REVSTAT – Statistical Journal
Volume 5, Number 2, June 2007, 177–207

IMPROVING SECOND ORDER REDUCED BIAS EXTREME VALUE INDEX ESTIMATION

Authors: M. Ivette Gomes
– Universidade de Lisboa, D.E.I.O. and C.E.A.U.L., Portugal
ivette.gomes@fc.ul.pt

M. João Martins
– D.M., I.S.A., Universidade Técnica de Lisboa, Portugal
mjmartins@isa.utl.pt

Manuela Neves
– D.M., I.S.A., Universidade Técnica de Lisboa, Portugal
manela@isa.utl.pt

Received: November 2006 Accepted: February 2007

Abstract:

- Classical extreme value index estimators are known to be quite sensitive to the number k of top order statistics used in the estimation. The recently developed second order reduced-bias estimators show much less sensitivity to changes in k. Here, we are interested in the improvement of the performance of reduced-bias extreme value index estimators based on an exponential second order regression model applied to the scaled log-spacings of the top k order statistics. In order to achieve that improvement, the estimation of a "scale" and a "shape" second order parameters in the bias is performed at a level k_1 of a larger order than that of the level k at which we compute the extreme value index estimators. This enables us to keep the asymptotic variance of the new estimators of a positive extreme value index γ equal to the asymptotic variance of the Hill estimator, the maximum likelihood estimator of γ, under a strict Pareto model. These new estimators are then alternatives to the classical estimators, not only around optimal and/or large levels k, but for other levels too. To enhance the interesting performance of this type of estimators, we also consider the estimation of the "scale" second order parameter only, at the same level k used for the extreme value index estimation. The asymptotic distributional properties of the proposed class of γ-estimators are derived and the estimators are compared with other similar alternative estimators of γ recently introduced in the literature, not only asymptotically, but also for finite samples through Monte Carlo techniques. Case-studies in the fields of finance and insurance will illustrate the performance of the new second order reduced-bias extreme value index estimators.

Key-Words:

- *statistics of extremes; semi-parametric estimation; bias estimation; heavy tails; maximum likelihood.*

AMS Subject Classification:

- 62G32, 62H12; 65C05.

1. INTRODUCTION AND MOTIVATION FOR THE NEW CLASS OF EXTREME VALUE INDEX ESTIMATORS

Examples of heavy-tailed models are quite common in the most diversified fields. We may find them in computer science, telecommunication networks, insurance, economics and finance, among other areas of application. In the area of *extreme value theory*, a model F is said to be *heavy-tailed* whenever the *tail function*, $\overline{F} := 1-F$, is a regularly varying function with a negative index of regular variation equal to $-1/\gamma$, $\gamma>0$, denoted $\overline{F} \in RV_{-1/\gamma}$, where the notation RV_α stands for the class of *regularly varying* functions at infinity with an *index of regular variation* equal to α, i.e., positive measurable functions g such that $\lim_{t\to\infty} g(tx)/g(t) = x^\alpha$, for all $x>0$. Equivalently, the quantile function $U(t) = F^{\leftarrow}(1-1/t),\ t\geq 1$, with $F^{\leftarrow}(x) = \inf\{y\colon F(y)\geq x\}$, is of regular variation with index γ, i.e.,

$$\text{(1.1)} \qquad F \text{ is heavy-tailed} \quad\Longleftrightarrow\quad \overline{F}\in RV_{-1/\gamma} \quad\Longleftrightarrow\quad U\in RV_\gamma\ ,$$

for some $\gamma>0$. Then, we are in the domain of attraction for maxima of an *extreme value* distribution function (d.f.),

$$EV_\gamma(x) \;=\; \begin{cases} \exp\bigl(-(1+\gamma x)^{-1/\gamma}\bigr), & 1+\gamma x\geq 0 \qquad \text{if } \gamma\neq 0\,, \\ \exp\bigl(-\exp(-x)\bigr), & x\in\mathbb{R} \qquad\qquad\ \ \text{if } \gamma=0\ , \end{cases}$$

but with $\gamma>0$, and we write $F\in\mathcal{D}_{\mathcal{M}}(EV_{\gamma>0})$. The parameter γ is the *extreme value index*, one of the primary parameters of extreme or even rare events.

The *second order parameter* ρ rules the rate of convergence in the first order condition (1.1), let us say the rate of convergence towards zero of $\ln U(tx) - \ln U(t) - \gamma\ln x$, and is the non-positive parameter appearing in the limiting relation

$$\text{(1.2)} \qquad \lim_{t\to\infty} \frac{\ln U(tx)-\ln U(t)-\gamma\ln x}{A(t)} \;=\; \frac{x^\rho-1}{\rho}\ ,$$

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 $\rho<0$. The second order condition (1.2) has been widely accepted as an appropriate condition to specify the tail of a Pareto-type distribution in a semi-parametric way, and it holds for most common Pareto-type models.

Remark 1.1. For Hall–Welsh class of Pareto-type models (Hall and Welsh, 1985), i.e., models such that, with $C>0$, $D_1\neq 0$ and $\rho<0$,

$$\text{(1.3)} \qquad U(t) \;=\; C\,t^\gamma\bigl(1+D_1 t^\rho + o(t^\rho)\bigr), \qquad \text{as}\ \ t\to\infty\ ,$$

condition (1.2) holds and we may choose $A(t) = \rho\, D_1 t^\rho$.

Here, although not going into a general third order framework, as the one found in Gomes et al. (2002) and Fraga Alves et al. (2003), in papers on the estimation of ρ, as well as in Gomes et al. (2004a), in a paper on the estimation of a positive extreme value index γ, we shall further specify the term $o(t^\rho)$ in Hall–Welsh class of models, and, for some particular details in the paper, we shall assume to be working with a Pareto-type class of models with a quantile function

$$U(t) \;=\; C\,t^\gamma \left(1 + D_1 t^\rho + D_2\, t^{2\rho} + o(t^{2\rho})\right) , \tag{1.4}$$

as $t \to \infty$, with $C > 0$, $D_1, D_2 \neq 0$, $\rho < 0$. Consequently, we may obviously choose, in (1.2),

$$A(t) \;=\; \rho\, D_1\, t^\rho \;=:\; \gamma\, \beta\, t^\rho \;, \qquad \beta \neq 0, \quad \rho < 0 \;, \tag{1.5}$$

and, with

$$B(t) \;=\; (2\,D_2/D_1 - D_1)\, t^\rho \;=:\; \beta'\, t^\rho \;=\; \frac{\beta' A(t)}{\beta\,\gamma} \;, \tag{1.6}$$

we may write

$$\ln U(t\,x) - \ln U(t) - \gamma \ln x \;=\; A(t) \left(\frac{x^\rho - 1}{\rho}\right) + A(t)\, B(t) \left(\frac{x^{2\rho} - 1}{2\,\rho}\right) \left(1 + o(1)\right) \;.$$

The consideration of models in (1.4) enables us to get full information on the asymptotic bias of the so-called second-order reduced-bias extreme value index estimators, the type of estimators under consideration in this paper.

Remark 1.2. Most common heavy-tailed d.f.'s, like the Fréchet, the Generalized Pareto (GP), the Burr and the Student's t belong to the class of models in (1.4), and consequently, to the class of models in (1.3) or, to the more general class of parents satisfying (1.2).

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

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

and with $X_{i:n}$ denoting the i-th ascending order statistic (o.s.), $1 \le i \le n$, associated to an independent, identically distributed (i.i.d.) random sample $(X_1, X_2, ..., X_n)$, 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,

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

We have a strong obvious link between the log-excesses and the scaled log-spacings provided by the equation, $\sum_{i=1}^{k} V_{ik} = \sum_{i=1}^{k} U_i$.

It is well known that for intermediate k, and whenever we are working with models in (1.1), the log-excesses 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, under the same conditions, the scaled log-spacings U_i, $1 \le i \le k$, are approximately i.i.d. and exponential with mean value γ. Consequently, the Hill estimator of γ (Hill, 1975),

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

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 (1.2) the asymptotic distributional representation

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

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 an asymptotically standard normal random variable. Consequently, $\sqrt{k}\,(H_n(k) - \gamma)$ converges weakly towards a normal r.v. with variance γ^2 and a non-null mean value equal to $\lambda/(1-\rho)$, whenever $\sqrt{k}\,A(n/k) \to \lambda \ne 0$, finite.

The adequate accommodation of the bias of Hill's estimator has been extensively addressed in recent years by several authors. Beirlant et al. (1999) and Feuerverger and Hall (1999) consider exponential regression techniques, based on the exponential approximations $U_i \approx \gamma\big(1 + b(n/k)\,(k/i)^{\rho}\big)\,E_i$ and $U_i \approx \gamma \exp\big(\beta\,(n/i)^{\rho}\big)\,E_i$, respectively, $1 \le i \le k$. They then proceed to the joint maximum likelihood (*ML*) estimation of the three unknown parameters or functionals at the same level k. Considering also the scaled log-spacings U_i in (1.9) to be approximately exponential with mean value $\mu_i = \gamma \exp\big(\beta(n/i)^{\rho}\big)$, $1 \le i \le k$, $\beta \ne 0$, Gomes and Martins (2002) advance with the so-called "external" estimation of the second order parameter ρ, i.e., an adequate estimation of ρ at a level k_1 higher than the level k used for the extreme value index estimation, together with a first order approximation for the *ML* estimator of β. They then obtain "quasi-*ML*" explicit estimators of γ and β, both computed at the same level k, and through that "external" estimation of ρ, are then able to reduce the asymptotic variance of the extreme value index estimator proposed, comparatively to the asymptotic variance of the extreme value index estimator in Feuerverger and Hall (1999), where the three parameters γ, β and ρ are estimated at the same level k. With the notation

$$d_k(\alpha) = \frac{1}{k}\sum_{i=1}^{k}\Big(\frac{i}{k}\Big)^{\alpha-1}, \qquad D_k(\alpha) = \frac{1}{k}\sum_{i=1}^{k}\Big(\frac{i}{k}\Big)^{\alpha-1} U_i \,, \tag{1.12}$$

for any real $\alpha \geq 1$ [$D_k(1) \equiv H(k)$ in (1.10)], and with $\hat\rho$ any consistent estimator of ρ, such estimators are

$$\widehat{\gamma}_n^{ML}(k) \;=\; H(k) - \hat\beta(k;\hat\rho)\left(\frac{n}{k}\right)^{\hat\rho} D_k(1-\hat\rho) \tag{1.13}$$

and

$$\hat\beta(k;\hat\rho) \;:=\; \left(\frac{k}{n}\right)^{\hat\rho} \frac{d_k(1-\hat\rho)\times D_k(1) - D_k(1-\hat\rho)}{d_k(1-\hat\rho)\times D_k(1-\hat\rho) - D_k(1-2\hat\rho)} \;, \tag{1.14}$$

for γ and β, respectively. This means that β, in (1.5), which is also a second order parameter, is estimated at the same level k at which the γ-estimation is performed, being $\hat\beta(k;\hat\rho)$ — not consistent for the estimation of β whenever $\sqrt{k}\,A(n/k) \to \lambda$, finite, but consistent for models in (1.2) and intermediate k such that $\sqrt{k}\,A(n/k) \to \infty$ (Gomes and Martins, 2002), — plugged in the extreme value index estimator in (1.13). In all the above mentioned papers, authors have been led to the now called "classical" second order reduced-bias extreme value index estimators with an asymptotic variance larger or equal to $\gamma^2\big((1-\rho)/\rho\big)^2$, the minimal asymptotic variance of an asymptotically unbiased estimator in Drees class of functionals (Drees, 1998).

