Revisiting the second order reduced bias "maximum likelihood" tail index estimators*

M. Ivette Gomes

Universidade de Lisboa, D.E.I.O. and C.E.A.U.L.

M. João Martins and Manuela Neves

Universidade Técnica de Lisboa

October 19, 2005

Abstract. Classical tail 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 reduced bias' tail index estimation, based on a "classical" exponential second order regression model, applied to the scaled log-spacings of the top order statistics. The estimation of a "scale" and a "shape" second order parameters in the bias, at a level k_1 of a larger order than that of the level k at which we compute the tail index estimators, enables us to keep the asymptotic variance of the new estimators of a positive tail index γ equal to the asymptotic variance of the Hill estimator, the maximum likelihood estimator of γ, under a strict Pareto model. 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 tail index estimation. The asymptotic distributional properties of the proposed class of γ-estimators are derived and the estimators are compared with other similar alternative

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

estimators of γ recently introduced in the literature, not only asymptotically, but also for finite samples through Monte Carlo techniques. A case-study in the field of finance will illustrate the performance of the new second order reduced bias' tail index estimators.

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

Keywords and phrases. *Statistics of Extremes*, *Semi-parametric estimation*, *Bias estimation*, *Heavy tails*, *Maximum likelihood.*

1 Introduction and motivation for the new class of tail 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 $\{-1/\gamma\}$, $\gamma > 0$, i.e., if and only if $\overline{F} \in RV_{-1/\gamma}$, 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 $F^{\leftarrow}(x) = \inf\{y : F(y) \geq x\}$, is of regular variation with index γ, i.e.,

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

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

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

but with $\gamma > 0$, and we write $F \in \mathcal{D}_{\mathcal{M}}(EV_\gamma)$, $\gamma > 0$. The parameter γ is the *tail 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), and is the non-positive parameter appearing in the limiting relation

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

which we assume to hold for every $x > 0$, and where $|A(t)|$ must then be of regular variation with index ρ (Geluk and de Haan, 1987). We shall assume everywhere that $\rho < 0$. This condition 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, like the Fréchet, the Generalized Pareto, the Burr and the Student's t.

Remark 1.1. *Note that $\mathcal{D}_{\mathcal{M}}(EV_0)$ also contains d.f.'s with tails quite similar to the tails of the models in $\mathcal{D}_{\mathcal{M}}(EV_\gamma)$, $\gamma > 0$. Such a kind of behaviour was detected a long time ago by Fisher and Tippett* (1928)*, who have first spoken about the so-called "penultimate" behaviour of maxima of models in $\mathcal{D}_{\mathcal{M}}(EV_0)$, or equivalently, the penultimate behaviour of excesses over a high threshold, further studied by Gomes* (1984)*, Gomes and de Haan* (1999)*, Kaufmann* (2000)*, Raoult and Worms* (2003) *and Diebolt and Guillou* (2005)*, among others. Very popular tails in insurance and finance are tails of the type, $1 - F(x) = \exp\{-H(x)\}$, $\quad H \in RV_{1/\theta}$, $\theta > 1$. In a context of Extreme Value Theory we have then a tail index $\gamma = 0$ and a second order parameter $\rho = 0$, i.e., we are working with tails in the domain of attraction for maxima of Gumbel's law $\Lambda(x) \equiv EV_0(x) = \exp(-\exp(-x))$, $x \in \mathbb{R}$, which exhibit a penultimate behaviour, looking more similar to Pareto tails than to exponential tails. These distributions belong to the class of sub-exponential models, another possible class of heavy-tailed models, and for relations among the different classes of heavy-*

tailed distributions see Embrechts et al. (1997)*, section* 1.4.

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

$$U(t) = Ct^{\gamma}\left(1 + Dt^{\rho} + o\left(t^{\rho}\right)\right), \quad as \quad t \to \infty,$$

$C > 0$*,* $D \neq 0$*,* $\rho < 0$*,* (1.2) *holds and we may choose* $A(t) = \rho\ D\ t^{\rho}$*.*

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 ρ-estimation, as well as in Gomes *et al.* (2004a), in a paper on the estimation of a positive tail index γ, we shall further specify the term $\{o\left(t^{\rho}\right)\}$ in Hall's class of models, and we shall sometimes assume to be working with a Pareto-type class of models with a quantile function

$$U(t) = Ct^{\gamma}\left(1 + D_1 t^{\rho} + D_2 t^{2\rho} + o\left(t^{2\rho}\right)\right), \tag{1.3}$$

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

$$A(t) = \rho\ D_1\ t^{\rho} =: \gamma\ \beta\ t^{\rho}, \quad \beta \neq 0, \quad \rho < 0, \tag{1.4}$$

and, with

$$B(t) = \left(2D_2/D_1 - D_1\right)\ t^{\rho} =: \beta'\ t^{\rho}, \tag{1.5}$$

we may write

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

Remark 1.3. *As noticed before in Caeiro et al.* (2004)*, most of the common heavy-tailed d.f.'s, like the Fréchet, the Generalized Pareto, the Burr and the Student's t belong to the class of models in* (1.3)*.*

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

$$k = k_n \to \infty, \quad k_n = o(n), \text{ as } n \to \infty, \tag{1.6}$$

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

$$V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n}, \quad 1 \leq i \leq k < n, \tag{1.7}$$

and the scaled log-spacings,

$$U_i := i\left\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\right\}, \quad 1 \leq i \leq k < n, \tag{1.8}$$

where $X_{i:n}$ denotes, as usual, the i-th ascending order statistic (o.s.), $1 \leq i \leq n$, associated to an independent, identically distributed (i.i.d.) random sample $(X_1, X_2, \cdots, X_n)$. We have a strong obvious link between the log-excesses and the scaled log-spacings provided by the equation, $\sum\limits_{i=1}^{k} V_{ik} = \sum\limits_{i=1}^{k} U_i$.

It is well known that for intermediate k, and whenever we are under the first order framework in (1.1), the log-excesses V_{ik}, $1 \leq i \leq k$, in (1.7), 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, $1 \leq i \leq k$, in (1.8), are approximately i.i.d. and exponential with mean value γ. Consequently the Hill estimator of γ (Hill, 1975),

$$H(k) \equiv H_n(k) \; = \; \frac{1}{k}\sum_{i=1}^{k} V_{ik} \; = \; \frac{1}{k}\sum_{i=1}^{k} U_i, \tag{1.9}$$

is consistent for the estimation of γ whenever (1.1) holds and k is intermediate, i.e., (1.6) holds. Under the second order framework in (1.2) the asymptotic distributional representation

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

holds, where $Z_k^{(1)} = \sqrt{k}\left(\sum\limits_{i=1}^{k} E_i/k - 1\right)$, with $\{E_i\}$ i.i.d. standard exponential random variables (r.v.'s), is an asymptotically standard normal r.v.

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 $U_i \approx \gamma\,(1+b(n/k)(k/i)^\rho)\,E_i$ and $U_i \approx \gamma \exp\left(\beta\ (n/i)^\rho\right) E_i$, respectively, $1 \le i \le k$. They then proceed to the joint estimation of the three unknown parameters or functionals at the same level k. Considering also the scaled log-spacings U_i in (1.8) to be approximately exponential with mean value $\mu_i = \gamma\ \exp\left(\beta (n/i)^\rho\right)$, $1 \le i \le k$, $\beta \neq 0$, Gomes and Martins (2002) advance with the "external" estimation of the second order parameter ρ, together with a first order approximation for the maximum likelihood β-estimator. They then obtain "quasi-maximum likelihood" explicit estimators of γ and β, and through that "external" estimation of ρ, are then able to reduce the asymptotic variance of the tail index estimator proposed. With the notation,

$$d_k(\alpha) = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1},\ D_k(\alpha) = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1} U_i, \tag{1.10}$$

for any real $\alpha \ge 1$ [$D_k(1) \equiv H(k)$ in (1.9)], and with $\widehat{\rho}$ any consistent estimator of ρ, such a tail index estimator is

$$\widehat{\gamma}_n^{ML}(k) = D_k(1) - \widehat{\beta}(k;\widehat{\rho})\ \left(\frac{n}{k}\right)^{\widehat{\rho}}\ D_k\left(1-\widehat{\rho}\right), \tag{1.11}$$

with

$$\widehat{\beta}(k;r) := \left(\frac{k}{n}\right)^{r} \frac{d_k(1-r)\times D_k(1) - D_k(1-r)}{d_k(1-r)\times D_k(1-r) - D_k(1-2r)}, \tag{1.12}$$

for an adequate choice $r \le 0$. This means that β in (1.4), which is also a second order parameter, is estimated at the same level k at which the γ-estimation is performed, being the estimator $\widehat{\beta}(k;\widehat{\rho})$ plugged in the tail index estimator in (1.11). We here propose an "external" estimation of both β and ρ, through $\widehat{\beta}$ and $\widehat{\rho}$, respectively, both using a larger number of top o.s.'s than the one used for the tail index estimation. We shall thus consider the estimator

$$ML_{\widehat{\beta},\widehat{\rho}}(k) := D_k(1) - \widehat{\beta}\ \left(\frac{n}{k}\right)^{\widehat{\rho}}\ D_k\left(1-\widehat{\rho}\right), \tag{1.13}$$

for adequate consistent estimators $\widehat{\beta}$ and $\widehat{\rho}$ of the second order parameters β and ρ, respectively, to be specified in subsection 2.2. of this paper.

