a Student model with ν degrees of freedom (Figures 5, 7 and 9), and whenever we assume β and ρ known, the most stable sample path around the target value γ is achieved by the statistic $\widetilde{\widetilde{H}}_{\beta,\rho}$, presented in the figures. And such a fact leads this statistic to have the smallest mean squared error, followed by the $\overline{\overline{H}}$ and next the WH statistics, for all values of ν. If we need to estimate β and ρ, the $\widetilde{\widetilde{H}}$-statistic is the one with the smallest mean squared error at the optimal level, also for every ν. Next comes the WH-statistic, quite close to the $\overline{\overline{H}}$-statistic when $\nu < 2$, i.e., when $\rho < -1$.

4.4. Relative efficiency and bias reduction indicators

In Table 1 we present the *REFF* indicators, in (4.1) and (4.3), and the *BRI* indicators in (4.2) and (4.4). For each model, the first, second and third rows are related to the WH-estimator in (1.13), the $\widetilde{\widetilde{H}}$-estimator in (1.15) and the $\overline{\overline{H}}$-estimator in (1.16), respectively. Each entry has two numbers: the first one is either the indicator in (4.1) or in (4.2) and the second one is either the indicator in (4.3) or in (4.4), according as we refer to the *REFF*-indicators (left hand-side table) or to the *BRI*-indicators (right hand-side table).

5. OVERALL CONCLUSIONS

- Generally, we may say that there is not a big difference between the estimators, WH, $\overline{H}$, $\widetilde{\widetilde{H}}$ and $\overline{\overline{H}}$ in (1.13), (1.14) (1.15) and (1.16), respectively. Anyway, whenever confronted with real data, the drawing of a few sample paths may help us in the choice of the most adequate estimate of the tail index γ.

- The $\widetilde{\widetilde{H}}_{\widehat{\beta}_{\widehat{\rho}}(k),\widehat{\rho}}$ statistic may perhaps help us in the choice of the optimal sample fraction of Hill's estimator, and for some of the models exhibits sample paths more stable around the target value γ for a wider region of k-values. This is however a topic which deserves further investigation, being outside the scope of the present paper.

- The main advantage of these estimators lies on the fact that we may estimate β and ρ adequately through $\widehat{\beta}$ and $\widehat{\rho}$ so that the *MSE* of the new estimator is smaller than the *MSE* of Hill's estimator for all k, even when $|\rho| > 1$, a region where it has been difficult to find alternatives for the Hill estimator. And this happens together with a higher stability of the sample paths around the target value γ.

Table 1: *REFF* and *BRI* indicators.