We here propose an "external" estimation of both β and ρ, through $\hat\beta$ and $\hat\rho$, respectively, both using a number of top o.s.'s, k_1, larger than the number of top o.s.'s, k, used for the extreme value index estimation. We shall thus consider the estimator

$$ML_{\hat\beta,\hat\rho}(k) \;:=\; H(k) - \hat\beta\left(\frac{n}{k}\right)^{\hat\rho} D_k(1-\hat\rho) \;, \tag{1.15}$$

for adequate consistent estimators $\hat\beta$ and $\hat\rho$ of the second order parameters β and ρ, respectively, to be specified in subsection 3.3 of this paper. Additionally, we shall also deal with the estimator

$$\overline{ML}_{\hat\beta,\hat\rho}(k) \;=\; \frac{1}{k}\sum_{i=1}^{k} U_i \exp\!\big(-\hat\beta(n/i)^{\hat\rho}\big) \;, \tag{1.16}$$

the estimator directly derived from the likelihood equation for γ with β and ρ fixed and based upon the exponential approximation, $U_i \approx \gamma \exp\!\big(\beta(n/i)^{\rho}\big)\,E_i$, $1 \leq i \leq k$. Doing this, we are 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. The estimators are thus better than the Hill estimator for all k.

Remark 1.3. If, in (1.15), we estimate β at the same level k used for the estimation of γ, we may be led to $\widehat{\gamma}_n^{ML}(k)$ in (1.13). Indeed, $\widehat{\gamma}_n^{ML}(k) = ML_{\hat\beta(k;\hat\rho),\hat\rho}(k)$, with $\hat\beta(k;\hat\rho)$ defined in (1.14).

Remark 1.4. The ML estimator in (1.15) may be obtained from the estimator in (1.16) through the use of the first order approximation, $\big\{1-\hat\beta(n/i)^{\hat\rho}\big\}$, for the exponential weight, $e^{-\hat\beta(n/i)^{\hat\rho}}$, of the scaled log-spacing U_i, $1 \leq i \leq k$.

Remark 1.5. The estimators in (1.15) and (1.16) have been inspired in the recent papers of Gomes et al. (2004*b*) and Caeiro et al. (2005). These authors consider, in different ways, the joint external estimation of both the "scale" and the "shape" parameters in the A function in (1.2), parameterized as in (1.5), 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. Those estimators are also going to be considered here for comparison with the new estimators in (1.15) and (1.16). The reduced-bias extreme value index estimator in Gomes et al. (2004*b*) is based on a linear combination of the log-excesses V_{ik} in (1.8), and is given by

$$(1.17) \qquad WH_{\hat\beta,\hat\rho}(k) := \frac{1}{k}\sum_{i=1}^{k} e^{-\hat\beta(n/k)^{\hat\rho}\,\psi_{\hat\rho}(i/k)}\, V_{ik}\ , \qquad \psi_\rho(x) = -\frac{x^{-\rho}-1}{\rho\ln x}\ ,$$

with the notation *WH* standing for *Weighted Hill* estimator. Caeiro et al. (2005) consider the estimator

$$(1.18) \qquad \overline{H}_{\hat\beta,\hat\rho}(k) := H(k)\left(1-\frac{\hat\beta}{1-\hat\rho}\left(\frac{n}{k}\right)^{\hat\rho}\right)\ ,$$

where the dominant component of the bias of Hill's estimator $H(k)$ in (1.10), given by $A(n/k)/(1-\rho) = \beta\gamma(n/k)^\rho/(1-\rho)$, is thus estimated through $H(k)\,\hat\beta(n/k)^{\hat\rho}/(1-\hat\rho)$, and directly removed from Hill's classical extreme value index estimator. As before, both in (1.17) and (1.18), $\hat\beta$ and $\hat\rho$ need to be adequate consistent estimators of the second order parameters β and ρ, respectively, so that the new estimators are better than the Hill estimator for all k.

In section 2 of this paper, and assuming first that only γ is unknown, we shall state a theorem that provides an obvious technical motivation for the estimators in (1.15) and (1.16). Next, in section 3, we consider the derivation of the asymptotic behavior of the classes of estimators in (1.15) and (1.16), for an appropriate estimation of β and ρ at a level k_1 larger than the value k used for the extreme value index estimation. We also do that only with the estimation of ρ, estimating β at the same level k used for the extreme value index estimation. In this same section, we finally briefly review the estimation of the two second order parameters β and ρ. In section 4, using simulation techniques, we exhibit the performance of the ML estimator in (1.15) and the $\overline{ML}$ estimator in (1.16), comparatively to the other *"Unbiased Hill"* (*UH*) estimators, *WH* and $\overline{H}$, in (1.17) and (1.18), respectively, to the classical Hill estimator H in (1.10) and to the "asymptotically unbiased" estimator $\widehat\gamma_n^{ML}(k)$ in (1.13), studied in Gomes and Martins (2002), or equivalently, $ML_{\hat\beta(k;\hat\rho),\hat\rho}$, with $ML_{\hat\beta,\hat\rho}$ the estimator in (1.15). Section 5 is devoted to the illustration of the behavior of these estimators for the Daily Log-Returns of the *Euro* against the *UK* Pound and automobile claims gathered from several European insurance companies co-operating with the same re-insurer (Secura Belgian Re).

2. ASYMPTOTIC BEHAVIOR OF THE ESTIMATORS (ONLY γ IS UNKNOWN)

For real values $\alpha \geq 1$, and denoting again $\{E_i\}$ a sequence of i.i.d. standard exponential r.v.'s, let us introduce the following notation:

$$Z_k^{(\alpha)} = \sqrt{(2\alpha-1)\,k}\,\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1} E_i - \frac{1}{\alpha}\right). \tag{2.1}$$

With the same kind of reasoning as in Gomes et al. (2005a), we state:

Lemma 2.1. *Under the second order framework in* (1.2), *for intermediate k-sequences, i.e., whenever* (1.7) *holds, and with U_i given in* (1.9), *we may guarantee that, for any real $\alpha \geq 1$, and $D_k(\alpha)$ given in* (1.12),

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

where $Z_k^{(\alpha)}$, given in (2.1), *is an asymptotically standard normal random variable. If we further assume to be working with models in* (1.4), *and with the same notation as before, we may write for any $\alpha, \beta \geq 1$, $\alpha \neq \beta$, the joint distribution*

$$\begin{aligned}\big(D_k(\alpha), D_k(\beta)\big) \stackrel{d}{=}\; & \left(\frac{\gamma}{\alpha}, \frac{\gamma}{\beta}\right) + \frac{\gamma}{\sqrt{k}}\left(\frac{Z_k^{(\alpha)}}{\sqrt{(2\alpha-1)}}, \frac{Z_k^{(\beta)}}{\sqrt{(2\beta-1)}}\right)\\ & + A(n/k)\left(\frac{1}{\alpha-\rho}, \frac{1}{\beta-\rho}\right) + \frac{\beta'\,A^2(n/k)}{\beta\,\gamma}\left(\frac{1}{\alpha-2\rho}, \frac{1}{\beta-2\rho}\right)\\ & + O_p\!\left(\frac{A(n/k)}{\sqrt{k}}\right) + o_p\big(A^2(n/k)\big)\;,\end{aligned} \tag{2.2}$$

with β and β' given in (1.5) *and* (1.6), *respectively.*

Let us assume that only the extreme value index parameter γ is unknown, and generally denote $\widetilde{ML}$ either ML or $\overline{ML}$. This case obviously refers to a situation which is rarely encountered in practice, but reveals the potential of the classes of estimators in (1.15) and (1.16).

2.1. Known β and ρ

We may state:

Theorem 2.1. *Under the second order framework in* (1.2), *further assuming that* $A(t)$ *may be chosen as in* (1.5), *and for levels* k *such that* (1.7) *holds, we get asymptotic distributional representations of the type*

$$\widetilde{ML}_{\beta,\rho}(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + o_p\big(A(n/k)\big) , \tag{2.3}$$

where $Z_k^{(1)}$ *is the asymptotically standard normal r.v. in* (2.1) *for* $\alpha = 1$. *Consequently,* $\sqrt{k}\,\big(\widetilde{ML}_{\beta,\rho}(k) - \gamma\big)$ *is 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* $n \to \infty$.

For models in (1.4), *we may further specify the term* $o_p\big(A(n/k)\big)$, *writing*

$$ML_{\beta,\rho}(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + \frac{(\beta'-\beta)\, A^2(n/k)}{\beta\,\gamma\,(1-2\rho)}\, \big(1+o_p(1)\big) , \tag{2.4}$$

$$\overline{ML}_{\beta,\rho}(k) \overset{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + \frac{(2\beta'-\beta)\, A^2(n/k)}{2\,\beta\,\gamma\,(1-2\rho)}\, \big(1+o_p(1)\big) , \tag{2.5}$$

with β *and* β' *given in* (1.5) *and* (1.6), *respectively. Consequently, even if* $\sqrt{k}\,A(n/k) \to \infty$, *with* $\sqrt{k}\,A^2(n/k) \to \lambda_A$, *finite,* $\sqrt{k}\,(ML_{\beta,\rho}(k) - \gamma)$ *and* $\sqrt{k}\,(\overline{ML}_{\beta,\rho}(k) - \gamma)$ *are asymptotically normal with variance equal to* γ^2 *and asymptotic bias equal to*

$$b_{ML} = \frac{(\beta'-\beta)\,\lambda_A}{\beta\,\gamma\,(1-2\rho)} \qquad \text{and} \qquad b_{\overline{ML}} = \frac{(2\beta'-\beta)\,\lambda_A}{2\,\beta\,\gamma\,(1-2\rho)} , \tag{2.6}$$

respectively.

Proof: If all parameters are known, apart from the extreme value index γ, we get directly from Lemma 2.1,

$$\begin{aligned} ML_{\beta,\rho}(k) \;&:=\; D_k(1) - \beta\Big(\frac{n}{k}\Big)^{\rho} D_k(1-\rho) \\ &\overset{d}{=}\; \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + \frac{A(n/k)}{1-\rho} \\ &\quad - \frac{A(n/k)}{\gamma}\left(\frac{\gamma}{1-\rho} + \frac{\gamma}{\sqrt{(1-2\rho)k}}\, Z_k^{(1-\rho)} + \frac{A(n/k)}{1-2\rho}\,\big(1+o_p(1)\big)\right) \\ &\overset{d}{=}\; \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + o_p\big(A(n/k)\big) . \end{aligned}$$

Similarly, since we may write

$$\begin{aligned} \overline{ML}_{\beta,\rho}(k) \;&=\; ML_{\beta,\rho}(k) + \frac{A^2(n/k)}{2\,\gamma^2}\, D_k(1-2\rho)\,\big(1+o_p(1)\big) \\ &=\; ML_{\beta,\rho}(k) + \frac{A^2(n/k)}{2\,\gamma\,(1-2\rho)}\,\big(1+o_p(1)\big) , \end{aligned} \tag{2.7}$$

(2.3) holds for $\overline{ML}$ as well. For models in (1.4), and directly from (2.2), we get

$$\begin{aligned} ML_{\beta,\rho}(k) \;&\stackrel{d}{=}\; \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + \frac{A(n/k)}{1-\rho} + \frac{\beta'\,A^2(n/k)}{\beta\,\gamma\,(1-2\rho)}\,\big(1+o_p(1)\big) + O_p\!\left(\frac{A(n/k)}{\sqrt{k}}\right) \\ &\quad - \frac{A(n/k)}{\gamma}\left(\frac{\gamma}{1-\rho} + \frac{\gamma}{\sqrt{(1-2\rho)\,k}}\, Z_k^{(1-\rho)} + \frac{A(n/k)}{1-2\rho}\,\big(1+o_p(1)\big)\right). \end{aligned}$$

Working this expression, we finally obtain

$$ML_{\beta,\rho}(k) \;\stackrel{d}{=}\; \gamma + \frac{\gamma}{\sqrt{k}}\, Z_k^{(1)} + O_p\!\left(\frac{A(n/k)}{\sqrt{k}}\right) + \frac{A^2(n/k)}{\gamma(1-2\rho)}\left(\frac{\beta'}{\beta}-1\right)\big(1+o_p(1)\big)\;,$$

i.e., (2.4) holds. Also directly from (2.4) and (2.7), (2.5) follows. Note that since $\sqrt{k}\; O_p\big(A(n/k)/\sqrt{k}\big) = O_p\big(A(n/k)\big) \to 0$, the summand $O_p\big(A(n/k)/\sqrt{k}\big)$ is totally irrelevant for the asymptotic bias in (2.6), that follows straightforwardly from the above obtained distributional representations. □

Remark 2.1. We know that the asymptotic variances of ML and $\overline{ML}$ are the same. Since $\lambda_A \geq 0$, $b_{\overline{ML}} = b_{ML} + \lambda_A/\big(2\,\gamma\,(1-2\rho)\big) \geq b_{ML}$. We may thus say that, asymptotically, the ML-statistic is expected to exhibit a better performance than $\overline{ML}$, provided the bias are both positive. Things work the other way round if the bias are both negative, i.e., the sample paths of $\overline{ML}$ are expected to be in average above the ones of ML.