Remark 1.4. *Note that* $\widehat{\gamma}_n^{ML}(k)$ *in* (1.11) *is equal to* $ML_{\widehat{\beta}(k;\widehat{\rho}),\ \widehat{\rho}}(k)$ *in* (1.13), *when both* γ *and* β *are estimated at the same level* k.

Remark 1.5. *Note also that the estimator in* (1.13) *has been inspired in the recent papers of Gomes et al.* (2004*b*) *and Caeiro et al.* (2004). *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* (1.4), *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 estimator in* (1.13). *The reduced bias' tail index estimator in Gomes et al.* (2004*b*) *is based on a linear combination of the log-excesses* V_{ik} *in* (1.7), *and is given by*

$$WH_{\widehat{\beta},\widehat{\rho}}(k) := \frac{1}{k}\sum_{i=1}^{k} e^{-\widehat{\beta}\ (n/k)^{\widehat{\rho}}\ \psi_{\widehat{\rho}}(i/k)}\ V_{ik}, \qquad \psi_\rho(x) = -\frac{x^{-\rho}-1}{\rho\ln x}, \tag{1.14}$$

with the notation WH *standing for Weighted Hill estimator. Caeiro et al.* (2004) *consider the estimator*

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

where the dominant component of the bias of Hill's estimator $H(k)$ *in* (1.9), *given by* $A(n/k)/(1-\rho) = \gamma\ \beta\ (n/k)^\rho/(1-\rho)$, *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. As before, and both in* (1.14) *and* (1.15), $\widehat{\beta}$ *and* $\widehat{\rho}$ *are adequate consistent estimators of the second order parameters* β *and* ρ, *respectively.*

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 (1.13), similar to the one provided before for the estimators in (1.14) (Gomes *et al.*, 2004b) and (1.15) (Caeiro *et al.*, 2004). We then shall briefly review the estimation of the two second order parameters β and ρ. Next, we consider the derivation of the asymptotic behaviour of the $ML_{\widehat{\beta},\widehat{\rho}}(k)$ class of estimators in (1.13), estimating β and ρ at a level k_1 larger than the value k 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 3, and through the use of simulation techniques, we shall exhibit the performance of the ML estimator in (1.13), comparatively to the other *"Unbiased Hill"* (UH) estimators, WH and $\overline{H}$, in (1.14) and (1.15), respectively, to the classical Hill estimator H in (1.9) and to the "asymptotically unbiased" estimator $\widehat{\gamma}_n^{ML}(k)$ in (1.11) $\Big[$or equivalently, the $ML_{\widehat{\beta}(k;\widehat{\rho}),\ \widehat{\rho}}$ estimator in (1.13)$\Big]$, studied in Gomes and Martins (2002). Whenever there is no distinction between the *"Unbiased Hill"* estimators in (1.13), (1.14) and (1.15), or the corresponding r.v.'s, we shall often use the notation UH, generically denoting either ML or WH or $\overline{H}$. We shall restrict ourselves to an "external" estimation of the second order parameter ρ at a level k_1 larger than the level k on which we base the tail index estimation. Section 4 is devoted to the illustration of the behaviour of these estimators for the Daily Log-Returns of the *Euro* against the *UK* Pound. Finally, in section 5, we shall provide the proofs of the main results in section 2.

2 Asymptotic behaviour of the reduced bias' tail index estimators

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:

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

2.1 Asymptotic behaviour of the reduced bias tail index estimators — known β and ρ

If we assume that only the tail index parameter γ is unknown:

Theorem 2.1. *Under the second order framework in* (1.2)*, further assuming that $A(t)$ may be chosen as in* (1.4)*, and for levels k such that* (1.6) *holds, we get, for $ML_{\beta,\rho}(k)$, an asymptotic distributional representation of the type*

$$ML_{\beta,\rho}(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}} \, \overline{Z}_k^{(1)} + o_p(A(n/k),$$

where $\overline{Z}_k^{(1)}$ is the asymptotically standard normal r.v. in (2.1) *for $\alpha = 1$, being $ML_{\widehat{\beta},\widehat{\rho}}(k)$ given in* (1.13)*. Consequently, $\sqrt{k}\ (ML_{\beta,\rho}(k) - \gamma)$ 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 (1.3)*, we may further specify the term $o_p(A(n/k))$, writing*

$$ML_{\beta,\rho}(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}} \, \overline{Z}_k^{(1)} + \frac{(\beta' - \beta)A^2(n/k)}{\gamma\ \beta(1-2\rho)}\ (1 + o_p(1), \qquad (2.2)$$

with β and β' given in (1.4) *and* (1.5)*, respectively. Consequently, even if $\sqrt{k}\ A(n/k) \to \infty$, with $\sqrt{k}\ A^2(n/k) \to \lambda_{_A}$, finite, $\sqrt{k}\ (ML_{\beta,\rho}(k) - \gamma)$ is asymptotically normal with variance equal to γ^2 and an asymptotic bias equal to*

$$b_{_{ML}} = (\beta' - \beta)\ \lambda_{_A}/(\gamma\ \beta(1-2\rho)). \qquad (2.3)$$

Remark 2.1. *For the Burr model, with d.f. $F(x) = 1 - \left(1 + x^{-\rho/\gamma}\right)^{1/\rho},\ x \geq 0$, we have $U(t) = t^\gamma\,(1 - t^\rho)^{-\gamma/\rho} = t^\gamma\left(1 + \gamma\ t^\rho/\rho + \gamma(\gamma+\rho)t^{2\rho}/(2\ \rho^2) + o(t^{2\rho})\right)$, for $t \geq 1$. Consequently,* (1.3) *holds with $D_1 = \gamma/\rho$, $D_2 = \gamma(\gamma+\rho)/(2\rho^2)$, $\beta' = \beta = 1$ and $b_{_{ML}} = 0$. A similar results holds for the Generalized Pareto d.f. $F(x) = 1 - (1 + \gamma\ x)^{-1/\gamma},\ 1 + \gamma\ x \geq 0$. For this d.f., $U(t) = (t^\gamma - 1)/\gamma$, and* (1.3) *holds with $\rho = -\gamma$, $D_1 = -1$ and $D_2 = 0$. Hence $\beta = \beta' = 1$ and $b_{_{ML}} = 0$.*

2.2 A brief review of the second order parameters' estimators

2.2.1 The estimation of ρ

We have nowadays a general class of ρ-estimators which work well in practice, the one in Fraga Alves *et al.* (2003). We shall 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, the number of top o.s.'s used, for a wide range of large k-values. Such a class of estimators has been first parameterised in a tuning parameter $\tau > 0$, but τ may be more generally considered as a real number (Caeiro and Gomes, 2004), and is defined as,

$$\widehat{\rho}(k;\tau) \equiv \widehat{\rho}_\tau(k) \equiv \widehat{\rho}_n^{(\tau)}(k) := -\left|\frac{3(T_n^{(\tau)}(k)-1)}{T_n^{(\tau)}(k)-3}\right|, \tag{2.4}$$

where

$$T_n^{(\tau)}(k) := \begin{cases} \dfrac{\left(M_n^{(1)}(k)\right)^{\tau} - \left(M_n^{(2)}(k)/2\right)^{\tau/2}}{\left(M_n^{(2)}(k)/2\right)^{\tau/2} - \left(M_n^{(3)}(k)/6\right)^{\tau/3}} & \text{if} \quad \tau \neq 0 \\ \\ \dfrac{\ln\left(M_n^{(1)}(k)\right) - \frac{1}{2}\ln\left(M_n^{(2)}(k)/2\right)}{\frac{1}{2}\ln\left(M_n^{(2)}(k)/2\right) - \frac{1}{3}\ln\left(M_n^{(3)}(k)/6\right)} & \text{if} \quad \tau = 0 \end{cases},$$

with

$$M_n^{(j)}(k) := \frac{1}{k}\sum_{i=1}^{k}\left\{\ln\frac{X_{n-i+1:n}}{X_{n-k:n}}\right\}^j, \; j \geq 1 \quad \left[M_n^{(1)} \equiv H \text{ in } (1.9)\right].$$

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

Proposition 2.1 (Fraga Alves *et al.*, 2003)**.** *Under the second order framework in* (1.2)*, if* (1.6) *holds, and if* $\sqrt{k}\; A(n/k) \to \infty$*, as* $n \to \infty$*, the statistics* $\widehat{\rho}_n^{(\tau)}(k)$ *in* (2.4) *converge in probability towards* ρ*, as* $n \to \infty$*, for any real* τ*. Moreover, for models in* (1.3)*, if we further assume that* $\sqrt{k}\; A^2(n/k) \longrightarrow \lambda_{_A}$*, finite,* $\widehat{\rho}_n^{(\tau)}(k)$ *is asymptotically normal, and* $\left\{\widehat{\rho}_n^{(\tau)}(k)-\rho\right\} = O_p\left(1/\left(\sqrt{k}\; A(n/k)\right)\right)$*. If* $\sqrt{k}\; A^2(n/k) \to \infty$*,* $\left\{\widehat{\rho}_n^{(\tau)}(k)-\rho\right\} = O_p\left(A(n/k)\right)$*.*

Remark 2.2. *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 $k_1 = \min(n-1, [2n/\ln\ln n])$, to estimate ρ. Here, and inspired on the results in Gomes et al.* (2004*b*) *for the estimator in* (1.14)*, we consider the level*

$$k_1 = \min\left(n-1, \left[n^{1-\epsilon}\right]+1\right), \quad \textit{for some } \epsilon > 0, \textit{ small} \tag{2.5}$$