***REFF* indicators**			***BRI* indicators**		
n			n		
200	500	1000	200	500	1000
Fréchet parent: $\rho = -\mathbf{1}$, $\gamma = \mathbf{1}$					
1.07/1.53	1.11/1.69	1.12/1.86	4.56/25.40	7.16/13.12	11.50/10.22
1.06/1.39	1.10/1.53	1.12/1.67	4.18/45.93	6.33/31.87	9.99/69.15
1.08/1.52	1.12/1.67	1.12/1.85	5.85/5.56	8.92/47.70	14.28/22.89
Burr parent: $\rho = -\mathbf{0.5}$, $\gamma = \mathbf{1}$					
1.20/1.39	1.26/1.35	1.23/1.33	1.83/1.32	1.76/1.23	1.67/1.24
1.19/1.39	1.26/1.35	1.22/1.33	1.73/1.25	1.73/1.22	1.64/1.24
1.18/1.35	1.25/1.32	1.22/1.31	1.69/1.30	1.69/1.20	1.62/1.22
Burr parent: $\rho = -\mathbf{1}$, $\gamma = \mathbf{1}$					
1.19/2.09	1.23/2.42	1.23/2.69	2.25/11.26	1.91/10.31	1.70/30.21
1.18/2.27	1.22/2.63	1.21/2.94	2.14/306.29	1.74/22.55	1.55/28.80
1.20/2.02	1.24/2.34	1.24/2.61	2.67/9.35	2.07/24.51	1.79/13.63
Burr parent: $\rho = -\mathbf{2}$, $\gamma = \mathbf{1}$					
1.05/1.18	1.08/1.18	1.11/1.21	2.75/1.08	2.29/1.12	2.09/1.10
1.05/1.17	1.08/1.18	1.10/1.21	2.60/1.09	2.25/1.10	2.07/1.09
1.05/1.16	1.08/1.17	1.10/1.20	2.54/1.10	2.22/1.10	2.05/1.06
Student parent: $\rho = -\mathbf{0.5}$, $\gamma = \mathbf{0.25}$					
1.35/1.42	1.25/1.35	1.22/1.32	2.27/1.54	1.90/1.47	1.73/1.30
1.32/1.41	1.24/1.34	1.21/1.32	2.17/1.49	1.87/1.39	1.68/1.26
1.30/1.36	1.23/1.31	1.20/1.30	1.99/1.41	1.80/1.42	1.64/1.26
Student parent: $\rho = -\mathbf{1}$, $\gamma = \mathbf{0.5}$					
1.07/1.56	1.02/1.70	1.15/1.86	3.17/3.40	1.68/3.94	1.25/4.85
1.03/1.51	1.02/1.75	1.14/1.97	2.77/7.88	2.03/7.15	1.25/8.71
1.08/1.51	1.02/1.65	1.16/1.82	4.58/2.84	2.26/3.31	1.24/4.65
Student parent: $\rho = -\mathbf{2}$, $\gamma = \mathbf{1}$					
0.87/1.18	1.07/1.16	1.04/1.16	2.00/7.00	13.80/1.67	3.81/1.48
0.91/1.12	1.06/1.16	1.04/1.15	1.71/15.65	11.52/1.71	3.64/1.47
0.88/1.18	1.07/1.15	1.04/1.15	2.43/8.52	8.83/1.51	3.50/1.45

REFERENCES

[1] Beirlant, J.; Dierckx, G.; Goegebeur, Y. and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes*, **2**, 177–200.

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

[3] 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.

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

[5] Gomes, M.I.; Caeiro, F. and Figueiredo, F. (2004a). Bias reduction of a tail index estimator through an external estimation of the second order parameter. *Statistics*, **38**:6, 497–510.

[6] Gomes, M.I. and Figueiredo, F. (2003). Bias reduction in risk modelling: semi-parametric quantile estimation. Accepted at *Test*.

[7] Gomes, M.I.; Figueiredo, F. and Mendonça, S. (2005). Asymptotically best linear unbiased tail estimators under a second order regular variation condition. *J. Statist. Planning and Inference*, **134**:2, 409–433.

[8] Gomes, M.I.; Haan, L. de and Henriques, L. (2004b). *Tail index estimation through accommodation of bias in the weighted log-excesses*. Notas e Comunicações C.E.A.U.L. 14/2004. *Submitted*.

[9] 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.

[10] 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.

[11] Gomes, M.I. and Martins, M.J. (2004). Bias reduction and explicit estimation of the tail index. *J. Statist. Planning and Inference*, **124**, 361–378.

[12] Gomes, M.I.; Martins, M.J. and Neves, M. (2000). Alternatives to a semi-parametric estimator of parameters of rare events — the Jackknife methodology. *Extremes*, **3**:3, 207–229.

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

[14] Gomes, M.I. and Pestana, D. (2004). A simple second order reduced bias' tail index estimator. Accepted at *J. Statist. Comp. and Simulation*.

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

[16] Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. Royal Statist. Soc.*, **B44**, 37–42.

[17] Hall, P. and Welsh, A.H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.*, **13**, 331–341.

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

[19] Rényi, A. (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hung.*, **4**, 191–231.