Remark 2.2. For the Burr d.f. $F(x) = 1-(1+x^{-\rho/\gamma})^{1/\rho}$, $x \geq 0$, we have $U(t) = t^\gamma(1-t^\rho)^{-\gamma/\rho} = t^\gamma\big(1+\gamma\, t^\rho/\rho + \gamma(\gamma+\rho)\, t^{2\rho}/(2\,\rho^2) + o(t^{2\rho})\big)$, for $t \geq 1$. Consequently, (1.4) holds with $D_1 = \gamma/\rho$, $D_2 = \gamma(\gamma+\rho)/(2\,\rho^2)$, $\beta' = \beta = 1$ and $b_{ML} = 0$. A similar result holds for the GP d.f. $F(x) = 1-(1+\gamma\, x)^{-1/\gamma}$, $x \geq 0$. For this d.f., $U(t) = (t^\gamma-1)/\gamma$, and (1.4) holds with $\rho = -\gamma$, $D_1 = -1$ and $D_2 = 0$. Hence $\beta = \beta' = 1$ and $b_{ML} = 0$. We thus expect a better performance of ML, comparatively to $\overline{ML}$, WH and $\overline{H}$ whenever the model underlying the data is close to Burr or to GP models, a situation that happens often in practice, and that is another point in favour of the ML-statistic.

2.2. Known ρ

We may state the following:

Theorem 2.2. *For models in* (1.4), *if* $k = k_n$ *is a sequence of intermediate integers, i.e.,* (1.7) *holds, and if* $\sqrt{k}\; A(n/k) \to \infty$, *with* $\sqrt{k}\; A^2(n/k)$ *converging towards* λ_A, *finite, as* $n \to \infty$, *then, with* $\hat\beta(k;\hat\rho)$, $ML_{\hat\beta,\hat\rho}(k)$ *and* $\overline{ML}_{\hat\beta,\hat\rho}(k)$ *given in* (1.14), (1.15) *and* (1.16), *respectively, the asymptotic variance of both* $ML^*(k) =$

$ML_{\hat\beta(k;\rho),\rho}(k)$ and $\overline{ML}^*(k) = \overline{ML}_{\hat\beta(k;\rho),\rho}(k)$ is equal to $\big(\gamma(1-\rho)/\rho\big)^2$, being their asymptotic bias given by

$$(2.8)\qquad b^*_{ML} = \frac{(\beta-\beta')\,(1-\rho)\,\lambda_A}{\beta\,\gamma\,(1-2\rho)\,(1-3\rho)} \quad \text{and} \quad b^*_{\overline{ML}} = \frac{\Big(\beta\,(3-5\rho)-2\,\beta'(1-\rho)\Big)\,\lambda_A}{2\,\beta\,\gamma\,(1-2\rho)\,(1-3\rho)}\ ,$$

respectively, again with β and β' given in (1.5) *and* (1.6)*, respectively.*

Proof: Following the steps in Gomes and Martins (2002), but working now with models in (1.4) and the distributional representation (2.2), we may write:

$$\begin{aligned} ML^*(k) \;&=\; H(k) - \frac{D_k(1-\rho)\,\Big\{D_k(1)\,\big(1+o(1)\big) - (1-\rho)\,D_k(1-\rho)\Big\}}{D_k(1-\rho)\,\big(1+o(1)\big) - (1-\rho)\,D_k(1-2\rho)} \\ &=:\; H(k) - \frac{\varphi_k(\rho)}{\psi_k(\rho)}\ , \end{aligned}$$

with $D_k(\alpha)$ given in (1.12). Directly from (2.2), we get

$$\frac{1}{\psi_k(\rho)} \;=\; -\frac{(1-\rho)\,(1-2\rho)}{\gamma\,\rho^2}\left(1-\left\{\frac{2\,(1-\rho)\,A(n/k)}{\gamma\,(1-3\rho)} + O_p\Big(\frac{1}{\sqrt{k}}\Big)\right\}\big(1+o_p(1)\big)\right)$$

and, under the conditions on k imposed,

$$\begin{aligned} \varphi_k(\rho) \;&=\; \frac{\gamma^2}{\sqrt{k}}\left(\frac{Z_k^{(1)}}{1-\rho} - \frac{Z_k^{(1-\rho)}}{\sqrt{1-2\rho}}\right) - \frac{\gamma\,\rho^2\,A(n/k)}{(1-\rho)^2\,(1-2\rho)} \\ &\quad - \frac{\rho^2\,A^2(n/k)}{(1-\rho)\,(1-2\rho)}\left(\frac{2\,\beta'}{\beta\,(1-3\rho)} + \frac{1}{1-2\rho}\right)\big(1+o_p(1)\big)\ . \end{aligned}$$

Consequently,

$$\begin{aligned} \frac{\varphi_k(\rho)}{\psi_k(\rho)} \;&=\; -\frac{\gamma}{\rho^2\sqrt{k}}\Big((1-2\rho)\,Z_k^{(1)} - (1-\rho)\,\sqrt{1-2\rho}\;Z_k^{(1-\rho)}\Big) + \frac{A(n/k)}{1-\rho} \\ &\quad + \frac{A^2(n/k)}{\gamma}\left(\frac{2\,(\beta'-\beta)}{\beta\,(1-3\rho)} + \frac{1}{1-2\rho}\right)\big(1+o_p(1)\big)\ . \end{aligned}$$

Then, with

$$\overline{Z}_k \;=\; \left(\frac{1-\rho}{\rho}\right)^2 Z_k^{(1)} - \left(\frac{(1-\rho)\,\sqrt{1-2\rho}}{\rho^2}\right) Z_k^{(1-\rho)}\ ,$$

$$ML^*(k) \;=\; ML_{\hat\beta(k;\rho),\rho}(k) \;\overset{d}{=}\; \gamma + \frac{\gamma}{\sqrt{k}}\,\overline{Z}_k - \frac{(\beta'-\beta)\,(1-\rho)\,A^2(n/k)}{\beta\,\gamma\,(1-2\rho)\,(1-3\rho)}\,\big(1+o_p(1)\big)\ ,$$

and the result in (2.8) follows for $ML^*(k)$. Also, since the asymptotic covariance between $Z_k^{(1)}$ and $Z_k^{(1-\rho)}$ is given by $\sqrt{1-2\rho}/(1-\rho)$, the asymptotic variance of $\overline{Z}_k$ is given by

$$\left(\frac{1-\rho}{\rho}\right)^4 + \frac{(1-\rho)^2\,(1-2\rho)}{\rho^4} - \frac{2\,(1-\rho)^3\,\sqrt{1-2\rho}}{\rho^4} \times \frac{\sqrt{1-2\rho}}{1-\rho} \;=\; \left(\frac{1-\rho}{\rho}\right)^2.$$

Hence, the asymptotic variance $\gamma^2\big\{(1-\rho)/\rho\big\}^2$, stated in the theorem. If we consider $\overline{ML}_{\hat\beta(k;\rho),\rho}(k)$, since $\sqrt{k}\,A(n/k)\to\infty$, $\hat\beta(k;\rho)$ converges in probability towards β and a result similar to (2.7) holds, i.e.,

$$\overline{ML}^*(k) \;=\; \overline{ML}_{\hat\beta(k;\rho),\rho}(k) \;=\; ML_{\hat\beta(k;\rho),\rho}(k) + \frac{A^2(n/k)}{2\gamma\,(1-2\rho)}\,\big(1+o_p(1)\big)\;.$$

The result in the theorem follows thus straightforwardly. □

Remark 2.3. For models in (1.4) and $\lambda_A \neq 0$ in Theorem 2.2, $b^*_{ML}=0$ if and only if $\beta=\beta'$. Again, this holds for Burr and *GP* underlying models.

Remark 2.4. When we look at Theorems 2.1 and 2.2, we see that, for (β,ρ) known, despite the increasing in the asymptotic variance, $(b_{ML}/b^*_{ML})^2 = \big((1-3\rho)/(1-\rho)\big)^2$ is an increasing function of $|\rho|$, always greater than one, for $\rho<0$, i.e., there is here again a compromise between bias and variance.

2.3. Asymptotic comparison at optimal levels

We now proceed to an asymptotic comparison of *ML* and *ML** at their optimal levels in the lines of de Haan and Peng (1998), Gomes and Martins (2001) and Gomes et al. (2005b, 2006), among others, but now for second order reduced-bias estimators. Suppose $\widehat\gamma^\bullet_n(k)$ is a general semi-parametric estimator of the extreme value index γ, for which the distributional representation

$$\widehat\gamma^\bullet_n(k) \;\overset{d}{=}\; \gamma + \frac{\sigma_\bullet}{\sqrt{k}}\,Z^\bullet_n + b_\bullet\,A^2(n/k) + o_p\big(A^2(n/k)\big) \tag{2.9}$$

holds for any intermediate k, and where $Z^\bullet_n$ is an asymptotically standard normal random variable. Then we have

$$\sqrt{k}\,\big[\widehat\gamma^\bullet_n(k)-\gamma\big] \;\overset{d}{\rightarrow}\; N(\lambda_A b_\bullet, \sigma^2_\bullet), \qquad \text{as } n\to\infty\;,$$

provided k is such that $\sqrt{k}\,A^2(n/k)\to\lambda_A$, finite, as $n\to\infty$. In this situation we may write $Bias_\infty[\widehat\gamma^\bullet_n(k)] := b_\bullet\,A^2(n/k)$ and $Var_\infty[\widehat\gamma^\bullet_n(k)] := \sigma^2_\bullet/k$. The so-called Asymptotic Mean Squared Error (*AMSE*) is then given by

$$AMSE\,[\widehat\gamma^\bullet_n(k)] \;:=\; \frac{\sigma^2_\bullet}{k} + b^2_\bullet\,A^4(n/k)\;.$$

Using regular variation theory (Bingham et al., 1987), it may be proved that, whenever $b_\bullet\neq 0$, there exists a function $\varphi(n)$, dependent only on the underlying model, and not on the estimator, such that

$$\lim_{n\to\infty}\varphi(n)\,AMSE\,[\widehat\gamma^\bullet_{n0}] \;=\; C(\rho)\,(\sigma^2_\bullet)^{-\frac{4\rho}{1-4\rho}}\,(b^2_\bullet)^{\frac{1}{1-4\rho}} \;=:\; LMSE\,[\widehat\gamma^\bullet_{n0}]\;, \tag{2.10}$$

where $\widehat{\gamma}^{\bullet}_{n0} := \widehat{\gamma}^{\bullet}_{n}(k^{\bullet}_0(n))$, $k^{\bullet}_0(n) := \arg\inf_k AMSE\,[\widehat{\gamma}^{\bullet}_n(k)]$, is the estimator $\widehat{\gamma}^{\bullet}_n(k)$ computed at its optimal level, the level where its $AMSE$ is minimum.