(*not chosen in any optimal way*) *and where* $[x]$ *denotes, as usual, the integer part of x. We shall also consider the tuning parameters $\tau = 0$ for the region $\rho \in [-1, 0)$ and $\tau = 1$ for the region $\rho \in (-\infty, -1)$. We however think that practitioners should not choose blindly the value of τ in* (2.4)*. It is sensible to draw some sample paths of $\widehat{\rho}(k;\tau)$, as functions of k and for a few τ-values, electing the value of $\tau \equiv \tau_U$ which provides the highest stability for large k, by means of any stability criterion, like the ones suggested in Gomes et al.* (2005)*, Gomes et al.* (2004*a*) *and Gomes and Pestana* (2004)*. Anyway, in all the Monte Carlo simulations we have considered the level k_1 in* (2.5)*, with $\epsilon = 0.001$, and the ρ-estimators*

$$\widehat{\rho}_\tau := -\left|\frac{3\left(T_n^{(\tau)}(k_1)-1\right)}{T_n^{(\tau)}(k_1)-3}\right|, \qquad \tau = \begin{cases} 0 & \textit{if} \quad \rho \geq -1 \\ 1 & \textit{if} \quad \rho < -1 \end{cases}. \tag{2.6}$$

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 $\tau = \tau_U$ mentioned before, is much more relevant than the choice of the level k_1, in the ρ-estimation and everywhere in the paper.*

From now on we shall use the notation $\widehat{\rho} \equiv \widehat{\rho}_\tau = \widehat{\rho}(k_1;\tau)$ for any of the estimators in (2.4) computed at the level k_1 in (2.5).

Remark 2.3. *When we consider the level k_1 in* (2.5)*, $\sqrt{k_1}\ A^2(n/k_1) \to \infty$, 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, provided we choose a small value of ϵ. For instance, if we choose $\epsilon = 0.001$, we get $\rho > -249.75$. Then, if we work with any of the ρ-estimators in this section, computed at the level k_1, $\{\widehat{\rho} - \rho\}$ is of the order of $A(n/k_1) = O\left(n^{\epsilon \times \rho}\right)$, which is of smaller order than $1/\ln n$, a condition needed in some of the results in the paper.*

Remark 2.4. *Consequently, under the conditions in Remark 2.3 and for any intermediate k, $(\widehat{\rho} - \rho)\ln(n/k) = o_p(1)$ and then, $\sqrt{k}\ A(n/k)\,(\widehat{\rho} - \rho)\ln(n/k) = o_p(1)$ whenever $\sqrt{k}\ A(n/k) \to \lambda$, finite, a result needed later on in the proof of Theorem 2.2.*

2.2.2 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 (1.12), and based on the scaled log-spacings U_i in (1.8), $1 \le i \le k$. Then, with $d_k(\alpha)$ and $D_k(\alpha)$ given in (1.10), since the denominator in (1.12) converges in probability towards

$$\gamma/(1-r)^2 - \gamma/(1-2r) = -\gamma\ r^2/((1-r)^2(1-2r))$$

(see Lemma 5.1), $\widehat{\beta}(k;\widehat{\rho})$ is asymptotically equivalent to

$$-\frac{(1-\rho)^2(1-2\rho)}{\gamma\ \rho^2}\left(\frac{k}{n}\right)^{\widehat{\rho}}\left\{d_k(1-\widehat{\rho}) \times D_k(1) - D_k(1-\widehat{\rho})\right\},$$

provided that $\widehat{\rho}$ is consistent for the estimation of ρ.

The following result has been proved in Gomes and Martins (2002) and Gomes *et al.* (2004b):

Proposition 2.2 (Gomes and Martins, 2002; Gomes *et al.*, 2004b)**.** *If the second order condition* (1.2) *holds, with $A(t) = \gamma\ \beta\ t^\rho$, $\rho < 0$, if $k = k_n$ is a sequence* *of intermediate positive integers, i.e.* (1.6) *holds, and if* $\sqrt{k}\ A(n/k) \underset{n\to\infty}{\longrightarrow} \infty$, *then, with* $\widehat{\beta}(k;r)$ *given in* (1.12), $\widehat{\beta}(k;\rho)$ *converges in probability towards* β, *as* $n \to \infty$. *Moreover, if*

$$(\widehat{\rho}-\rho)\ln n = o_p(1), \ \ as\ n \to \infty, \tag{2.7}$$

$\widehat{\beta}(k;\widehat{\rho})$ *is consistent for the estimation of* β. *We may further say that*

$$\widehat{\beta}(k;\widehat{\rho}(k;\tau)) - \beta \ \overset{p}{\sim}\ -\beta\ \ln(n/k)\,(\widehat{\rho}(k;\tau)-\rho)\,, \tag{2.8}$$

with $\widehat{\rho}(k;\tau)$ *given in* (2.4). *Consequently, the statistic* $\widehat{\beta}(k;\widehat{\rho}(k;\tau))$ *is consistent for the estimation of* β *whenever* (1.6) *holds and* $\sqrt{k}\ A(n/k)/\ln(n/k) \to \infty$. *For models in* (1.3), $\widehat{\beta}(k;\widehat{\rho}(k;\tau)) - \beta = O_p\left(\ln(n/k)/\left(\sqrt{k}\ A(n/k)\right)\right)$ *whenever* $\sqrt{k}\ A^2(n/k) \to \lambda_{_A}$, *finite. If* $\sqrt{k}\ A^2(n/k) \to \infty$, *then* $\widehat{\beta}(k;\widehat{\rho}(k;\tau)) - \beta = O_p\left(\ln(n/k)\ A(n/k)\right)$.

Remark 2.5. *Note that for models in* (1.3), *when we consider the level* k_1 *in* (2.5), *with* $\sqrt{k_1}\ A^2(n/k_1) \to \infty$, *and* $\widehat{\beta} \equiv \widehat{\beta}(k_1;\widehat{\rho})$, *with* $\widehat{\rho}$ *any of the estimators in* (2.4), *computed also at the same level* k_1, *Proposition 2.2 enables us to say that* $\left\{\widehat{\beta}-\beta\right\}$ *is of the order of* $\ln(n/k_1)\ A(n/k_1) = O\left((\ln n)\ n^{\epsilon\times\rho}\right) = o(1/\ln n)$, *as* $n \to \infty$.

2.3 Tail index estimation based on the estimation of ρ at a lower threshold

Let us assume first that we estimate both β and ρ externally at the level k_1 in (2.5). We may state the following:

Theorem 2.2. *Under the conditions of Theorem 2.1, and for any of the estimators* $\widehat{\rho} = \widehat{\rho}(k_1;\tau)$ *and* $\widehat{\beta} = \widehat{\beta}(k_1;\widehat{\rho})$, *with* $\widehat{\beta}(k;r)$, $\widehat{\rho}(k;\tau)$ *and* k_1 *given in* (1.12), (2.4) *and* (2.5), *respectively, let us consider the class of tail index estimators* $ML_{\widehat{\beta},\widehat{\rho}}(k)$ *in* (1.13). *Let us further assume that* (2.7) *holds.*

Then, $\sqrt{k}\left\{ML_{\widehat{\beta},\widehat{\rho}}(k)-\gamma\right\}$ *is asymptotically normal with null mean value and variance equal to* γ^2*, not only when* $\sqrt{k}\ A(n/k)\ \rightarrow\ 0$*, but also whenever* $\sqrt{k}\ A(n/k)\rightarrow\lambda\neq 0$*, finite.*

If we consider γ and β estimated at the same level, we are going to have an increase in the asymptotic variance of our final tail index estimators. Indeed, as stated in Corollary 2.1 of Theorem 2.1 in Gomes and Martins (2002), for the tail index estimator in (1.11), Theorem 3.2 in Gomes *et al.* (2004b), for the tail index estimator $WH_{\widehat{\beta}(k;\widehat{\rho}),\ \widehat{\rho}}$ and Theorem 3.2 in Caeiro *et al.* (2004), for the tail index estimator $\overline{H}_{\widehat{\beta}(k;\widehat{\rho}),\ \widehat{\rho}}$, we may state:

Proposition 2.3 (Gomes and Martins, 2002; Gomes *et al.*, 2004b; Caeiro *et al.*, 2004)**.** *Under the second order framework in* (1.2)*, if* $k=k_n$ *is a sequence of intermediate integers, i.e.,* (1.6) *holds, and if* $\sqrt{k}\ A(n/k)\underset{n\rightarrow\infty}{\longrightarrow}\lambda$*, finite, non necessarily null, then, with* UH *denoting any of the statistics* ML*,* WH *or* $\overline{H}$ *in* (1.13)*,* (1.14) *and* (1.15)*, respectively, and* $\widehat{\rho}$ *any consistent estimator of the second order parameter* ρ*,*

$$\sqrt{k}\left(UH_{\widehat{\beta}(k;\widehat{\rho}),\ \widehat{\rho}}\ (k)-\gamma\right)\underset{n\rightarrow\infty}{\overset{d}{\longrightarrow}}\ Normal\left(0,\ \sigma_2^2:=\gamma^2\ \left(\frac{1-\rho}{\rho}\right)^2\right), \qquad (2.9)$$

i.e., the asymptotic variance of $UH_{\widehat{\beta}(k;\widehat{\rho}),\ \widehat{\rho}}(k)$ *increases of a factor* $\left((1-\rho)/\rho\right)^2>1$ *for every* $\rho\leq 0$*.*

Assuming again $(\beta,\ \rho)$ known, we may further state the following result for models in (1.3) and for the ML-estimators:

Theorem 2.3. *For models in* (1.3)*, with* $(\beta,\ \rho)$ *known, if* $k=k_n$ *is a sequence of intermediate integers, i.e.,* (1.6) *holds, and if* $\sqrt{k}\ A(n/k)\rightarrow\infty$*, with* $\sqrt{k}\ A^2(n/k)$ *converging towards* λ_A*, finite, as* $n\rightarrow\infty$*, the asymptotic variance of* $ML_{\widehat{\beta}(k;\rho),\rho}(k)$ *is kept equal to* $\left(\gamma(1-\rho)/\rho\right)^2$*, being its asymptotic bias given* *by*

$$b^*_{_{ML}} = (\beta - \beta')(1-\rho)\lambda_{_A}/(\gamma\ \beta(1-2\rho)(1-3\rho)), \tag{2.10}$$

again with β *and* β' *given in* (1.4) *and* (1.5), *respectively.*

Remark 2.6. *For models in* (1.3) *and* $\lambda_{_A} \neq 0$ *in Theorem 2.3,* $b^*_{_{ML}} = 0$ *if and only if* $\beta = \beta'$, *and again, this holds for Burr and Generalized Pareto underlying models.*

Remark 2.7. *If we compare Theorem 2.2 and Proposition 2.3, 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* $((1-\rho)/\rho)^2$, *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 also the three parameters are computed at the same level* k). *These authors get an asymptotic variance ruled by* $\sigma^2_{_{FH}} := \gamma^2\left((1-\rho)/\rho\right)^4$, *and we have* $\sigma_1 < \sigma_2 < \sigma_{_{FH}}$ *for all* $\rho \le 0$. *Consequently, 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 tail index* γ.

Note however that when we look at Theorems 2.2 and 2.3, we see that, for $(\beta,\ \rho)$ *known,* $\left(b_{_{ML}}/b^*_{_{ML}}\right)^2 = ((1-3\rho)/(1-\rho))^2$ *is an increasing function of* $|\rho|$, *always greater than one, for* $\rho < 0$, *i.e., there may be here again a compromise between bias and variance.*

3 Finite sample behaviour of the estimators

3.1 Simulated models

In the simulations we have considered the following models:

- 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 *Generalized Pareto* (GP) model, with d.f. $F(x) = 1 - (1 + \gamma x)^{-1/\gamma}$, $x \geq 0$, $\gamma > 0$, for which $\rho = -\gamma$, $\beta = 1$, $\beta' = 1$.

3.2 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 (2.5), with $\epsilon = 0.001$, the same level we have used for the estimation of ρ. We use the notation $\widehat{\beta}_{j1} = \widehat{\beta}(k_1; \widehat{\rho}_j)$, $j = 0, 1$, with $\beta(k; r)$ and $\widehat{\rho}_j$, $j = 0, 1$, given in (1.12) and (2.6), respectively. Similarly to what has been done in Gomes *et al.* (2004b) for the WH-estimator, in (1.14), and in Caeiro *et al.* (2004) for the $\overline{H}$-estimator, in (1.15), these estimators of ρ and β have been also incorporated in the ML-estimator, leading to $ML_0(k) \equiv ML_{\widehat{\beta}_{01},\ \widehat{\rho}_0}(k)$ or to $ML_1(k) \equiv ML_{\widehat{\beta}_{11},\ \widehat{\rho}_1}(k)$.

The simulations show that the tail index estimators $ML_j(k) \equiv ML_{\widehat{\beta}_{j1},\ \widehat{\rho}_j}(k)$, $WH_j(k) \equiv WH_{\widehat{\beta}_{j1},\ \widehat{\rho}_j}(k)$ and $\overline{H}_j(k) \equiv \overline{H}_{\widehat{\beta}_{j1},\ \widehat{\rho}_j}(k)$, j equal to either 0 or 1, according as $|\rho| \leq 1$ or $|\rho| > 1$, seem to work reasonably well, as illustrated in Figures from 1 till 4. 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 ML_j, WH_j, and $\overline{H}_j$ (*left*), $\widetilde{ML}_j \equiv ML_{\widehat{\beta}(k;\widehat{\rho}_j),\ \widehat{\rho}_j}$, $\widetilde{WH}_j \equiv WH_{\widehat{\beta}(k;\widehat{\rho}_j),\ \widehat{\rho}_j}$ and $\widetilde{H}_j \equiv \overline{H}_{\widehat{\beta}(k;\widehat{\rho}_j),\ \widehat{\rho}_j}$ (*right*), with $j = 0$ or $j = 1$, according as $|\rho| \leq 1$ or $|\rho| > 1$ and the r.v.'s $ML \equiv ML_{\beta,\rho}$, $WH \equiv WH_{\beta,\rho}$ and $\overline{H} \equiv \overline{H}_{\beta,\rho}$ (*center*). The discrepancy, in some of the models, between the behaviour of the estimators under study, 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.

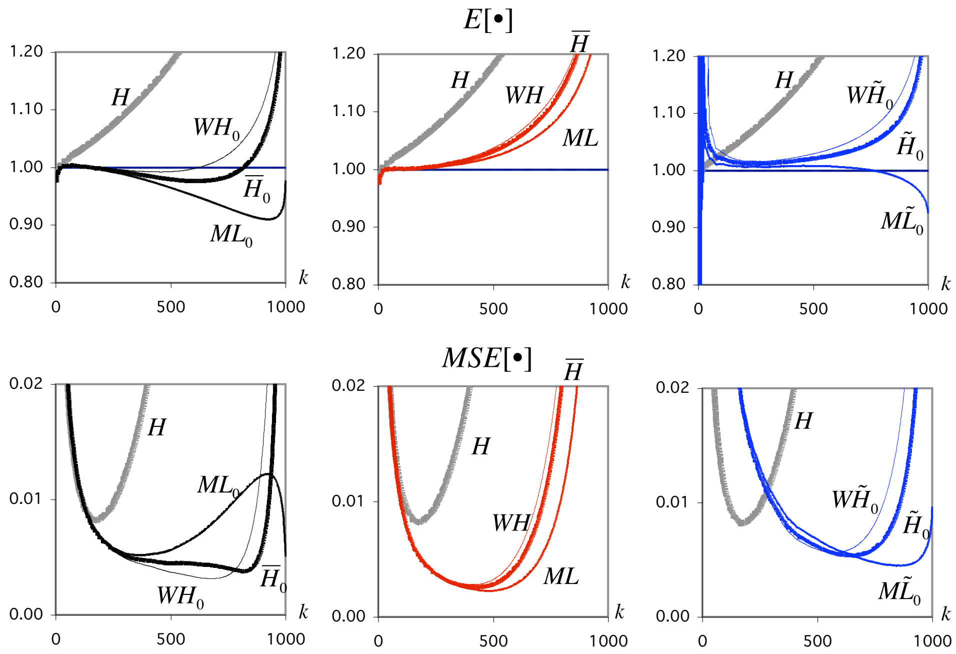

Figure 1: Underlying *Fréchet* parent with $\gamma = 1$ ($\rho = -1$).

Remark 3.1. *For the Fréchet model (Figure 1), the* $UH_{\widehat{\beta},\widehat{\rho}}$ *estimators exhibit a negative bias up to moderate values of* k *and the* ML_0 *statistic is the one exhibiting the worst performance in terms of bias and minimum mean squared error. The* WH_0 *estimator exhibits the best performance among the three statistics considered. Things work the other way round, either with the r.v.'s* UH *(Figure 1, center) or with the statistics* $\widetilde{UH}_0$ *(Figure 2, right).*

Remark 3.2. *For a Generalized Pareto model, we may draw the following comments:*

- *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, 3 and 4). Indeed* $b_{ML} = 0$ *(see Remark 2.1), 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 Generalized Pareto models.*

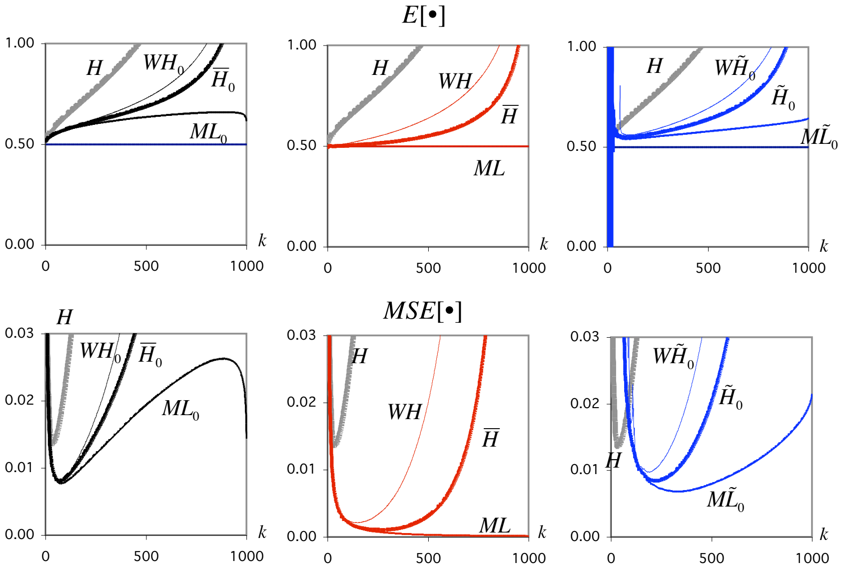

Figure 2: Underlying GP parent with $\gamma = 0.5$ ($\rho = -0.5$).