It is then sensible to consider the usual:

Definition 2.1. Given two second order reduced-bias estimators, $\widehat{\gamma}^{(1)}_n(k)$ and $\widehat{\gamma}^{(2)}_n(k)$, for which distributional representations of the type (2.9) hold, with constants (σ_1, b_1) and (σ_2, b_2), $b_1, b_2 \neq 0$, respectively, both computed at their optimal levels, the Asymptotic Root Efficiency ($AREFF$) of $\widehat{\gamma}^{(1)}_{n0}$ relatively to $\widehat{\gamma}^{(2)}_{n0}$ is $AREFF_{1|2} \equiv AREFF_{\widehat{\gamma}^{(1)}_{n0}|\widehat{\gamma}^{(2)}_{n0}} := \left(LMSE\left[\widehat{\gamma}^{(2)}_{n0}\right]/LMSE\left[\widehat{\gamma}^{(1)}_{n0}\right]\right)^{1/2}$, with $LMSE$ given in (2.10).

Remark 2.5. This measure was devised so that the higher the $AREFF$ measure the better the estimator 1 is, comparatively to the estimator 2.

Proposition 2.1. *For every $\beta \neq \beta'$, if we compare $ML = ML_{\beta,\rho}$ and $ML^* = ML_{\hat{\beta}(k;\rho),\rho}$, we get $AREFF_{ML|ML^*} = (1-\rho)^2\left((1-3\rho)\,\rho^{-4\rho}\right)^{-1/(1-4\rho)} > 1$ for all $\rho < 0$.*

We may also say that $AREFF_{ML|\overline{ML}} > 1$, for all ρ, β and β'. This indicator depends then not only of ρ, but also of β and β'. This result, together with the result in Proposition 2.1, provides again a clear indication on an overall better performance of the ML estimator, comparatively to $\overline{ML}$ and ML^*.

3. EXTREME VALUE INDEX ESTIMATION BASED ON THE ESTIMATION OF THE SECOND ORDER PARAMETERS β AND ρ

Again for $\alpha \geq 1$, let us further introduce the following extra notations:

$$W^{(\alpha)}_k = (2\alpha-1)\sqrt{(2\alpha-1)\,k/2}\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1}\ln\!\left(\frac{i}{k}\right)E_i + \frac{1}{\alpha^2}\right), \tag{3.1}$$

$$D'_k(\alpha) = \frac{d\,D_k(\alpha)}{d\,\alpha} := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1}\ln\!\left(\frac{i}{k}\right)U_i\,, \tag{3.2}$$

with U_i and $D_k(\alpha)$ given in (1.9) and (1.12), respectively.

Again with the same kind of reasoning as in Gomes et al. (2005a), we state:

Lemma 3.1. *Under the second order framework in* (1.2), *for intermediate* k*-sequences, i.e., whenever* (1.7) *holds, and with* U_i *given in* (1.9), *we may guarantee that, for any real* $\alpha \geq 1$ *and with* $D'_k(\alpha)$ *given in* (3.2),

$$D'_k(\alpha) \overset{d}{=} -\frac{\gamma}{\alpha^2} + \frac{\gamma\, W_k^{(\alpha)}}{(2\alpha-1)\sqrt{(2\alpha-1)\,k/2}} - \frac{A(n/k)}{(\alpha-\rho)^2}\left(1+o_p(1)\right) , \tag{3.3}$$

where $W_k^{(\alpha)}$, *in* (3.1), *are asymptotically standard normal r.v.'s.*

3.1. Estimation of both second order parameters β and ρ at a lower threshold

Let us assume first that we estimate both β and ρ externally at a level k_1 of a larger order than the level k at which we compute the extreme value index estimator, now assumed to be an intermediate level k such that $\sqrt{k}\, A(n/k) \to \lambda$, finite, as $n \to \infty$, with $A(t)$ the function in (1.2). We may state the following:

Theorem 3.1. *Under the initial conditions of Theorem 2.1, let us consider the class of extreme value index estimators* $\widetilde{ML}_{\hat\beta,\hat\rho}(k)$, *with* $\widetilde{ML}$ *denoting again either the ML estimator in* (1.15) *or the* $\overline{ML}$ *estimator in* (1.16), *with* $\hat\beta$ *and* $\hat\rho$ *consistent for the estimation of* β *and* ρ, *respectively, and such that*

$$(\hat\rho-\rho)\,\ln n = o_p(1), \qquad \text{as } n\to\infty . \tag{3.4}$$

Then, $\sqrt{k}\left\{\widetilde{ML}_{\hat\beta,\hat\rho}(k)-\gamma\right\}$ *is asymptotically normal with null mean value and variance* $\sigma_1^2=\gamma^2$, *not only when* $\sqrt{k}\,A(n/k)\to 0$, *but also whenever* $\sqrt{k}\,A(n/k)\to \lambda\neq 0$, *finite.*

Proof: With the usual notation $X_n \overset{p}{\sim} Y_n$ if and only if X_n/Y_n goes in probability to 1, as $n\to\infty$, we may write

$$\frac{\partial \widetilde{ML}_{\beta,\rho}}{\partial\beta} \overset{p}{\sim} -\left(\frac{n}{k}\right)^{\rho} D_k(1-\rho) = -\frac{A(n/k)\, D_k(1-\rho)}{\beta\,\gamma} \overset{p}{\sim} -\frac{A(n/k)}{\beta(1-\rho)}$$

and

$$\begin{aligned}\frac{\partial \widetilde{ML}_{\beta,\rho}}{\partial\rho} &\overset{p}{\sim} -\frac{A(n/k)}{\gamma}\left(\ln\left(\frac{n}{k}\right) D_k(1-\rho) - D'_k(1-\rho)\right)\\ &\overset{p}{\sim} -\frac{A(n/k)}{1-\rho}\left(\ln\left(\frac{n}{k}\right) + \frac{1}{1-\rho}\right) .\end{aligned}$$

If we estimate consistently ρ and β through the estimators $\hat\beta$ and $\hat\rho$ in the conditions of the theorem, we may use Taylor's expansion series, and we obtain

$$\widetilde{ML}_{\hat\beta,\hat\rho}(k)-\widetilde{ML}_{\beta,\rho}(k) \overset{p}{\sim} -\frac{A(n/k)}{1-\rho}\left\{\left(\frac{\hat\beta-\beta}{\beta}\right)+(\hat\rho-\rho)\left(\ln(n/k)+\frac{1}{1-\rho}\right)\right\} . \tag{3.5}$$

Consequently, taking into account the conditions in the theorem,

$$\widetilde{ML}_{\hat\beta,\hat\rho}(k) - \widetilde{ML}_{\beta,\rho}(k) \;=\; o_p\big(A(n/k)\big) \;.$$

Hence, if $\sqrt{k}\;A(n/k) \to \lambda$, finite, Theorem 2.1 enables us to guarantee the results in the theorem. □

3.2. Estimation of the second order parameter ρ only at a lower threshold

If we consider γ and β estimated at the same level k, we are going to have an increase in the asymptotic variance of our final extreme value index estimators, but we no longer need to assume that condition (3.4) holds. Indeed, as stated in Corollary 2.1 of Theorem 2.1 in Gomes and Martins (2002), for the estimator in (1.13), Theorem 3.2 in Gomes et al. (2004b), for the estimator $WH_{\hat\beta(k;\hat\rho),\hat\rho}$ and Theorem 3.2 in Caeiro et al. (2005), for the estimator $\overline{H}_{\hat\beta(k;\hat\rho),\hat\rho}$, we may state:

Theorem 3.2. (Gomes and Martins, 2002; Gomes et al., 2004b; Caeiro et al., 2005) *Under the second order framework in* (1.2)*, if $k = k_n$ is a sequence of intermediate integers, i.e.,* (1.7) *holds, and if $\lim_{n\to\infty} \sqrt{k}\;A(n/k) = \lambda$, finite, then, with UH denoting any of the statistics ML, $\overline{ML}$, WH or $\overline{H}$ in* (1.15)*,* (1.16)*,* (1.17) *and* (1.18)*, respectively, $\hat\rho$ any consistent estimator of the second order parameter ρ, and $\hat\beta(k;\hat\rho)$ the β-estimator in* (1.14)*,*

$$\text{(3.6)}\qquad \sqrt{k}\Big(UH_{\hat\beta(k;\hat\rho),\hat\rho}(k) - \gamma\Big) \;\underset{n\to\infty}{\overset{d}{\longrightarrow}}\; Normal\bigg(0,\;\sigma_2^2 := \gamma^2\Big(\frac{1-\rho}{\rho}\Big)^2\bigg)\;,$$

i.e., the asymptotic variance of $UH_{\hat\beta(k;\hat\rho),\hat\rho}(k)$ increases of a factor $\big((1-\rho)/\rho\big)^2 > 1$ for every $\rho < 0$.

Remark 3.1. If we compare Theorem 3.1 and Theorem 3.2, we see that, as expected, 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 $\big((1-\rho)/\rho\big)^2$, greater than 1. The estimation of the three parameters γ, β and ρ at the same level k may still induce an extra increase in the asymptotic variance of the final γ-estimator, as may be seen in Feuerverger and Hall (1999) (where the three parameters are indeed computed at the same level k). These authors get an asymptotic variance ruled by $\sigma^2_{FH} := \gamma^2\big((1-\rho)/\rho\big)^4$, and we have $\sigma_1 < \sigma_2 < \sigma_{FH}$ for all $\rho < 0$. Consequently, and taking into account asymptotic variances, it seems convenient to estimate both β and ρ "externally", at a level k_1 of a larger order than the level k used for the estimation of the extreme value index γ.

3.3. How to estimate the second order parameters

We now provide some details on the type of second order parameters' estimators we think sensible to use in practice, together with their distributional properties.

3.3.1. The estimation of ρ

Several classes of ρ-estimators are available in the literature. Among them, we mention the ones introduced in Hall and Welsh (1985), Drees and Kaufman (1998), Peng (1998), Gomes et al. (2002) and Fraga Alves et al. (2003). The one working better in practice and for the most common heavy-tailed models, is the one in Fraga Alves et al. (2003). We shall thus consider here particular members of this class of estimators. Under adequate general conditions, and for $\rho < 0$, they are semi-parametric asymptotically normal estimators of ρ, which show highly stable sample paths as functions of k_1, the number of top o.s.'s used, for a wide range of large k_1-values. Such a class of estimators has been first parameterized by a tuning parameter $\tau > 0$, but τ may be more generally considered as a real number (Caeiro and Gomes, 2004), and is defined as

$$\hat\rho(k_1;\tau) \equiv \hat\rho_\tau(k_1) \equiv \hat\rho_n^{(\tau)}(k_1) := -\left|\frac{3\left(T_n^{(\tau)}(k_1)-1\right)}{T_n^{(\tau)}(k_1)-3}\right| , \tag{3.7}$$

where

$$T_n^{(\tau)}(k_1) := \frac{\left(M_n^{(1)}(k_1)\right)^\tau - \left(M_n^{(2)}(k_1)/2\right)^{\tau/2}}{\left(M_n^{(2)}(k_1)/2\right)^{\tau/2} - \left(M_n^{(3)}(k_1)/6\right)^{\tau/3}} , \qquad \tau\in\mathbb{R} ,$$

with the notation $a^{b\tau} = b\ln a$, whenever $\tau = 0$ and with

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

We shall here summarize a particular case of the results proved in Fraga Alves et al. (2003):

Proposition 3.1 (Fraga Alves et al., 2003)**.** *Under the second order framework in* (1.2)*, if k_1 is an intermediate sequence of integers, and if $\sqrt{k_1}\,A(n/k_1)\to\infty$, as $n\to\infty$, the statistics $\hat\rho_n^{(\tau)}(k_1)$ in* (3.7) *converge in probability towards ρ, as $n\to\infty$, for any real τ. Moreover, for models in* (1.4)*, if we further assume* *that* $\sqrt{k_1}\, A^2(n/k_1) \longrightarrow \lambda_{A_1}$, *finite,* $\hat\rho_\tau(k_1) \equiv \hat\rho_n^{(\tau)}(k_1)$ *is asymptotically normal with a bias proportional to* λ_{A_1}, *and* $\{\hat\rho_\tau(k_1) - \rho\} = O_p\big(1/(\sqrt{k_1}\, A(n/k_1))\big)$. *If* $\sqrt{k_1}\, A^2(n/k_1) \to \infty$, $\{\hat\rho_\tau(k_1) - \rho\} = O_p\big(A(n/k_1)\big)$.