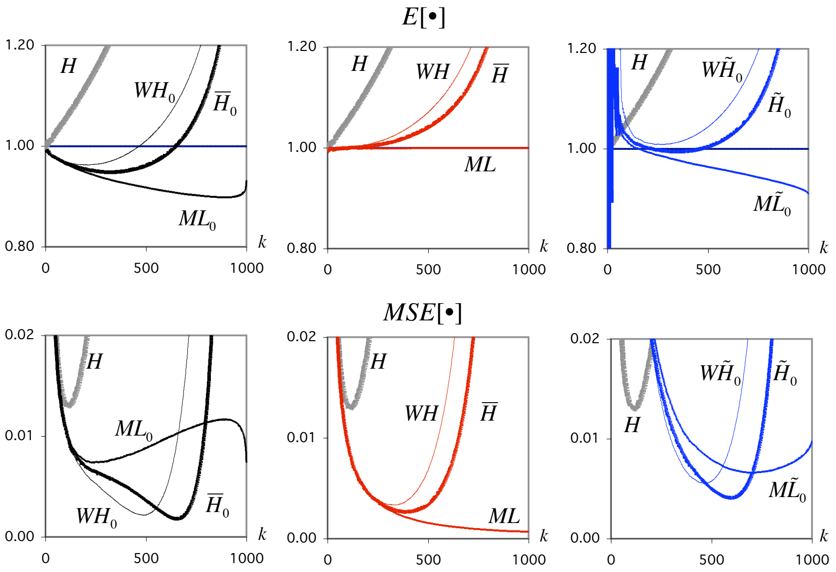

Figure 3: Underlying GP parent with $\gamma = 1$ ($\rho = -1$).

- *For values of* $\rho > -1$ *(Figure 2), the estimators exhibit a positive bias, overestimating the true value of the parameter, and the* ML_0*-statistic is better than the* $\overline{H}_0$*-statistic, which on its turn behaves better than the*

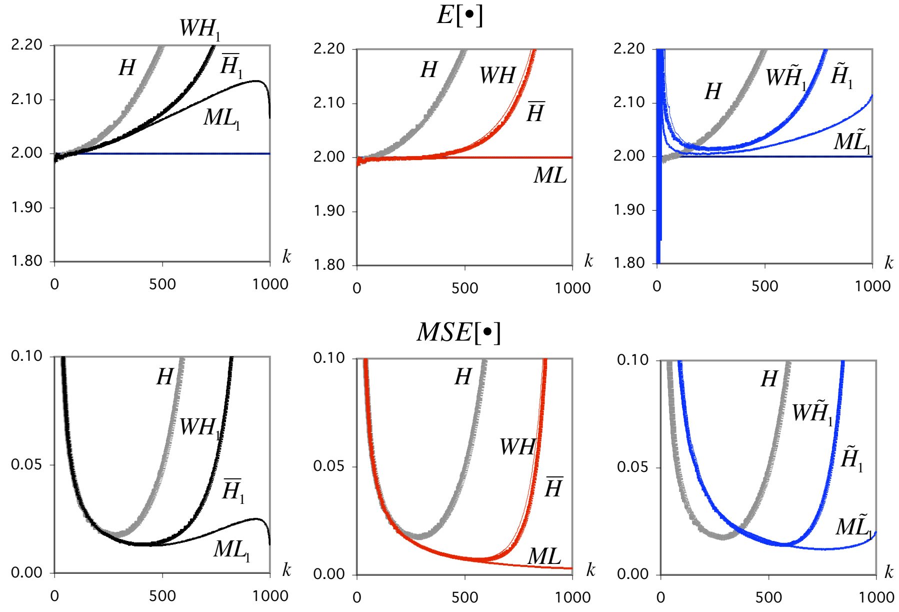

Figure 4: Underlying GP parent with $\gamma = 2$ ($\rho = -2$).

WH_0-statistic, both regarding bias and mean squared error 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 tail index).

- *For $\rho = -1$ (Figure 3), a similar pattern appears if we consider β and ρ known. But if we estimate both β and ρ through $\widehat{\beta}_{01}$ and $\widehat{\rho}_{01}$, the ML_0-statistic turns out to be the worst one, like in the Fréchet situation; the $\overline{H}_0$ statistic is then the best one regarding MSE at the optimal level, but the WH-statistic is the one with the smallest bias for not too large values of k. If we estimate only ρ through $\widehat{\rho}_0$, the $\widetilde{H}_0$ statistic is the best one, followed by the $\widetilde{WH}_0$-statistic, being the $\widetilde{ML}_0$-statistic the worst one, both in terms of bias, as well as minimum mean squared error.*
- *For $\rho < -1$ (Figure 4), we need to use $\widehat{\rho}_1$ (instead of $\widehat{\rho}_0$) 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 the $\overline{H}_1$ and*

the WH_1-statistics almost equivalent.

3.3 Comparative behaviour at optimal levels

In Tables 1, 2, 3 and 4, for the above mentioned $Frechet(\gamma = 1)$, $GP(\gamma = 0.5)$, $GP(\gamma = 1)$ and $GP(\gamma = 2)$ models, respectively, and for each entry related to the Hill estimator H in (1.9), the tail index estimators UH_j and $\widetilde{UH}_j$, $j = 0$ or 1 according as $|\rho| \le 1$ or $|\rho > 1$, as well as 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. For details on multi-sample simulation, see Gomes and Oliveira (2002) and references therein. The optimal sample fraction is, for any estimator $T_n(k)$,

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

and, relatively to the Hill estimator $H_n(k)$ in (1.9), the $REFF$ indicator is

$$REFF_T := \frac{\sqrt{MSE\left[H_n\left(k_0^{(H)}(n)\right)\right]}}{\sqrt{MSE\left[T_n\left(k_0^{(T)}(n)\right)\right]}}.$$

For any value of n, and among the seven estimators considered, the minimum mean squared error and the minimum squared bias are underlined.

Table 1: Simulated mean values (first row) and MSE's (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	0.3260 / 1.0257	0.2810 / 1.0686	0.2222 / 1.0559	0.1743 / 1.0551	0.1377 / 1.0307
	0.0435 / 1.0000	0.0261 / 1.0000	0.0134 / 1.0000	0.0081 / 1.0000	0.0050 / 1.0000
ML_0	0.5690 / 0.8199	0.5920 / 0.9661	0.8260 / 0.9769	0.8075 / 1.0100	0.9995 / 0.9853
	0.0370 / 1.0838	0.0211 / 1.1125	0.0095 / 1.1852	0.0051 / 1.2685	0.0026 / 1.4024
WH_0	0.8160 / 0.9629	0.7560 / 1.0137	0.7016 / 1.0038	0.6777 / 1.0302	0.6503 / 1.0007
	0.0195 / 1.4944	0.0121 / 1.4666	0.0058 / 1.5168	0.0031 / 1.6160	0.0017 / 1.7306
$\overline{H}_0$	0.8770 / 0.9514	0.8405 / 1.0047	0.8186 / 0.9980	0.8079 / 1.0263	0.8082 / 0.9732
	0.0236 / 1.3582	0.0148 / 1.3305	0.0071 / 1.3762	0.0038 / 1.4686	0.0020 / 1.5759
$\widetilde{ML}_0$	0.9470 / 0.8493	0.9200 / 0.9732	0.8698 / 0.9917	0.8554 / 1.0187	0.8340 / 0.9794
	0.0365 / 1.0919	0.0201 / 1.1385	0.0087 / 1.2387	0.0045 / 1.3489	0.0023 / 1.4799
$\widetilde{WH}_0$	0.8110 / 0.9915	0.7360 / 1.0621	0.6474 / 1.0694	0.5665 / 1.0568	0.5108 / 1.0031
	0.0297 / 1.2112	0.0183 / 1.1942	0.0094 / 1.1943	0.0056 / 1.2080	0.0034 / 1.2244
$\widetilde{H}_0$	0.8560 / 0.9725	0.7945 / 1.0483	0.7112 / 1.0586	0.6432 / 1.0572	0.5786 / 0.9943
	0.0307 / 1.1907	0.0187 / 1.1831	0.0092 / 1.2046	0.0054 / 1.2313	0.0032 / 1.2609
ML	0.6420 / 0.9860	0.5985 / 1.0170	0.5170 / 1.0369	0.4732 / 1.0391	0.4292 / 1.0119
	0.0154 / 1.6777	0.0087 / 1.7343	0.0040 / 1.8320	0.0022 / 1.9093	0.0013 / 2.0006
WH	0.5800 / 0.9603	0.5125 / 1.0188	0.4504 / 1.0518	0.3950 / 1.0407	0.3573 / 1.0027
	0.0184 / 1.5393	0.0105 / 1.5769	0.0049 / 1.6579	0.0027 / 1.7226	0.0015 / 1.8049
$\overline{H}$	0.5870 / 0.9625	0.5365 / 1.0121	0.4822 / 1.0477	0.4355 / 1.0408	0.3788 / 1.0084
	0.0179 / 1.5597	0.0101 / 1.6088	0.0046 / 1.7102	0.0026 / 1.7857	0.0014 / 1.8736

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

n	100	200	500	1000	2000
H	0.1030 / 0.7418	0.0770 / 0.6455	0.0512 / 0.6322	0.0400 / 0.6018	0.0278 / 0.5853
	0.0579 / 1.0000	0.0366 / 1.0000	0.0203 / 1.0000	0.0138 / 1.0000	0.0091 / 1.0000
ML_0	0.3060 / 0.6361	0.2160 / 0.6333	0.1074 / 0.6056	0.0761 / 0.5827	0.0510 / 0.5581
	0.0234 / 1.5720	0.0169 / 1.4741	0.0106 / 1.3825	0.0077 / 1.3392	0.0056 / 1.2740
WH_0	0.2020 / 0.6691	0.1440 / 0.6140	0.0996 / 0.6071	0.0714 / 0.5862	0.0485 / 0.5577
	0.0289 / 1.4156	0.0192 / 1.3818	0.0114 / 1.3361	0.0081 / 1.3084	0.0058 / 1.2567
$\overline{H}_0$	0.2340 / 0.6405	0.1650 / 0.6400	0.1030 / 0.6067	0.0725 / 0.5878	0.0485 / 0.5573
	0.0288 / 1.4176	0.0191 / 1.3838	0.0113 / 1.3387	0.0081 / 1.3103	0.0058 / 1.2568
$\widetilde{ML}_0$	0.7950 / 0.6518	0.6355 / 0.6278	0.4214 / 0.6019	0.3102 / 0.5778	0.2397 / 0.5682
	0.0223 / 1.6115	0.0158 / 1.5250	0.0096 / 1.4522	0.0069 / 1.4201	0.0049 / 1.3704
$\widetilde{WH}_0$	0.4500 / 0.7322	0.3335 / 0.6486	0.2452 / 0.6121	0.1911 / 0.5997	0.1383 / 0.5764
	0.0508 / 1.0683	0.0297 / 1.1100	0.0154 / 1.1494	0.0098 / 1.1872	0.0063 / 1.2048
$\widetilde{H}_0$	0.4640 / 0.6969	0.3890 / 0.6335	0.2888 / 0.6003	0.2256 / 0.5987	0.1685 / 0.5581
	0.0395 / 1.2113	0.0238 / 1.2396	0.0128 / 1.2605	0.0085 / 1.2800	0.0056 / 1.2707
ML	0.9870 / 0.5071	0.9850 / 0.5125	0.9912 / 0.5042	0.9901 / 0.5044	0.9972 / 0.5031
	0.0017 / 5.8132	0.0009 / 6.5672	0.0003 / 7.8305	0.0002 / 9.1844	0.0001 / 10.4868
WH	0.2730 / 0.5731	0.2210 / 0.5659	0.1740 / 0.5372	0.1462 / 0.5332	0.1171 / 0.5299
	0.0115 / 2.2455	0.0067 / 2.3317	0.0035 / 2.4191	0.0021 / 2.5422	0.0013 / 2.6239
$\overline{H}$	0.3910 / 0.5494	0.3530 / 0.5372	0.3016 / 0.5360	0.2621 / 0.5198	0.2082 / 0.5211
	0.0068 / 2.9183	0.0037 / 3.1280	0.0018 / 3.3673	0.0011 / 3.5966	0.0006 / 3.8348

Table 3: Simulated mean values (first row) and MSE's (second row) at optimal levels of the different estimators and r.v.'s under study, for Generalized Pareto parents with $\gamma = 1$ ($\rho = -1,\ \beta = 1$).

n	100	200	500	1000	2000
H	0.2290 / 1.2013 0.0715 / 1.0000	0.1900 / 1.1803 0.0419 / 1.0000	0.1410 / 1.1126 0.0212 / 1.0000	0.1139 / 1.1050 0.0130 / 1.0000	0.0947 / 1.0604 0.0079 / 1.0000
ML_0	0.9110 / 0.9253 0.0250 / 1.6926	0.9875 / 0.9411 0.0166 / 1.5902	0.8662 / 0.9521 0.0102 / 1.4427	0.6152 / 0.9478 0.0072 / 1.3400	0.3382 / 0.9656 0.0040 / 1.2743
WH_0	0.5620 / 1.0505 0.0200 / 1.8924	0.5130 / 1.0636 0.0106 / 1.9935	0.4940 / 1.0267 0.0043 / 2.2274	0.4821 / 1.0164 0.0022 / 2.4350	0.4717 / $\underline{1.0164}$ 0.0011 / 2.6354
$\overline{H}_0$	0.6600 / $\underline{1.0014}$ $\underline{0.0177}$ / 2.0138	0.6525 / $\underline{1.0090}$ $\underline{0.0091}$ / 2.1480	0.6516 / $\underline{1.0210}$ $\underline{0.0036}$ / 2.4389	0.6503 / $\underline{1.0080}$ $\underline{0.0018}$ / 2.6972	0.6441 / 1.0237 $\underline{0.0009}$ / 2.9622
$\widetilde{ML}_0$	0.9900 / 0.9485 0.0260 / 1.6570	0.9890 / 0.9442 0.0166 / 1.5910	0.8178 / 0.9646 0.0098 / 1.4686	0.7015 / 0.9623 0.0066 / 1.4034	0.5915 / 0.9579 0.0044 / 1.3438
$\widetilde{WH}_0$	0.6360 / 1.1764 0.0419 / 1.3056	0.5710 / 1.1112 0.0223 / 1.3718	0.5070 / 1.0853 0.0100 / 1.4540	0.4616 / 1.0493 0.0056 / 1.5232	0.4272 / 1.0294 0.0031 / 1.5918
$\widetilde{H}_0$	0.7040 / 1.1356 0.0339 / 1.4524	0.6690 / 1.0972 0.0180 / 1.5269	0.6244 / 1.0612 0.0077 / 1.6552	0.5936 / 1.0385 0.0042 / 1.7704	0.5637 / 1.0190 0.0022 / 1.8794
ML	0.9880 / 1.0124 0.0072 / 3.1467	0.9855 / 1.0237 0.0036 / 3.4214	0.9924 / 1.0145 0.0014 / 3.8773	0.9914 / 1.0075 0.0007 / 4.3093	0.9972 / 1.0070 0.0004 / 4.7491
WH	0.4610 / 1.0876 0.0221 / 1.7988	0.4285 / 1.0780 0.0122 / 1.8528	0.3612 / 1.0604 0.0058 / 1.9151	0.3250 / 1.0351 0.0033 / 1.9752	0.2877 / 1.0151 0.0019 / 2.0472
$\overline{H}$	0.5200 / 1.0824 0.0195 / 1.9159	0.4855 / 1.0963 0.0103 / 2.0210	0.4248 / 1.0393 0.0047 / 2.1152	0.3873 / 1.0224 0.0026 / 2.2192	0.3530 / 1.0107 0.0015 / 2.3292

Table 4: Simulated mean values (first row) and MSE's (second row) at optimal levels of the different estimators and r.v.'s under study, for Generalized Pareto parents with $\gamma = 2$ ($\rho = -2,\ \beta = 1$).

n	100	200	500	1000	2000
H	0.4150 / 2.1787 0.1165 / 1.0000	0.3585 / 1.9682 0.0641 / 1.0000	0.3186 / $\underline{2.0175}$ 0.0302 / 1.0000	0.2902 / 2.0681 0.0175 / 1.0000	0.2505 / 2.0688 0.0100 / 1.0000
ML_1	0.8170 / 2.1840 $\underline{0.0708}$ / 1.2829	0.6470 / 2.0120 $\underline{0.0430}$ / 1.2213	0.6630 / 2.0483 0.0212 / 1.1942	0.6567 / 2.0770 0.0127 / 1.1729	0.9995 / 2.0943 0.0071 / 1.1803
WH_1	0.6230 / $\underline{2.1552}$ 0.0813 / 1.1970	0.5535 / 2.0240 0.0467 / 1.1707	0.4700 / 2.0479 0.0225 / 1.1586	0.3958 / 2.0512 0.0131 / 1.1530	0.3494 / 2.0410 0.0075 / 1.1493
$\overline{H}_1$	0.6180 / 2.1670 0.0828 / 1.1863	0.5445 / 2.0407 0.0472 / 1.1647	0.4700 / 2.0495 0.0226 / 1.1558	0.3958 / $\underline{2.0509}$ 0.0132 / 1.1515	0.3494 / $\underline{2.0409}$ 0.0076 / 1.1483
$\widetilde{ML}_1$	0.9900 / 2.1939 0.0721 / 1.2719	0.9345 / $\underline{2.0004}$ 0.0437 / 1.2111	0.8278 / 2.0337 $\underline{0.0208}$ / 1.2041	0.7677 / 2.0773 $\underline{0.0122}$ / 1.1972	0.6809 / 2.0554 $\underline{0.0070}$ / 1.1907
$\widetilde{WH}_1$	0.7110 / 2.2404 0.1001 / 1.0786	0.6515 / 2.0024 0.0542 / 1.0871	0.5948 / 2.0380 0.0251 / 1.0975	0.5483 / 2.0699 0.0143 / 1.1047	0.5097 / 2.0454 0.0080 / 1.1150
$\widetilde{H}_1$	0.7100 / 2.2404 0.1002 / 1.0780	0.6570 / 2.0007 0.0542 / 1.0876	0.6040 / 2.0409 0.0249 / 1.1009	0.5610 / 2.0708 0.0142 / 1.1093	0.5129 / 2.0412 0.0079 / 1.1198
ML	0.9900 / 2.0650 0.0315 / 1.9232	0.9935 / 1.9213 0.0156 / 2.0299	0.9948 / 1.9924 0.0062 / 2.2110	0.9934 / 2.0109 0.0031 / 2.3819	0.9992 / 2.0146 0.0015 / 2.5412
WH	0.6590 / 2.0911 0.0578 / 1.4202	0.6325 / 1.9766 0.0305 / 1.4495	0.5760 / 2.0357 0.0135 / 1.4957	0.5401 / 2.0572 0.0075 / 1.5282	0.5051 / 2.0616 0.0040 / 1.5732
$\overline{H}$	0.6690 / 2.1030 0.0575 / 1.4230	0.6470 / 1.9762 0.0297 / 1.4696	0.6044 / 2.0474 0.0130 / 1.5253	0.5738 / 2.0532 0.0071 / 1.5703	0.5325 / 2.0567 0.0038 / 1.6217

3.4 An overall conclusion

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 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.