Remark 3.2. Note that if we choose for the estimation of ρ a level k_1 under the conditions that assure, in Proposition 3.1, asymptotic normality with a non-null bias, we may guarantee that $k_1 = O\big(n^{-4\rho/(1-4\rho)}\big)$ and consequently $\sqrt{k_1}\, A(n/k_1) = O\big(n^{-\rho/(1-4\rho)}\big)$. Hence, $\hat\rho_\tau(k_1) - \rho = O_p\big(1/(\sqrt{k_1}\, A(n/k_1))\big) = O_p\big(n^{\rho/(1-4\rho)}\big) = o_p(1/\ln n)$ provided that $\rho < 0$, i.e., (3.4) holds whenever we assume $\rho < 0$.

Remark 3.3. The adaptive choice of the level k_1 suggested in Remark 3.2 is not straightforward in practice. The theoretical and simulated results in Fraga Alves et al. (2003), together with the use of these ρ-estimators in the Generalized Jackknife statistics of Gomes et al. (2000), as done in Gomes and Martins (2002), has led these authors to advise the choice $k_1 = \min\big(n-1, [2n/\ln\ln n]\big)$, to estimate ρ. Note however that with such a choice of k_1, $\sqrt{k_1}\, A^2(n/k_1) \to \infty$ and $\{\hat\rho_\tau(k_1) - \rho\} = O_p\big(A(n/k_1)\big) = O_p\big((\ln\ln n)^\rho\big)$. Consequently, without any further restrictions on the behavior of the ρ-estimators, we may no longer guarantee that (3.4) holds.

Remark 3.4. Here, and inspired in the results in Gomes et al. (2004*b*) for the estimator in (1.17), we advise the consideration of a level of the type

$$k_1 = \big[n^{1-\epsilon}\big], \qquad \text{for some } \ \epsilon > 0, \ \text{ small}\ , \tag{3.8}$$

where $[x]$ denotes, as usual, the integer part of x. When we consider the level k_1 in (3.8), $\sqrt{k_1}\, A^2(n/k_1) \to \infty$, if and only if $\rho > \frac{1}{4} - \frac{1}{4\epsilon} \to -\infty$, as $\epsilon \to 0$, and such a condition is an almost irrelevant restriction in the underlying model, provided we choose a small value of ϵ. For instance, if we choose $\epsilon = 0.001$, we get $\rho > -249.75$. Then, and with such an irrelevant restriction in the models in (1.4), if we work with any of the ρ-estimators in this section, computed at the level k_1, $\{\hat\rho - \rho\}$ is of the order of $A(n/k_1) = O(n^{\epsilon\times\rho})$, which is of smaller order than $1/\ln n$. This means that, again, condition (3.4) holds, being the choice in (3.8) a very adequate choice in practice.

We advise practitioners not to choose blindly the value of τ in (3.7). It is sensible to draw some sample paths of $\hat\rho(k;\tau)$, as functions of k and for a few τ-values, electing the value of $\tau \equiv \tau^*$ which provides the highest stability for large k, by means of any stability criterion, like the ones suggested in Gomes et al. (2004*a*), Gomes and Pestana (2004) and Gomes et al. (2005*a*). Anyway, in all the Monte Carlo simulations we have considered the level k_1 in (3.8), with

$\epsilon = 0.001$, and

$$\hat{\rho}_\tau := -\left|\frac{3\left(T_n^{(\tau)}(k_1)-1\right)}{T_n^{(\tau)}(k_1)-3}\right|, \qquad \tau = \begin{cases} 0 & \text{if } \rho \geq -1\,, \\ 1 & \text{if } \rho < -1\,. \end{cases} \tag{3.9}$$

Indeed, an adequate stability criterion, like the one used in Gomes and Pestana (2004), has practically led us to this choice for all models simulated, whenever the sample size n is not too small. Note also that the choice of the most adequate value of τ, let us say the tuning parameter $\tau = \tau^*$ mentioned before, is much more relevant than the choice of the level k_1, in the ρ-estimation and everywhere in the paper, whenever we use second order parameters' estimators in order to estimate the extreme value index.

From now on we shall generally use the notation $\hat{\rho} \equiv \hat{\rho}_\tau = \hat{\rho}(k_1;\tau)$ for any of the estimators in (3.7) computed at a level k_1 in (3.8).

3.3.2. The estimation of β based on the scaled log-spacings

We have here considered the estimator of β obtained in Gomes and Martins (2002), already defined in (1.14), and based on the scaled log-spacings U_i in (1.9), $1 \leq i \leq k$. The first part of the following result has been proved in Gomes and Martins (2002) and the second part, related to the behavior of $\hat{\beta}(k;\hat{\rho}(k;\tau))$, has been proved in Gomes et al. (2004b):

Proposition 3.2 (Gomes and Martins, 2002; Gomes et al., 2004b)**.** *If the second order condition* (1.2) *holds, with* $A(t) = \beta\,\gamma\,t^\rho$, $\rho < 0$, *if* $k = k_n$ *is a sequence of intermediate positive integers, i.e.* (1.7) *holds, and if* $\lim_{n\to\infty} \sqrt{k}\,A(n/k) = \infty$, *then* $\hat{\beta}(k;\rho)$, *defined in* (1.14), *converges in probability towards* β, *as* $n \to \infty$. *Moreover, if* (3.4) *holds,* $\hat{\beta}(k;\hat{\rho})$ *is consistent for the estimation of* β. *We may further say that*

$$\hat{\beta}\big(k;\hat{\rho}(k;\tau)\big) - \beta \;\overset{p}{\sim}\; -\beta\,\ln(n/k)\,\big(\hat{\rho}(k;\tau) - \rho\big)\,, \tag{3.10}$$

with $\hat{\rho}(k;\tau)$ *given in* (3.7). *Consequently,* $\hat{\beta}\big(k;\hat{\rho}(k;\tau)\big)$ *is consistent for the estimation of* β *whenever* (1.7) *holds and* $\sqrt{k}\,A(n/k)/\ln(n/k) \to \infty$. *For models in* (1.4), $\hat{\beta}\big(k;\hat{\rho}(k;\tau)\big) - \beta = O_p\big(\ln(n/k)/(\sqrt{k}\,A(n/k))\big)$ *whenever* $\sqrt{k}\,A^2(n/k) \to \lambda_A$, *finite. If* $\sqrt{k}\,A^2(n/k) \to \infty$, *then* $\hat{\beta}\big(k;\hat{\rho}(k;\tau)\big) - \beta = O_p\big(\ln(n/k)\,A(n/k)\big)$.

An algorithm for second order parameter estimation, in a context of high quantiles estimation, can be found in Gomes and Pestana (2005).

4. FINITE SAMPLE BEHAVIOR OF THE ESTIMATORS

4.1. Simulated models

In the simulations we have considered the following underlying parents: the *Fréchet* model, with d.f. $F(x) = \exp(-x^{-1/\gamma})$, $x \geq 0$, $\gamma > 0$, for which $\rho = -1$, $\beta = 1/2$, $\beta' = 5/6$; and the *GP* model, with d.f. $F(x) = 1-(1+\gamma x)^{-1/\gamma}$, $x \geq 0$, $\gamma > 0$, for which $\rho = -\gamma$, $\beta = 1$, $\beta' = 1$.

4.2. Mean values and mean squared error patterns

We have here implemented simulation experiments with 5000 runs, based on the estimation of β at the level k_1 in (3.8), with $\epsilon = 0.001$, the same level we have used for the estimation of ρ. We use the notation $\hat\beta_{j1} = \hat\beta(k_1; \hat\rho_j)$, $j = 0, 1$, with $\beta(k; \hat\rho)$ and $\hat\rho_\tau$, $\tau = 0, 1$, given in (1.14) and (3.9), respectively. Similarly to what has been done in Gomes et al. (2004b) for the *WH*-estimator, in (1.17), and in Caeiro et al. (2005) for the $\overline{H}$-estimator, in (1.18), these estimators of ρ and β have been also incorporated in the $\widetilde{ML}$-estimators, leading to $\widetilde{ML}_0(k) \equiv \widetilde{ML}_{\hat\beta_{01},\hat\rho_0}(k)$ or to $\widetilde{ML}_1(k) \equiv \widetilde{ML}_{\hat\beta_{11},\hat\rho_1}(k)$, with $\widetilde{ML}$ denoting both ML and $\overline{ML}$ in (1.15) and (1.16), respectively.

The simulations show that the extreme value index estimators $UH_j(k) \equiv UH_{\hat\beta_{j1},\hat\rho_j}(k)$, with UH denoting again either ML or $\overline{ML}$ or WH or $\overline{H}$, j equal to either 0 or 1, according as $|\rho| \leq 1$ or $|\rho| > 1$, seem to work reasonably well, as illustrated in Figures 1, 2 and 3. In these figures we picture for the above mentioned underlying models, 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 UH_j (*left*), $UH_j^* \equiv UH_{\beta(k;\hat\rho_j),\hat\rho_j}$ (*right*), with $j = 0$ or $j = 1$, according as $|\rho| \leq 1$ or $|\rho| > 1$ and the r.v.'s $UH \equiv UH_{\beta,\rho}$ (*center*). The discrepancy, in some of the models, between the behavior of the estimators proposed in this paper, the ones in the left figures, and the r.v.'s, in the central ones, suggests that some improvement in the estimation of second order parameters β and ρ is still welcome.

Remark 4.1. For the Fréchet model (Figure 1), the $UH_{\hat\beta,\hat\rho}$ estimators exhibit a negative bias up to moderate values of k and consequently, as hinted in Remark 2.1, the ML statistic is the one exhibiting the worst performance in terms of bias and minimum mean squared error. The $\overline{ML}_0$ estimator, always quite close to WH_0, exhibits the best performance among the statistics considered.

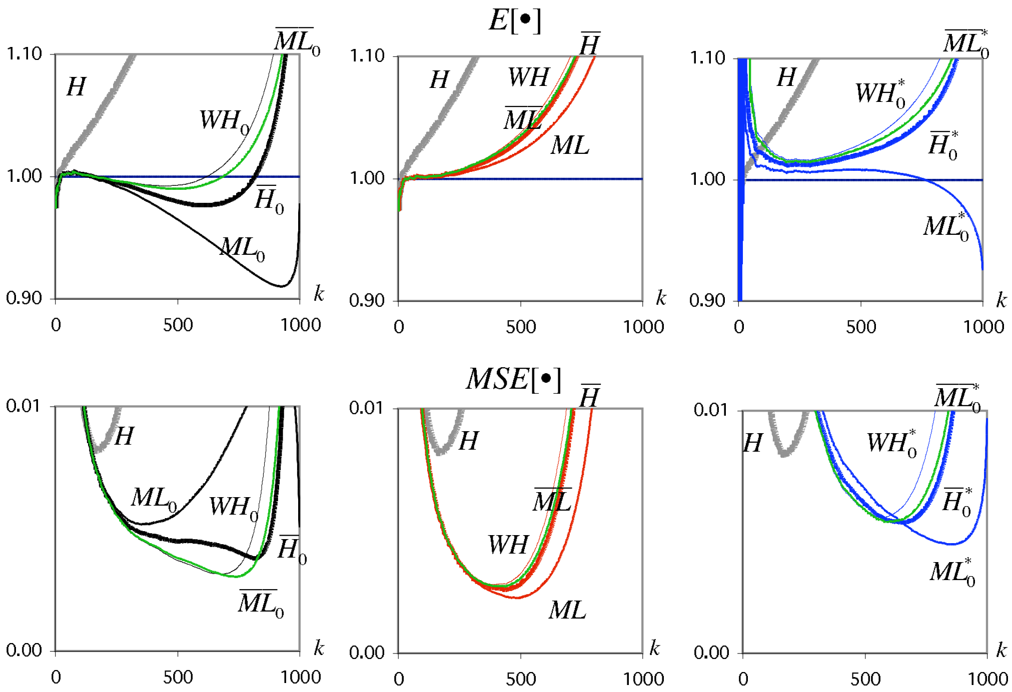

Figure 1: Underlying *Fréchet* parent with $\gamma = 1$ ($\rho = -1$).