4 A case-study

We shall here consider the performance of the above mentioned estimators in the analysis of 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 5 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.

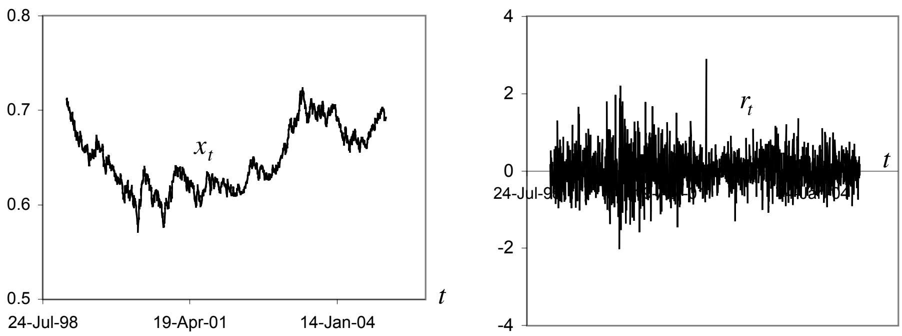

Figure 5: Daily Exchange Rates (*left*) and Daily Log-Returns (*right*) on Euro-UK Pound Exchange Rate.

In Figure 6, and working with the $n_0 = 725$ positive log-returns, we now picture the sample paths of $\widehat{\rho}(k;\tau)$ in (2.4) for $\tau = 0,$ and 1 (*left*), together with the sample paths of $\widehat{\beta}(k;\widehat{\rho}_0)$ in (1.12), for $\tau = 0$, as functions of k. The sample paths of the ρ-estimates associated to $\tau = 0$ and $\tau = 1$ lead us 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, $\widehat{\rho}_0 = \widehat{\rho}_0(721) = -0.65$ and $\widehat{\beta}_0 = \widehat{\beta}_{\widehat{\rho}_0}(721) = 1.03$.

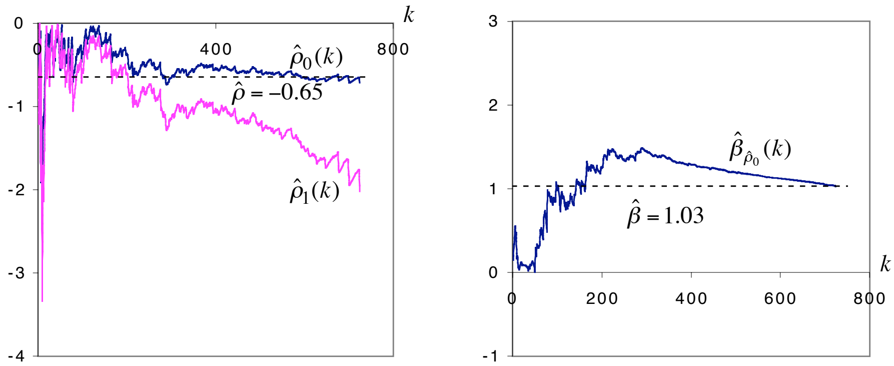

Figure 6: Estimates of the second order parameter ρ (*left*) and of the tail index γ (*right*) for the Daily Log-Returns on the Euro-UK Pound.

The sample paths of the classical Hill estimator in (1.9) (H) and of the three reduced bias, second order tail index estimates discussed in this paper, associated to $\widehat{\rho}_0 = -0.65$ and $\widehat{\beta}_0 = 1.03$, are pictured in Figure 7.

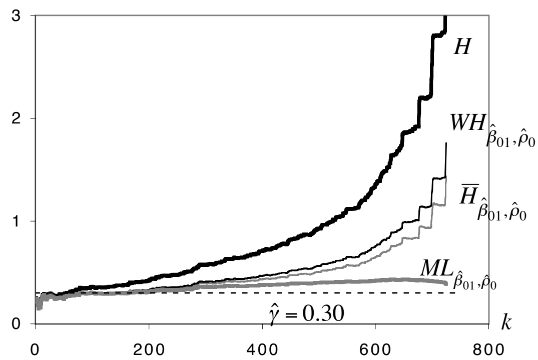

Figure 7: Estimates of the Hill estimator in (1.9) and the three UH-estimators in (1.13), (1.14) and (1.15), for the Daily Log-Returns on the Euro-UK Pound.

4.1 How to estimate γ?

The Hill estimator exhibits a relevant bias, as may be seen from Figure 7, and we are for sure a long way from the strict Pareto model. The other estimators, WH, $\overline{H}$ and ML, 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 herewith considered. More important than this: we know that any estimate considered on the basis of $ML_{\widehat{\beta}_{01},\, \widehat{\rho}_0}(k)$ (or any of the other two 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 $\left[n_0^{-2\widehat{\rho}/(1-2\widehat{\rho})}/4\right] = 10$ and $\left[4 \times n_0^{-2\widehat{\rho}/(1-2\widehat{\rho})}\right] = 165$, chosen in an heuristic way. If we use this same criterion on the estimates WH and $\overline{H}$ we are also led to the same estimate, $\widehat{\gamma}_{WH} \equiv \widehat{\gamma}_{\overline{H}} = 0.30$.

Since we have obtained estimates of the second order parameters β and ρ, we may also proceed to the estimation of the optimal k for the Hill estimator, i.e., the level k which minimizes the approximation to the asymptotic mean squared error, provided by $\gamma^2 \left(1/k + \beta^2\, (n_0/k)^{2\rho} / (1-\rho)^2\right)$. We then obtain,

$$\widehat{k}_0^H = \left[\left(\frac{(1-\widehat{\rho}_0)^2\; n_0^{-2\widehat{\rho}_0}}{-2\widehat{\rho}_0 \widehat{\beta}_0^2}\right)^{1/(1-2\widehat{\rho}_0)}\right] = 56 \quad \Longrightarrow \quad \widehat{\gamma} = H_{n_0}(56) = 0.2986,$$

not a long way from the estimate obtained before. The development of adequate techniques for the adaptive choice of the optimal threshold for this type of second order reduced bias tail index estimators is needed, being indeed an interesting topic of research, but is outside the scope of the present paper.

5 Proofs of theorems in section 2

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

$$\overline{W}_k^{(\alpha)} = (2\alpha - 1)\sqrt{(2\alpha - 1)\ k/2}\ \left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1}\ln\left(\frac{i}{k}\right)E_i + \frac{1}{\alpha^2}\right), \quad (5.1)$$

$$D_k'(\alpha) = \frac{d\ D_k(\alpha)}{d\alpha} := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\alpha-1}\ln\left(\frac{i}{k}\right)U_i, \quad (5.2)$$

with $D_k(\alpha)$ in (1.10). With the same kind of reasoning as in Gomes *et al.* (2005), we state:

Lemma 5.1. *Under the second order framework in* (1.2)*, for intermediate k-sequences, i.e., whenever* (1.6) *holds, and with* U_i *given in* (1.8)*, we may guarantee that, for any real* $\alpha \geq 1$*,*

$$D_k(\alpha) \ \stackrel{d}{=}\ \frac{\gamma}{\alpha} + \frac{\gamma\ \overline{Z}_k^{(\alpha)}}{\sqrt{(2\alpha-1)\ k}} + \frac{A(n/k)}{\alpha-\rho}(1+o_p(1)), \quad (5.3)$$

and

$$D_k'(\alpha) \ \stackrel{d}{=}\ -\frac{\gamma}{\alpha^2} + \frac{\gamma\ \overline{W}_k^{(\alpha)}}{(2\alpha-1)\sqrt{(2\alpha-1)\ k/2}} - \frac{A(n/k)}{(\alpha-\rho)^2}(1+o_p(1)), \quad (5.4)$$

with $D_k(\alpha)$ *and* $D_k'(\alpha)$ *given in* (1.10) *and* (5.2)*, respectively, and where* $\overline{Z}_k^{(\alpha)}$ *and* $\overline{W}_k^{(\alpha)}$ *in* (2.1) *and* (5.1)*, respectively, are asymptotically standard normal r.v.'s. If we further assume to be working with models in* (1.3)*, and with the same notation as before, we may write*

$$D_k(\alpha) \ \stackrel{d}{=}\ \frac{\gamma}{\alpha} + \frac{\gamma}{\sqrt{(2\alpha-1)\ k}}\ \overline{Z}_k^{(\alpha)} + \frac{A(n/k)}{\alpha-\rho} + O_p\left(\frac{A(n/k)}{\sqrt{k}}\right) + \frac{\beta'\ A^2(n/k)}{\gamma\ \beta(\alpha-2\rho)}\ (1+o_p(1)), \quad (5.5)$$

with β *and* β' *given in* (1.4) *and* (1.5)*, respectively.*

Proof. (Theorem 2.1). If all parameters are known, apart from the tail index γ, we get directly from (5.3),

$$\begin{aligned} ML_{\beta,\rho}(k) &:= D_k(1) - \beta \left(\frac{n}{k}\right)^{\rho} D_k(1-\rho) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, \overline{Z}_k^{(1)} + \frac{A(n/k)}{1-\rho} \\ &\quad - \frac{A(n/k)}{\gamma} \left(\frac{\gamma}{1-\rho} + \frac{\gamma}{\sqrt{(1-2\rho)k}} \overline{Z}_k^{(1-\rho)} + \frac{A(n/k)}{1-2\rho}(1+o_p(1)) \right) \\ &\stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, \overline{Z}_k^{(1)} + o_p(A(n/k)). \end{aligned}$$

For models in (1.3), and now directly from (5.5), we get

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

Working this expression, we finally obtain,

$$\begin{aligned} ML_{\beta,\rho}(k) &\stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}}\, \overline{Z}_k^{(1)} + \frac{A^2(n/k)}{\gamma(1-2\rho)} \left(\frac{\beta'}{\beta} - 1\right)(1+o_p(1)) \\ &\qquad + O_p\left(\frac{A(n/k)}{\sqrt{k}}\right)(1+o_p(1)), \end{aligned}$$

i.e., (2.2) holds. Note that since $\sqrt{k}\ O_p\left(A(n/k)/\sqrt{k}\right) = O_p(A(n/k)) \to 0$, the summand $O_p\left(A(n/k)/\sqrt{k}\right)$ is totally irrelevant for the asymptotic bias in (2.3), that follows straightforwardly from the above obtained distributional representation. □

Proof. (Theorem 2.2). We may write,

$$\frac{\partial ML_{\beta,\rho}}{\partial \beta} = -\left(\frac{n}{k}\right)^{\rho} D_k(1-\rho) = -\frac{A(n/k)\ D_k(1-\rho)}{\gamma\ \beta}$$

and

$$\begin{aligned} \frac{\partial ML_{\beta,\rho}}{\partial \rho} &= -\beta \left(\frac{n}{k}\right)^{\rho} \left(\ln\left(\frac{n}{k}\right) D_k(1-\rho) + D_k'(1-\rho) \right) \\ &= -\frac{A(n/k)}{\gamma} \left(\ln\left(\frac{n}{k}\right) D_k(1-\rho) + D_k'(1-\rho) \right). \end{aligned}$$

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

$$\begin{aligned} ML_{\widehat{\beta},\widehat{\rho}}(k) - ML_{\beta,\rho}(k) \overset{p}{\sim} & -\frac{A(n/k)}{\gamma}\Big\{\left(\widehat{\beta}-\beta\right) D_k(1-\rho)/\beta \\ & + (\widehat{\rho}-\rho)\left(\ln(n/k) D_k(1-\rho) + D'_k(1-\rho)\right)\Big\}. \end{aligned} \tag{5.6}$$

Consequently, taking into account (2.8) for $k = k_1$, (5.3) and (5.4), we get

$$ML_{\widehat{\beta},\widehat{\rho}}(k) - ML_{\beta,\rho}(k) = O_p\left((\widehat{\rho}-\rho)\, A(n/k) \ln(k/k_1)\right).$$

Hence, if $\sqrt{k}\ A(n/k) \to \lambda$, finite, condition (2.7) enables us to guarantee that $\sqrt{k}\ (\widehat{\rho}-\rho)\ A(n/k)\ \ln(k/k_1)$ converges towards zero, as $n \to \infty$, and the results in the theorem follow. □

Proof. (Theorem 2.3). Following the steps in Gomes and Martins (2002), but working now with models in (1.3) and the distributional representation (5.5), we may write:

$$\begin{aligned} ML_{\widehat{\beta}(k;\rho),\rho}(k) &= D_k(1) - \frac{D_k(1-\rho)\left\{D_k(1)(1+o(1)) - (1-\rho)D_k(1-\rho)\right\}}{D_k(1-\rho)(1+o(1)) - (1-\rho)D_k(1-2\rho)} \\ &=: D_k(1) - \frac{\varphi_k(\rho)}{\psi_k(\rho)} \end{aligned}$$

with $D_k(\alpha)$ given in (1.10). Directly from (5.5), we get

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

and

$$\begin{aligned} \varphi_k(\rho) &\overset{d}{=} \frac{\gamma^2}{\sqrt{k}}\left(\frac{\overline{Z}_k^{(1)}}{1-\rho} - \frac{\overline{Z}_k^{(1-\rho)}}{\sqrt{1-2\rho}}\right) - \frac{\gamma\ \rho^2\ A(n/k)}{(1-\rho)^2(1-2\rho)} \\ -&\frac{\rho^2 A^2(n/k)}{(1-\rho)(1-2\rho)}\left(\frac{2\beta'}{\beta(1-3\rho)} + \frac{1}{1-2\rho}\right)(1+o_p(1)) + O_p\left(\frac{A(n/k)}{\sqrt{k}}\right)(1+o_p(1)). \end{aligned}$$

Consequently,

$$\begin{aligned} \frac{\varphi_k(\rho)}{\psi_k(\rho)} &\overset{d}{=} -\frac{\gamma}{\rho^2\sqrt{k}}\left((1-2\rho)\overline{Z}_k^{(1)} - (1-\rho)\sqrt{1-2\rho}\ \overline{Z}_k^{(1-\rho)}\right) + \frac{A(n/k)}{1-\rho} \\ &+ \frac{A^2(n/k)}{\gamma}\left(\frac{2(\beta'-\beta)}{\beta(1-3\rho)} + \frac{1}{1-2\rho}\right)(1+o_p(1)) + O_p\left(\frac{A(n/k)}{\sqrt{k}}\right)(1+o_p(1)). \end{aligned}$$

Then, with

$$\begin{aligned}\overline{\overline{Z}}_k &:= \left(\overline{Z}_k^{(1)} + \frac{(1-\rho)(1-2\rho)}{\rho^2}\left(\frac{\overline{Z}_k^{(1)}}{1-\rho} - \frac{\overline{Z}_k^{(1-\rho)}}{\sqrt{1-2\rho}}\right)\right)\\ &= \left(\frac{1-\rho}{\rho}\right)^2 \overline{Z}_k^{(1)} - \left(\frac{(1-\rho)\sqrt{1-2\rho}}{\rho^2}\right)\overline{Z}_k^{(1-\rho)},\end{aligned}$$

$$\begin{aligned}ML_{\widehat{\beta}(k;\rho),\ \rho}(k) = \gamma + \frac{\gamma}{\sqrt{k}}\ \overline{\overline{Z}}_k + O_p\left(\frac{A(n/k)}{\sqrt{k}}\right)(1+o_p(1))\\ - \frac{(\beta'-\beta)(1-\rho)A^2(n/k)}{\gamma\ \beta(1-2\rho)(1-3\rho)}(1+o_p(1)),\end{aligned}$$

and the result in (2.10) follows for the r.v. $ML_{\widehat{\beta}(k;\rho),\ \rho}(k)$. Also, since the asymptotic variance between $\overline{Z}_k^{(1)}$ and $\overline{Z}_k^{(1-\rho)}$ is given by $\sqrt{1-2\rho}/(1-\rho)$, the asymptotic variance of $\overline{\overline{Z}}_k$ 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 $\gamma^2\left\{(1-\rho)/\rho\right\}^2$, stated in the theorem.

□

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] Caeiro, F. and Gomes, M. I. (2004). *A new class of estimators of the "scale" second order parameter*. Notas e Comunicações CEAUL 20/04. Submitted.

[3] Caeiro, F., Gomes, M. I. and Pestana, D. D. (2004). *Direct reduction of bias of the classical Hill estimator.* Notas e Comunicações CEAUL 18/04. Accepted at *RevStat*.

[4] Diebolt, J. and Guillou, A. (2005). Asymptotic behaviour of regular estimators. *RevStat* **3**:1, 19-44.

[5] Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*, Springer-Verlag.

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

[7] Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distributions of the largest or smallest member of a sample *Proc. Camb. Phil. Soc.* **24**, 180-190.

[8] 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.

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

[10] Gomes, M. I. (1984). Penultimate limiting forms in extreme value theory. *Ann. Inst. Statist. Math* **A 36**, 71-85.

[11] 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.

[12] 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.

[13] Gomes, M. I. and de Haan, L. (1999). Approximation by penultimate extreme value distributions. *Extremes* **2**:1, 71-85.

[14] 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.

[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. (2002). “Asymptotically unbiased” estimators of the tail index based on external estimation of the second order parameter. *Extremes* **5**:1, 5-31.

[17] 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.

[18] 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.

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

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

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

[22] Kaufmann, E. (2000). Penultimate approximations in extreme value theory. *Extremes* **3**:1, 39-55.

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

[24] Raoult, J. P. and Worms, R. (2003). Rate of convergence for the generalized Pareto approximation of the excesses. *Adv. Appl. Probab.* **35**(4), 1007-1027.