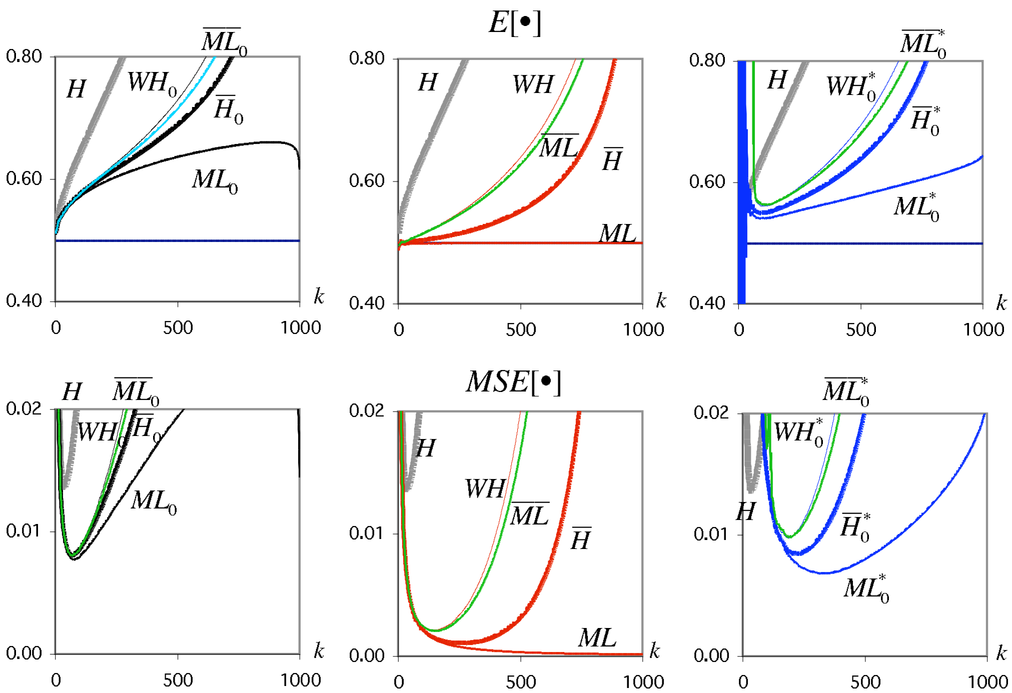

Figure 2: Underlying *GP* parent with $\gamma = 0.5$ ($\rho = -0.5$).

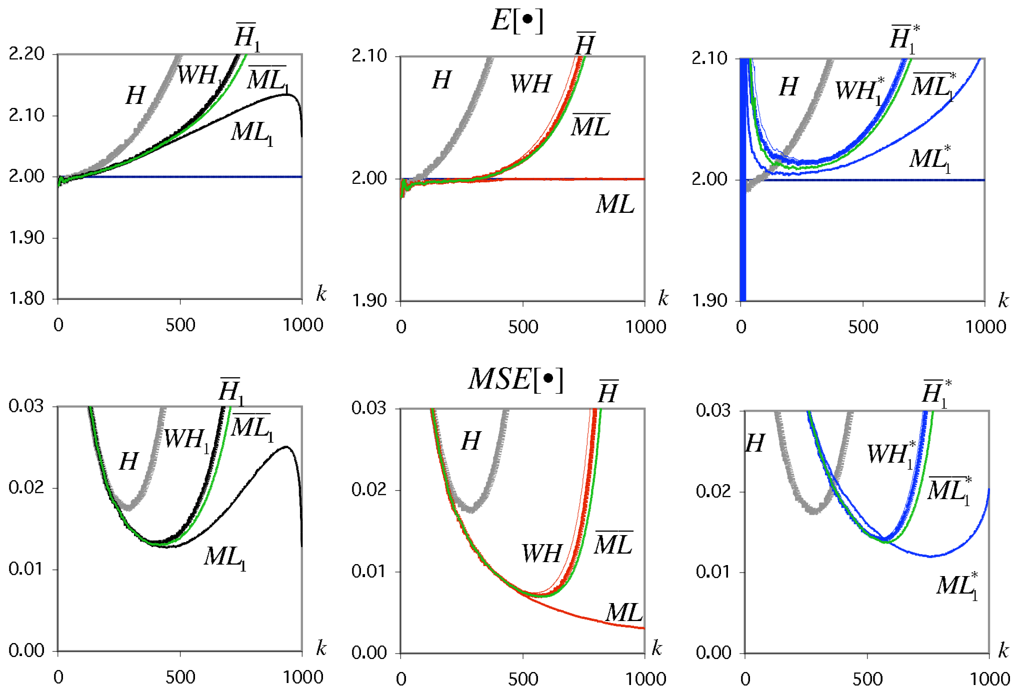

Figure 3: Underlying *GP* parent with $\gamma = 2$ ($\rho = -2$).

Things work the other way round, either with the r.v.'s UH (Figure 1, center) or with the statistics UH_0^* (Figure 1, right). The ML_0^* statistic is then the one with the best performance.

Remark 4.2. For a *GP* model, we make the following comments:

1) The ML statistic behaves indeed as a "really unbiased" estimator of γ, should we get to know the true values of β and ρ (see the central graphs of Figures 2 and 3). Indeed $b_{ML} = 0$ (see Remark 2.2), but we believe that more than this happens, although we have no formal proof of the unbiasedness of $ML(k)$ for all k and for Burr and *GP* models, among other possible parents.

2) For values of $\rho > -1$ (Figure 2), the estimators exhibit a positive bias, overestimating the true value of the parameter, and the ML-statistic is better than $\overline{H}$, which on its turn behaves better than $\overline{ML}$, this one better than WH, both regarding bias and mean squared error and in all situations (either when β and ρ are known or when β and ρ are estimated at the larger level k_1 or when only ρ is estimated at a larger level k_1, with β estimated at the same level than the extreme value index).

3) For $\rho < -1$ (Figure 3), we need to use $\hat{\rho}_1$ (instead of $\hat{\rho}_0$) or an hybrid estimator like the one suggested in Gomes and Pestana (2004). In all the simulated cases the ML_1-statistic is always the best one, being $\overline{ML}_1$, $\overline{H}_1$ and WH_1 almost equivalent.

4.3. Simulated comparative behavior at optimal levels

In Table 1, for the above mentioned Fréchet$(\gamma=1)$, $GP(\gamma=.5)$ and $GP(\gamma=2)$ parents and for the r.v.'s $UH \equiv UH_{\beta,\rho}$, we present the simulated values of the following characteristics at optimal levels: the optimal sample sample fraction (OSF)/ mean value (E) (*first row*) and the mean squared error (MSE)/ Relative Efficiency ($REFF$) indicator (*second row*). The simulated output is now based on a multi-sample simulation of size 1000×10, and standard errors, although not shown, are available from the authors. The OSF is, for any $T_n(k)$,

$$OSF_T \equiv \frac{k_0^{(T)}(n)}{n} := \frac{\arg\min_k MSE\big(T_n(k)\big)}{n} ,$$

and, relatively to the Hill estimator $H_n(k)$ in (1.10), the $REFF$ indicator is

$$REFF_T := \sqrt{MSE\Big[H_n\big(k_0^{(H)}(n)\big)\Big] \,/\, MSE\Big[T_n\big(k_0^{(T)}(n)\big)\Big]} .$$

For any value of n, and among the four r.v.'s, the largest $REFF$ (equivalent to smallest MSE) is **<u>in bold and underlined</u>**.

It is clear from Table 1 the overall best performance of ML estimator, whenever (β,ρ) is assumed to be known. Indeed, since $b_{ML}=0$, we were intuitively expecting this type of performance. The choice is not so clear-cut when we consider the estimation of the second order parameters, and either the statistics UH_j or the statistics UH_j^*. Tables 2, 3 and 4 are similar to Table 1, but for the extreme value index estimators UH_j and UH_j^*, $j=0$ or 1 according as $|\rho|\le 1$ or $|\rho > 1$. Again, for any value of n, and among any four estimators of the same type, the largest $REFF$ (equivalent to smallest MSE) is also in **<u>bold and underlined</u>** if it attains the largest value among all estimators, or only in **bold** if it attains the largest value among estimators of the same type.

A few remarks:

- For Fréchet parents, and among the UH_0^* estimators, the best performance is associated to $\overline{ML}_0^*$ for $n < 500$ and to ML_0^* for $n \ge 500$. Among the UH_0 estimators, $\overline{ML}_0$ exhibits the best performance for all n.

- For GP parents with $\gamma = 0.5$, ML_0 exhibits the best performance among the UH_0 statistics. ML_0^* is also the best among the UH_0^* statistics, behaving ML_0^* better than ML_0, for all n.
- For GP parents with $\gamma = 2$, ML_1 exhibits the best performance among the UH_1 statistics. ML_1^* is also the best among the UH_1^* statistics. Now, ML_1^* behaves better than ML_1, for $n \geq 500$ and for $n < 500$ ML_1 performs better than ML_1^*.

Table 1: Simulated OSF/E (first row) and $MSE/REFF$ (second row) at optimal levels of the r.v.'s under study.

n	100	200	500	1000	2000
Fréchet parent, $\gamma = 1$ $(\rho = -1)$					
ML	0.642 / 0.986 0.015 / **1.678**	0.599 / 1.017 0.009 / **1.734**	0.517 / 1.037 0.004 / **1.832**	0.473 / 1.039 0.002 / **1.909**	0.429 / 1.012 0.001 / **2.001**
$\overline{ML}$	0.608 / 0.971 0.016 / 1.647	0.544 / 1.008 0.010 / 1.662	0.477 / 1.045 0.005 / 1.727	0.416 / 1.040 0.003 / 1.782	0.367 / 1.007 0.002 / 1.855
WH	0.580 / 0.960 0.018 / 1.539	0.513 / 1.019 0.011 / 1.577	0.450 / 1.052 0.005 / 1.658	0.395 / 1.041 0.003 / 1.723	0.357 / 1.003 0.002 / 1.805
$\overline{H}$	0.587 / 0.963 0.018 / 1.560	0.537 / 1.012 0.010 / 1.609	0.482 / 1.048 0.005 / 1.710	0.436 / 1.041 0.003 / 1.786	0.379 / 1.008 0.001 / 1.874
GP **parent**, $\gamma = 0.5$ $(\rho = -0.5)$					
ML	0.987 / 0.507 0.002 / **5.813**	0.985 / 0.513 0.001 / **6.567**	0.991 / 0.504 0.000 / **7.831**	0.990 / 0.504 0.000 / **9.184**	0.997 / 0.503 0.000 / **10.487**
$\overline{ML}$	0.295 / 0.565 0.009 / 2.529	0.240 / 0.545 0.006 / 2.561	0.183 / 0.530 0.003 / 2.591	0.157 / 0.531 0.002 / 2.697	0.124 / 0.523 0.001 / 2.753
WH	0.273 / 0.573 0.012 / 2.246	0.221 / 0.566 0.007 / 2.332	0.174 / 0.537 0.004 / 2.419	0.146 / 0.533 0.002 / 2.542	0.117 / 0.530 0.001 / 2.624
$\overline{H}$	0.391 / 0.549 0.007 / 2.918	0.353 / 0.537 0.004 / 3.128	0.302 / 0.536 0.002 / 3.367	0.262 / 0.5200 0.001 / 3.597	0.208 / 0.521 0.001 / 3.835
GP **parent**, $\gamma = 2$ $(\rho = -2)$					
ML	0.990 / 2.065 0.032 / **1.923**	0.994 / 1.921 0.016 / **2.030**	0.995 / 1.992 0.006 / **2.211**	0.993 / 2.011 0.00 / **2.382**	0.999 / 2.015 0.002 / **2.541**
$\overline{ML}$	0.731 / 2.111 0.050 / 1.530	0.677 / 1.956 0.027 / 1.544	0.633 / 2.033 0.012 / 1.573	0.588 / 2.047 0.007 / 1.602	0.549 / 2.063 0.004 / 1.640
WH	0.659 / 2.091 0.058 / 1.420	0.633 / 1.977 0.031 / 1.450	0.576 / 2.036 0.014 / 1.496	0.540 / 2.057 0.008 / 1.528	0.505 / 2.062 0.004 / 1.573
$\overline{H}$	0.669 / 2.103 0.058 / 1.423	0.647 / 1.976 0.030 / 1.470	0.604 / 2.047 0.013 / 1.525	0.574 / 2.053 0.007 / 1.570	0.533 / 2.057 0.004 / 1.622

Table 2: Simulated OSF/E (first row) and $MSE/REFF$ (second row) at optimal levels of the different estimators and r.v.'s under study, for Fréchet parents with $\gamma = 1$ ($\rho = -1$, $\beta = 0.5$).

n	100	200	500	1000	2000
H	0.326 / 1.026 0.044 / 1.000	0.281 / 1.069 0.026 / 1.000	0.222 / 1.056 0.013 / 1.000	0.174 / 1.055 0.008 / 1.000	0.138 / 1.031 0.005 / 1.000
ML_0	0.569 / 0.820 0.037 / 1.084	0.592 / 0.966 0.021 / 1.113	0.826 / 0.977 0.010 / 1.185	0.808 / 1.010 0.005 / 1.269	0.999 / 0.985 0.003 / 1.402
$\overline{ML}_0$	0.847 / 0.959 0.019 / **<u>1.518</u>**	0.802 / 1.027 0.012 / **<u>1.485</u>**	0.758 / 1.008 0.006 / **<u>1.538</u>**	0.727 / 1.026 0.003 / **<u>1.641</u>**	0.709 / 0.998 0.002 / **<u>1.766</u>**
WH_0	0.816 / 0.963 0.020 / 1.494	0.756 / 1.014 0.012 / 1.467	0.702 / 1.004 0.006 / 1.517	0.678 / 1.030 0.003 / 1.616	0.650 / 1.001 0.001 / 1.731
$\overline{H}_0$	0.877 / 0.951 0.024 / 1.358	0.841 / 1.005 0.015 / 1.331	0.819 / 0.998 0.007 / 1.376	0.808 / 1.026 0.004 / 1.469	0.808 / 0.973 0.002 / 1.576
ML_0^*	0.947 / 0.849 0.037 / 1.092	0.920 / 0.973 0.020 / 1.139	0.870 / 0.992 0.009 / **1.239**	0.855 / 1.019 0.005 / **1.349**	0.834 / 0.979 0.002 / **1.480**
$\overline{ML}_0^*$	0.858 / 0.988 0.027 / **1.277**	0.787 / 1.054 0.017 / **1.234**	0.676 / 1.064 0.009 / 1.222	0.603 / 1.058 0.005 / 1.230	0.530 / 1.001 0.003 / 1.246
WH_0^*	0.811 / 0.992 0.030 / 1.211	0.736 / 1.062 0.018 / 1.194	0.647 / 1.069 0.009 / 1.194	0.567 / 1.057 0.006 / 1.208	0.511 / 1.003 0.003 / 1.224
H_0^*	0.856 / 0.973 0.031 / 1.191	0.795 / 1.048 0.019 / 1.183	0.711 / 1.059 0.009 / 1.205	0.643 / 1.057 0.005 / 1.231	0.579 / 0.994 0.003 / 1.261

Table 3: Simulated OSF/E (first row) and $MSE/REFF$ (second row) at optimal levels of the different estimators and r.v.'s under study, for GP parents with $\gamma = 0.5$ ($\rho = -0.5$, $\beta = 1$).

n	100	200	500	1000	2000
H	0.103 / 0.742 0.058 / 1.000	0.077 / 0.646 0.037 / 1.000	0.051 / 0.632 0.020 / 1.000	0.040 / 0.602 0.014 / 1.000	0.028 / 0.585 0.009 / 1.000
ML_0	0.306 / 0.636 0.023 / **1.572**	0.216 / 0.633 0.017 / **1.474**	0.107 / 0.606 0.011 / **1.383**	0.076 / 0.583 0.008 / **1.339**	0.051 / 0.558 0.006 / **1.274**
$\overline{ML}_0$	0.211 / 0.674 0.029 / 1.418	0.149 / 0.618 0.019 / 1.383	0.101 / 0.606 0.011 / 1.338	0.073 / 0.588 0.008 / 1.310	0.049 / 0.558 0.006 / 1.258
WH_0	0.202 / 0.669 0.029 / 1.416	0.144 / 0.614 0.019 / 1.382	0.100 / 0.607 0.011 / 1.336	0.071 / 0.586 0.008 / 1.308	0.049 / 0.558 0.006 / 1.257
$\overline{H}_0$	0.234 / 0.641 0.029 / 1.418	0.165 / 0.640 0.019 / 1.384	0.103 / 0.607 0.011 / 1.339	0.073 / 0.588 0.008 / 1.310	0.049 / 0.557 0.006 / 1.257
ML_0^*	0.795 / 0.652 0.022 / **<u>1.612</u>**	0.636 / 0.628 0.016 / **<u>1.525</u>**	0.421 / 0.602 0.010 / **<u>1.452</u>**	0.310 / 0.578 0.007 / **<u>1.420</u>**	0.240 / 0.568 0.005 / **<u>1.370</u>**
$\overline{ML}_0^*$	0.449 / 0.720 0.049 / 1.090	0.350 / 0.654 0.030 / 1.114	0.251 / 0.610 0.015 / 1.148	0.192 / 0.600 0.010 / 1.185	0.140 / 0.579 0.006 / 1.199
WH_0^*	0.450 / 0.732 0.051 / 1.068	0.334 / 0.649 0.030 / 1.110	0.245 / 0.612 0.015 / 1.149	0.191 / 0.600 0.010 / 1.187	0.138 / 0.576 0.006 / 1.205
H_0^*	0.464 / 0.697 0.040 / 1.211	0.389 / 0.634 0.024 / 1.240	0.289 / 0.600 0.012 / 1.261	0.226 / 0.599 0.009 / 1.280	0.169 / 0.558 0.006 / 1.271

Table 4: Simulated *OSF/E* (first row) and *MSE/REFF* (second row) at optimal levels of the different estimators and r.v.'s under study, for *GP* parents with $\gamma = 2$ ($\rho = -2,\ \beta = 1$).

n	100	200	500	1000	2000
H	0.415 / 2.179 0.117 / 1.000	0.359 / 1.968 0.064 / 1.000	0.319 / 2.018 0.030 / 1.000	0.290 / 2.068 0.018 / 1.000	0.251 / 2.069 0.010 / 1.000
ML_1	0.817 / 2.184 0.071 / <u>**1.282**</u>	0.647 / 2.012 0.043 / <u>**1.221**</u>	0.663 / 2.048 0.021 / **1.194**	0.657 / 2.077 0.013 / **1.173**	1.000 / 2.094 0.007 / **1.180**
$\overline{ML}_1$	0.631 / 2.140 0.079 / 1.215	0.558 / 2.008 0.046 / 1.184	0.478 / 2.044 0.022 / 1.168	0.399 / 2.050 0.013 / 1.158	0.358 / 2.040 0.008 / 1.153
WH_1	0.623 / 2.155 0.081 / 1.197	0.554 / 2.024 0.047 / 1.171	0.470 / 2.048 0.023 / 1.159	0.396 / 2.051 0.013 / 1.153	0.349 / 2.041 0.008 / 1.149
$\overline{H}_1$	0.618 / 2.167 0.083 / 1.186	0.545 / 2.041 0.047 / 1.165	0.470 / 2.050 0.023 / 1.156	0.396 / 2.051 0.013 / 1.152	0.349 / 2.041 0.008 / 1.148
ML_1^*	0.990 / 2.194 0.072 / **1.272**	0.935 / 2.000 0.044 / **1.211**	0.828 / 2.034 0.021 / <u>**1.204**</u>	0.768 / 2.077 0.012 / <u>**1.197**</u>	0.681 / 2.055 0.007 / <u>**1.191**</u>
$\overline{ML}_1^*$	0.751 / 2.199 0.089 / 1.143	0.696 / 1.993 0.050 / 1.129	0.624 / 2.044 0.024 / 1.123	0.571 / 2.065 0.014 / 1.125	0.519 / 2.041 0.008 / 1.130
WH_1^*	0.711 / 2.240 0.100 / 1.079	0.652 / 2.002 0.054 / 1.087	0.595 / 2.038 0.025 / 1.098	0.548 / 2.070 0.014 / 1.105	0.510 / 2.045 0.008 / 1.115
$\overline{H}_1^*$	0.710 / 2.240 0.10 / 1.0780	0.657 / 2.001 0.054 / 1.088	0.604 / 2.041 0.025 / 1.101	0.561 / 2.071 0.014 / 1.109	0.513 / 2.041 0.008 / 1.120

4.4. An overall conclusion

The main advantage of the estimators UH_j, and particularly of the ML_j estimators in this paper, the ones with an overall better performance, lies on the fact that we may estimate β and ρ adequately through $\hat{\beta}$ and $\hat{\rho}$ 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. These new estimators work indeed better than the Hill estimator for all values of k, contrarily to the alternatives so far available in the literature, like the alternatives UH_j^*, $j = 0$ or 1, also considered in this paper for comparison.

5. CASE-STUDIES IN THE FIELDS OF FINANCE AND INSURANCE

5.1. Euro-UK Pound daily exchange rates

We shall first consider the performance of the above mentioned estimators in the analysis of the Euro-UK Pound daily exchange rates from January 4, 1999 until December 14, 2004. This data has been collected by the European System of Central Banks, and was obtained from `http://www.bportugal.pt/rates/cambtx/`. In Figure 4 we picture the Daily Exchange Rates x_t over the above mentioned period and the Log-Returns, $r_t = 100 \times (\ln x_t - \ln x_{t-1})$, the data to be analyzed. Indeed, although conscious that the log-returns of any financial time-series are not i.i.d., we also know that the semi-parametric behavior of estimators of rare event parameters may be generalized to weak dependent data (see Drees, 2002, and references therein). Semi-parametric estimators of extreme events' parameters, devised for i.i.d. processes, are usually based on the tail empirical process, and remain consistent and asymptotically normal in a large class of weakly dependent data.

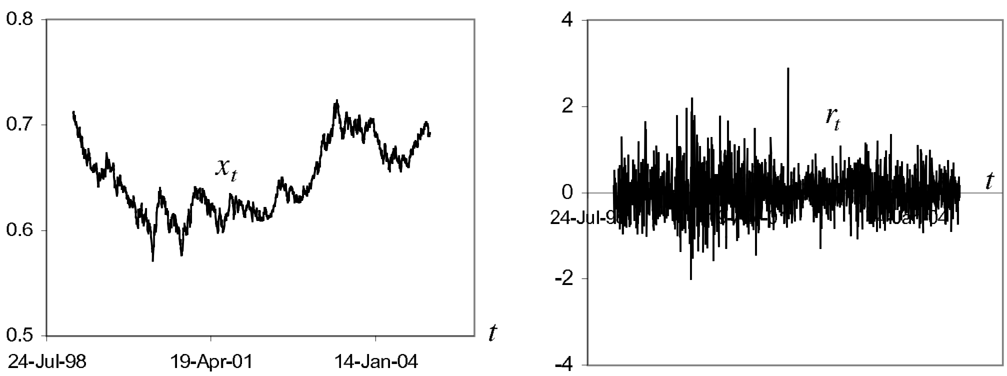

Figure 4: Daily Exchange Rates (*left*) and Daily Log-Returns (*right*) on Euro-UK Pound Exchange Rate.

The histogram in Figure 5 points to a heavy right tail. Indeed, the empirical counterparts of the usual skewness and kurtosis coefficients are $\hat{\beta}_1 = 0.424$ and $\hat{\beta}_2 = 1.835$, clearly greater than 0, the target value for an underlying normal parent.

In Figure 6, and working with the $n_0 = 725$ positive log-returns, we now picture the sample paths of $\hat{\rho}(k;\tau)$ in (3.7) for $\tau = 0$, and 1 (*left*), as functions of k. The sample paths of the ρ-estimates associated to $\tau = 0$ and $\tau = 1$ lead us

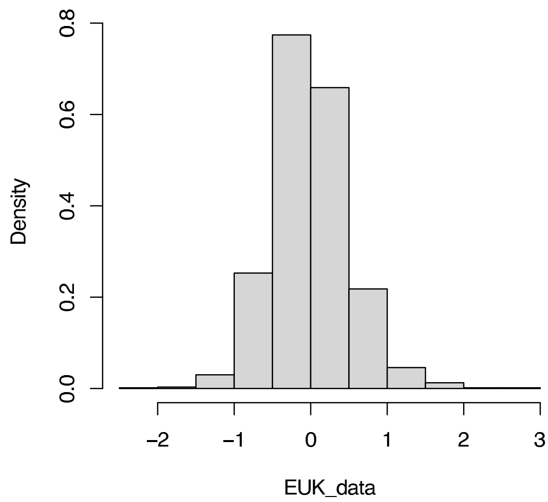

Figure 5: Histogram of the Daily Log-Returns on the Euro-UK Pound.

to choose, on the basis of any stability criterion for large values of k, the estimate associated to $\tau = 0$. In Figure 6 we thus present the associated second order parameters estimates, $\hat{\rho}_0 = \hat{\rho}_0(721) = -0.65$ (*left*) and $\hat{\beta}_0 = \hat{\beta}_{\hat{\rho}_0}(721) = 1.03$, together with the sample paths of $\hat{\beta}(k; \hat{\rho}_0)$ in (1.14), for $\tau = 0$ (*center*). The sample paths of the classical Hill estimator in (1.10) (H) and of three of reduced-bias, second order extreme value index estimates discussed in this paper, associated to $\hat{\rho}_0 = -0.65$ and $\hat{\beta}_0 = 1.03$, are also pictured in Figure 6 (*right*). We do not picture the statistic WH_0 because that statistic practically overlaps $\overline{ML}_0$.

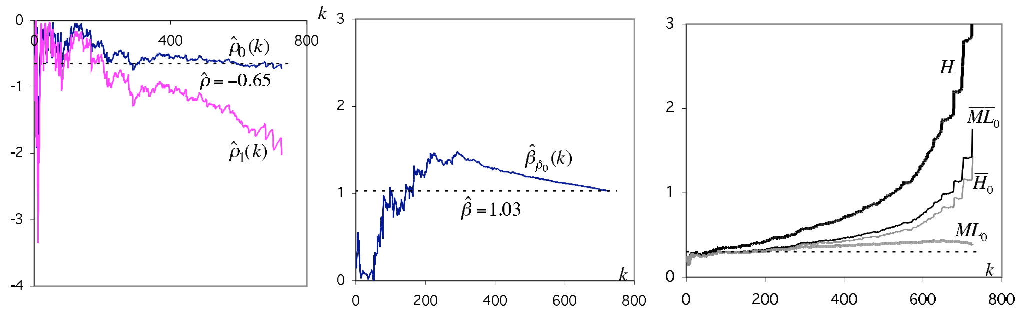

Figure 6: Estimates of the second order parameter ρ (*left*), of the second order parameter β (*center*) and of the extreme value index (*right*), for the Daily Log-Returns on the Euro-UK Pound.

The Hill estimator exhibits a relevant bias, as may be seen from Figure 6, and we are for sure a long way from the strict Pareto model. The other estimators, ML_0, $\overline{ML}_0$ and $\overline{H}_0$, which are "asymptotically unbiased", reveal without doubt a bias much smaller than that of the Hill. All these statistics enable us to take a decision upon the estimate of γ to be used, with the help of any stability criterion, but the *ML* statistic is without doubt the one with smallest bias, among the statistics considered. More important than this: we know that any estimate considered on the basis of $ML_0(k)$ (or any of the other three reduced-bias statistics) performs for sure better than the estimate based on $H(k)$ for any level k. Here, we represent the estimate $\widehat{\gamma} \equiv \widehat{\gamma}_{ML} = 0.30$, the median of the *ML* estimates, for thresholds k between $\left[n_0^{-2\hat{\rho}/(1-2\hat{\rho})}/4\right] = 10$ and $\left[4 \times n_0^{-2\hat{\rho}/(1-2\hat{\rho})}\right] = 165$, chosen in an heuristic way. If we use this same criterion on the estimates $\overline{ML}$, *WH* and $\overline{H}$ we are also led to the same estimate, $\widehat{\gamma}_{\overline{ML}} \equiv \widehat{\gamma}_{WH} \equiv \widehat{\gamma}_{\overline{H}} = 0.30$. The development of adequate techniques for the adaptive choice of the optimal threshold for this type of second order reduced-bias extreme value index estimators is needed, being indeed an interesting topic of research, but is outside the scope of the present paper.

5.2. Automobile claims

We shall next consider an illustration of the performance of the above mentioned estimators, through the analysis of automobile claim amounts exceeding 1,200.000 Euros, over the period 1988–2001, and gathered from several European insurance companies co-operating with the same re-insurer (Secura Belgian Re). This data set has already been studied, for instance, in Beirlant et al. (2004). Figure 7 is similar to Figure 5, but for the Secura data. It is now quite clear the heaviness of the right tail. The empirical skewness and kurtosis coefficients are $\hat{\beta}_1 = 2.441$ and $\hat{\beta}_2 = 8.303$. Here, the existence of left-censoring is also clear, begin the main reason for the high skewness and kurtosis values.

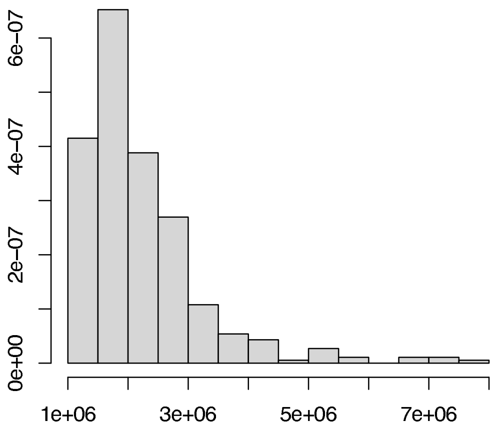

Figure 7: Histogram or the Secura data.

Finally, in Figure 8, working with the $n = 371$ automobile claims exceeding 1,200.000 Euro, we present the sample path of the $\hat{\rho}_\tau$ (*left*), $\hat{\rho}_\tau$ (*center*) estimates, as function of k, for $\tau = 0$ and $\tau = 1$, together with the sample paths of estimates of the extreme value index γ, provided by the Hill estimator, H, the M-estimator and the $\overline{M}$ estimator (*right*).

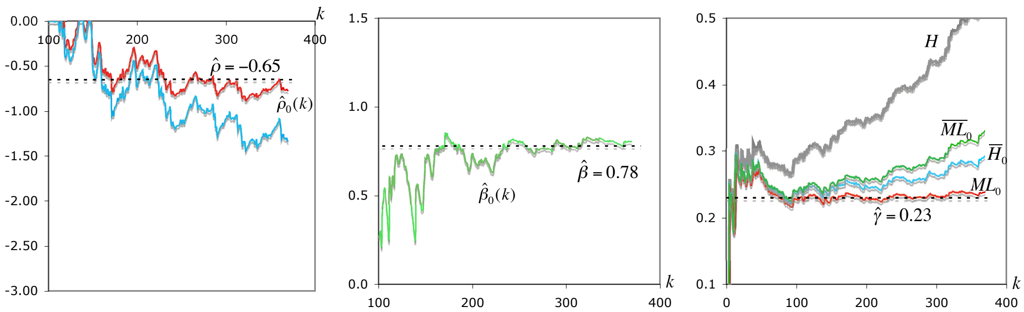

Figure 8: Estimates of the second order parameter ρ (*left*) and of the extreme value index γ (*right*) for the automobile claims.

Again, the ML_0 statistic is the one exhibiting the best performance, leading us to the estimate $\widehat{\gamma} = 0.23$.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Bingham, N.H.; Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press.

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

[3] Beirlant, J.; Goegebeur, Y.; Segers, J. and Teugels, J. *Statistics of Extremes. Theory and Applications*, Wiley, New York.

[4] Caeiro, F. and Gomes, M.I. (2004). A new class of estimators of the "scale" second order parameter. To appear in *Extremes*.

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

[6] Drees, H. (1998). A general class of estimators of the tail index, *J. Statist. Planning and Inference*, **98**, 95–112.

[7] Drees, H. (2002). *Tail empirical processes under mixing conditions.* In "Empirical Process Techniques for Dependent Data" (Dehling *et al.*, Eds.), Birkhăuser, Boston, 325–342.

[8] Drees, H. and Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation, *Stoch. Proc. and Appl.*, **75**, 149–172.

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

[10] 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.

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

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

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

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

[15] Gomes, M.I.; Haan, L. de and Rodrigues, L. (2004b). *Tail index estimation through accommodation of bias in the weighted log-excesses*, Notas e Comunicações, CEAUL, 14/2004 (submitted).

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

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

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

[19] 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.

[20] Gomes, M.I.; Miranda, C. and Pereira, H. (2005b). Revisiting the role of the Jackknife methodology in the estimation of a positive tail index, *Comm. in Statistics — Theory and Methods*, **34**, 1–20.

[21] Gomes, M.I.; Miranda, C. and Viseu, C. (2006). Reduced bias tail index estimation and the Jackknife methodology, *Statistica Neerlandica*, **60**, 4, 1–28.

[22] Gomes, M.I. and Oliveira, O. (2000). The bootstrap methodology in Statistical Extremes — choice of the optimal sample fraction, *Extremes*, **4**, 4, 331–358.

[23] Gomes, M.I. and Pestana, D. (2004). A simple second order reduced-bias extreme value index estimator. To appear in *J. Statist. Comput. and Simulation*.

[24] Gomes, M.I. and Pestana, D. (2005). A sturdy reduced-bias extreme quantile (*VaR*) estimator. To appear in *J. American Statist. Assoc.*.

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

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

[27] Peng, L. (1998). Asymptotically unbiased estimator for the extreme-value index, *Statistics and Probability Letters*, **38**, 2, 107–115.